

**Gender bias and stereotypes in linguistic example sentences**

Hadas KOTEK Massachusetts Institute of Technology hkotek@mit.edu

Rikker DOCKUM Swarthmore College rdockum1@swarthmore.edu

Sarah BABINSKI Yale University sarah.babinski@yale.edu

Christopher GEISSLER Yale University christopher.geissler@yale.edu

## **Gender bias and stereotypes in linguistic example sentences**

This paper examines the distribution of gender in arguments in example sentences in contemporary linguistics publications. Prior studies have shown that example sentences in syntax textbooks systematically under-represent women and perpetuate gender stereotypes (Macaulay and Brice 1994, 1997; Pabst, Cépeda, Kotek, and Syrett 2018). Here we examine example sentences in articles published over the past 20 years in *Language*, *Linguistic Inquiry*, and *Natural Language & Linguistic Theory*, and find striking similarities to this prior work. Among our findings, we show a stark imbalance of male (N=12006) to female (N=5670) arguments, where male-gendered arguments are more likely to be subjects, and female arguments non-subjects. We show that female-gendered arguments are more likely to be referred to using a kinship term, to exhibit positive emotions, and to be the object of affection, whereas male-gendered arguments are more likely to have occupations, to exhibit negative emotions, and to perpetrate violence. We show that this pattern has remained stable, with little change, over the course of the twenty years that we examine, leading up to the present day. We conclude with a brief discussion of possible remedies and suggestions for improvement.\*

**Keywords:** gender representation, implicit bias, syntax, publications, academia

\* We would like to thank audiences at Yale University, Brandeis, MIT, the University of Oregon, and the LSA 2020 Annual Meeting in New Orleans for questions and comments. We thank the Yale Women Faculty Forum and Claire Bovern for providing funding which supported this work. Thanks to Monica Macaulay and Katharina Pabst, as well as three anonymous reviewers and the editors Andries Coetzee and Megan Crowhurst, for their careful comments on the manuscript. The first author would like to additionally acknowledge the members of the LSA Committee on the Status of Women in Linguistics, and in particular Katharina Pabst, Paola Cépeda, and Kristen Syrett, for their support of this and related work. Finally, Special thanks to the broader community that assisted us to carry this project through: the linguists and native speakers who assisted with identification of the gender of proper names: Will Bennett, Robert Daland, Michael Yoshitaka Erlewine, Henrison Hsieh, Beste Kamali, Sonia Kasyanenko, Monica Macaulay, Maria Polinsky,

Norvin Richards, Eszter Rónai, Patricia Schneider-Zioga, Chaeyun Sheen, Jisu Sheen, Sverre Stausland, and Leah Velleman; and the Yale undergraduates who coded the data used in this study: Aarohi Srivastava, Abigail Fortier, Amelia Lake, Calvin Kaleel, Faren Roth, Georgia Michelman, Joe Class, Joshua Celli, Justin Yamamura, Karina Di Franco, Lena Venkatraman, Michael Gancz, Nanyan Wu, Nico Kidd, Oliver Shoulson, Prastik Mohanraj, Ronnie Rodriguez, Serena Puang, Shayley Martin, Slater Smith, Stella Fitzgerald, Stella Xu, and Zhiliang Fang.

**1. INTRODUCTION.** Constructed example sentences represent a main source of data for work in linguistics. As scientists, we are trained to consider our data to be an unbiased source of evidence for our claims about the language or construction under investigation. However, it has long been noted that such examples may reflect the biases of the researchers who construct them and of the world around us — examples tend to display hetero-normative gender roles, portray mainly Western protagonists such as *John* and *Mary*, and likewise reflect the dominant white culture far more than minoritized races and ethnicities (Macaulay and Brice 1994, 1997; Bergvall 1996a,b).

In a landmark study, Macaulay and Brice (1997) (henceforth M&B) analyzed example sentences in eleven then-current syntax textbooks, and concluded that “the majority of constructed example sentences in syntax textbooks are biased toward male-gendered NPs, and [ . . . ] contain highly stereotyped representations of both genders.” Twenty years later, Pabst and colleagues (2018) conducted a follow-up study of recently published textbooks in syntax to test whether the problem of gender bias has been rectified in the time since M&B published their work. These authors find that the majority of problems identified by M&B in 1997 still plague syntax textbooks today: male-gendered arguments are over-represented in linguistic examples and almost all of the stereotypes observed in M&B are still present in more recently published textbooks.

In this paper, we go beyond textbooks to test the distribution and role of arguments in constructed example sentences with regard to gender in journal papers published in three major theoretical linguistics journals over the past twenty years — covering the period between the publication of Macaulay & Brice 1997 and that of Pabst et al. 2018. In short, we observe the same issues of imbalance in example sentences used in the research literature as in textbooks. Although we observe a positive trend, at its current pace the field will not reach parity in this regard for several more decades. These findings affect all three journals we study to the same degree, suggesting that this is an entrenched way of thought in our field, which we argue must change.

Investigating the role of gender representation in example sentences fits into a broader current trend in linguistic research, turning its gaze inward and examining bias against underrepresented minorities in the field. In the past several years, within the LSA alone, we have seen the adoption of the *Guidelines for Inclusive Language* (2016);<sup>1</sup> several presentations (e.g. Pabst et al. 2018;

Kibbey 2019; Kurumada & Gardner 2019; Muller et al. 2019; Zimman 2019; Haugen & Margaris 2020; plenary events, including the panel *Our Linguistics Community: Addressing Bias, Power Dynamics, Harassment* (2018), the special film showing *Talking Black in America: The Story of African American English* (2018), and the talk *Fostering a Culture of Racial Inclusion in Linguistics: For the Children of the 9th Ward Circa 2005* (Charity Hudley 2020); special sessions such as *Sharing Our Views; Native Americans Speak About Language and Linguistics* (2018), *A Survey of Linguists and Language Researchers: Harassment, Bias, and What We Can Do About It* (2019), *Linguistic Discrimination on the University Campus* (2019), *Special Panel on Fostering a Culture of Inclusion in Linguistics* (2019), *Black Becoming for Language and Linguistics Researchers* (2020), *Queer and Trans Digital Modalities* (2020), *Hate Speech* (2020); and finally, the recent *LSA Statement on Race* and accompanying paper (Charity Hudley et al. 2018).

In short: Who we hire, who we cite, and who we signal to our students and early career researchers is a part of our field has a large impact on its makeup. The shape of the world our example sentences convey to readers — students and active researchers alike — implicitly and sometimes explicitly send powerful signals about who is welcome in our field and who is less so. This, in turn, affects the kinds of research questions that are welcome, and the kinds of answers that we expect and ultimately adopt. Limiting access to the field inevitably leads to a reduced richness of ideas, research topics, approaches, types of data collected, and more generally limits the reach and broadness of our field. It is thus in everyone's interest to increase the inclusivity of our field.

In what follows, we present the main results of M&B 1997 and Pabst et al. 2018. We then present our current study and its results, showing pervasive and expansive gender bias in the form of over-representation of male individuals and stereotypes of all genders. We conclude with a brief discussion of naturalistic and corpus data in the context of elicitation and discourse analysis, as well as implications for linguists and educators more broadly and suggestions for practical ways to improve the shape of the data we use.

**2. TERMINOLOGY AND BACKGROUND.** Before discussing prior literature which motivates this study and the study itself, we begin by describing the terminology we use in this paper, and the

choice to use a binary gender designation.

**2.1. TERMINOLOGY.** Like the authors of the papers cited above, we rely on the notion of perceived/conceptual gender, using the following definition from Ackerman (2019):

(1) **CONCEPTUAL GENDER:**

The gender that is expressed, inferred, and used by a perceiver to classify a referent (typically human, but can be extended to anthropomorphized non-humans).

Numerous studies have shown strong gender biases of certain noun phrases, for example *surgeon*, *CEO*, *nurse*, *cheerleader* (e.g. Garnham et al. 2002; Kennison & Trofe 2003; Duffy & Keir 2004; Kreiner et al. 2008; Gygax et al. 2008; Pyykkönen et al. 2010). Experimental participants tend to expect the referent of nouns such as *surgeon* and *CEO* to be male rather than any other gender, while nouns such as *nurse* and *cheerleader* are biased female. These trends are robust across many studies, but such inferences are of course defeasible, indicating that these biases are tied to conceptual gender rather than to grammatical gender, defined as follows:

(2) **GRAMMATICAL GENDER:**

Formal syntactic and/or semantic features that are morpho-syntactically defined.<sup>2</sup>

That is, grammatical gender is a formal syntactic property of a noun, is obligatory and cannot be overridden. On the other hand, conceptual gender is tied to the inferences speakers make, and may vary by societal norms, by context, and over time. For example, studies have shown that names — including those that are often considered to be ‘gender-neutral’ — change over time and may become associated stereotypically with either male or female gender (see Barry & Harper 1982, 1993; Van Fleet & Atwater 1997; Lieberman et al. 2000; Hahn & Bentley 2003, among others).

Since the papers we investigate (usually implicitly) assume that gender is a binary, we are forced into the same classification, as well. As we will expand on in more detail below, our classification of nouns into gender categories in example sentences relies extensively on the perception of individuals reading the examples. Consequently, any reference to ‘male’ arguments or ‘men’ is meant to identify those arguments whose conceptual or perceived gender is male, and

likewise for ‘female’ and ‘woman’. It is important to stress, however, that treating gender as if it is a binary classification is harmful to individuals who identify neither as male nor female; while there is diversity across individuals in the preferred term, such gender identities are often referred to collectively as nonbinary. We hope that one consequence of this work will be to move away from the binary classification, as well as from gender-normative roles, when there is no need for them in linguistic examples.

**2.2. MACAULAY AND BRICE 1997.** The study we present in section 3 of this paper is inspired by two prior studies which investigate the distribution of gender in example sentences in syntax textbooks. In what follows, we present a brief summary of these papers. We begin by describing the study in Macaulay & Brice 1997. M&B (1997) present two studies: the first is a comprehensive study of gender representation in the example sentences in a single textbook (also published in Macaulay & Brice 1994). The second is a comparative study designed to investigate whether the gender imbalance found in the first study generalizes across other textbooks. Here we concentrate on the latter study, whose design and results inspired both Pabst et al. 2018 and the study we present in section 3.

M&B (1997) present an investigation of 10 syntax textbooks published between 1969 and 1994. 7 textbooks had male authors, and 3 had female authors. 200 examples were randomly sampled from each textbook and manually coded for the following parameters:

- (3) *M&B's coding:*
- a. Grammatical function (subject, direct object, indirect object, etc.)
  - b. Thematic relations (agent, patient, experiencer, recipient, etc.)
  - c. Lexical choices (pronouns, proper names, violence, appearance, etc.)

Here we summarize some of the main findings. We refer the reader to M&B (1997) for a more comprehensive discussion and for examples illustrating each one of the findings. In short, M&B find that example sentences introduce male protagonists at higher rates than female ones, and that they perpetuate gender biases, as summarized in 4. When female-gendered arguments are

over-represented in example sentences, this, too, is done in a way that perpetuates stereotypes, as outlined in 5.

(4) *Male-gendered arguments in M&B ...*

- a. Appear more often as arguments
- b. Are more likely to be subjects and agents
- c. Are more often referred to using pronouns and proper names
- d. Engage more often in ‘intellectual activities’ such as book reading/handling, and feature more frequently in examples involving cars
- e. Have (a) more and (b) more varied occupations
- f. Are more often engaged in violent activities, especially as perpetrators

(5) *Female-gendered arguments in M&B ...*

- a. Are more often referred to with kinship terms (X’s wife, mother)
- b. Are more likely to have their appearances described

These findings are shown in the selected examples below, representing only a small portion of those shown by M&B to illustrate their observations:

(6) *Selected examples from M&B:*

- a. Harry watches the fights and his wife the soap operas.
- b. Bill is proud of his father and tired of his mother.
- c. Every painting of Maja and photograph of Debbie pleased Ben.
- d. The man is hitting the woman with a stick.
- e. The man who shot her believed there was someone else who was seeing Helen.

In addition, the syntax textbooks studied by M&B commonly used examples that contained explicit and suggestive language:

(7) *Explicit and suggestive language in example sentences from M&B:*

- a. Max doesn't beat his wife because he loves her.
- b. After Rambo as a lover, she was exhausted.
- c. She's fond of John naked.
- d. What a nice pear Mary's got!<sup>3</sup>
- e. John forced Mary to be kissed by Bill.

Finally, the gender of the textbook author played an important role: male authors were on average much more likely to choose biased examples, whereas female authors tended toward a more balanced sample.

M&B thus conclude that “Our results clearly illustrate the need for such scrutiny: females are simply not significant actors in the world constructed in most corpora of example sentences” (p. 816).

**2.3. PABST ET AL. 2018.** The majority of M&B's findings are replicated in the study of textbooks published between the years 2005–2017 presented in Pabst et al. 2018. Results for male-gendered arguments are summarized in 8. The only finding in M&B (4) that was not apparent in Pabst et al. 2018 (8) is that men often worked with cars.

(8) *Male-gendered arguments in Pabst et al. (2018) ...*

- a. Appear more often as arguments
- b. Are more likely to be subjects and agents
- c. Are more often referred to using pronouns and proper names
- d. Engage more often in ‘intellectual activities’ such as book reading/handling
- e. Have (a) more and (b) more varied occupations
- f. Are more often engaged in violent activities, especially as perpetrators

Some findings about female-gendered arguments are replicated as well, summarized in 9. Unlike M&B (1997), however, Pabst and colleagues (2018) no longer find that female-gendered

arguments have their appearance described more often than male ones — in fact, they find a very small number of examples that describe physical appearances in general. On the other hand, they find that female-gendered arguments exhibit a greater proportion of negative emotions in their sample, a finding that was not investigated in the original M&B study.

(9) *Female-gendered arguments in Pabst et al. (2018) ...*

- a. Are more often referred to with kinship terms (X's wife, mother)
- b. Exhibit more negative emotions than male-gendered arguments

In general, Pabst and colleagues (2018) find very little sexually explicit or suggestive language, but they find that the majority of stereotypes about both male- and female-gendered arguments is maintained in recent textbooks. Some examples are shown below, and the reader is referred to the original text for more data.

(10) *Selected examples from Pabst et al. (2018):*

- a. She snarled at the students who hadn't read the book.
- b. Bruce loved and Kelly hated phonology class.
- c. Joan believes he is a genius even more fervently than Bob's mother does.
- d. Mary entertained the men during each other's vacation.
- e. He drove her hard, he stole her fame or would have if he could have.
- f. Mohammed buys a house.
- g. The woman bought rice for the children.
- h. Slavko left his wife.
- i. Mary may wonder if John cheats on her.

Finally, Pabst and colleagues find some differences between male and female authors, such that female authors tended to use a higher proportion of non-gendered arguments than male authors. One female author used a more balanced proportion of gendered arguments, but the other two used similar proportions to the men. Overall there were no statistically significant differences between the proportions of male to female arguments based on the textbook authors' genders.

**2.4. SUMMARY.** To summarize, the vast majority of problems that afflicted example sentences in syntax textbooks 20 years ago are still present today. The main change has affected explicitly suggestive and stereotypical examples. While Pabst and colleagues (2018) find a small number of such examples, they are no longer as blatant or numerous. Instead, the discrepancies are made more difficult to detect, although they remain present: the skew now requires a broader lens to observe. Once this is done, however, Pabst and colleagues (2018) find a vast range of ways in which implicit gender biases are present in their sample, such that men are over-represented and presented more favorably, and women are suppressed. They additionally point out that non-binary identities are non-existent in their sample, and that there is a Western bias in choice of names and contexts.

**3. STUDY.** On the heels of these results, and following a recent study of French example sentences (Richy & Burnett 2019), in this study we examine all papers published in three leading theoretical linguistics journals: *Linguistic Inquiry*, *Natural Language & Linguistic Theory*, and *Language* between the years 1997–2018. To foreshadow our results, we systematically find similar findings to those reported in the above studies.

**3.1. METHODS AND DESIGN.** Unlike previous studies, instead of sampling, we automatically extracted all example sentences from all papers published during the time period under investigation in all three journals, using regular expressions.<sup>4</sup> This was made possible by the fact that constructed example sentences follow a standard formatting in linguistic work: examples are marked by a number enclosed in parentheses, and they are typically removed from the margin, perhaps also including a title, a gloss and a translation.<sup>5</sup>

Twenty-four Yale University undergraduate students were recruited to clean up the examples and provide coding following guidelines which we will detail immediately below. These students had all taken at least one linguistics course, and several were (or subsequently became) linguistics majors. The coding process was as follows: As a first step, coders were instructed to remove all data points which were not example sentences, but rather were mistakenly captured or were generalizations, rules, descriptions of data, or any other text that appeared in numbered examples.

Next, examples that did not contain third person arguments were removed. Next, examples with third person arguments that did not refer to humans or anthropomorphized individuals were removed from the study.<sup>6</sup> Finally, for each third person argument in each of the remaining example sentences, our coders provided information along the same lines as the studies surveyed above:

(11) *Argument coding in this study:*

- a. Grammatical function (subject, direct object, indirect object)
- b. Thematic relations (agent, patient, experiencer, recipient)
- c. Lexical choices (pronouns, proper names, violence, physical appearance, etc.)

We decided to focus on major syntactic and semantic roles, as spelled out in 11a–b. Unlike prior work, we did not separately code minor thematic relations such as beneficiary or instrument, anticipating the data to be too sparse for analysis, as was the case in those prior studies, and since our coders had minimal prior exposure to the notion of thematic relations. Such roles were grouped under ‘other’, and eventually excluded from our analysis.

Coders were instructed to flag any example that they considered stereotypical in any way. In addition, they were given a list of specific properties to look out for. Following prior work, we were interested in violence, appearances, emotions, romance, “intellectual activities” such as writing/reading, sexually suggestive/explicit language, and cars. In addition, data for kinship terms was extracted automatically using regular expressions, as the set of such terms is sufficiently limited to easily allow such a search.

In the case of non-English examples, our coders were instructed to rely on the English translation, and in addition to inspect the glosses to ensure that the translation was a good match.<sup>7</sup> We were specifically concerned with possible cases of mismatches between gloss and translation, such as a non-gendered third-person pronoun being translated into English as “he.” Further, as noted above, in cases of gendered languages, we did not include in our data any inanimate objects, even if they bore masculine or feminine agreement. Only human arguments are included in our analysis.

Coders received hands-on training in conducting the coding described here. Their work was inspected on an ongoing basis by one of the paper’s authors, and feedback was provided as needed

to improve their work. We additionally conducted some post-hoc tests to verify the accuracy of the coding, such as the use of regular expressions to find examples with relevant traits and confirm that they were correctly flagged by the coders, and spot-checking of random individual examples.

Throughout the coders' work, we emphasized speed as a guiding principle. We introduced a notation for decisions which the coders were not sure of. They were encouraged to flag and skip such cases, and not spend any time investigating further through googling or additional reading. This became particularly useful in the case of unfamiliar names, and is in part the reason for our choice to concentrate only on major thematic relations. In all cases where the 'unsure' notation was used, one of the paper's authors followed up to add the missing details.

Finally, we note that each example contained some metadata, such as the year of publication, journal name, and example number. However, data concerning the journal papers' authors was not shown to the coders, and they were not asked to identify whether the authors were male or female. This was done for two reasons: first, to avoid any bias that might stem from knowing this fact during the coding process, and second because we expected our coders to be unfamiliar with many of the authors. This information was added by the paper's authors after the coding process was completed. We provide additional details about the analysis of authors' gender in section 3.4.3.

**3.2. RESULTS.** In total, we aggregated all papers from the three journals under investigation published over the 20 years we were concerned with. Of those, 1,126 articles had example sentences, for a total of 30,591 examples. Overall, 806 papers contained at least one gendered human argument, resulting in a total of 25,085 third person human arguments for our analysis. We first show results for each of the properties listed in 11, and in the next section discuss meta-analyses (a) over time, (b) by language of the example, and (c) by the gender of the authors. All results shown here are statistically significant, as confirmed by Pearson's  $\chi^2$  tests with Yates' continuity correction, unless otherwise noted.

**3.2.1. DISTRIBUTION OF ARGUMENTS.** We begin by considering the overall distribution of arguments in our sample. Of the 25,085 third-person human arguments identified in our study, 7,409 were non-gendered or ambiguous (A) arguments.<sup>8</sup> 5,670 arguments were coded as female,

while 12,006 were coded as male. Therefore, female arguments make up 22% of all third-person arguments in our sample, ambiguous arguments make up 30%, and male arguments make up the remaining 48% of arguments.

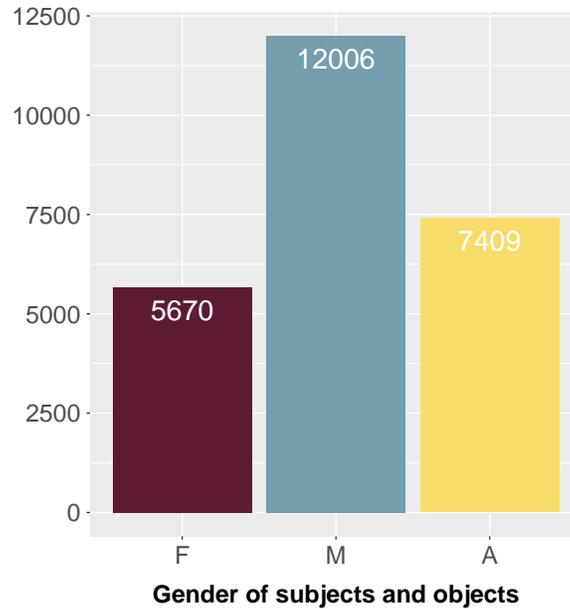


Figure 1: Gender distribution of arguments in all example sentences in this study

In the remainder of this paper we concentrate on arguments that were perceived as either male or female. Of the 17,676 male/female gendered arguments we coded, 32% were female and 68% were male, for a ratio of 2.1 male arguments for every 1 female argument. These ratios are consistent across the three journals that we studied, as shown in the next figure:

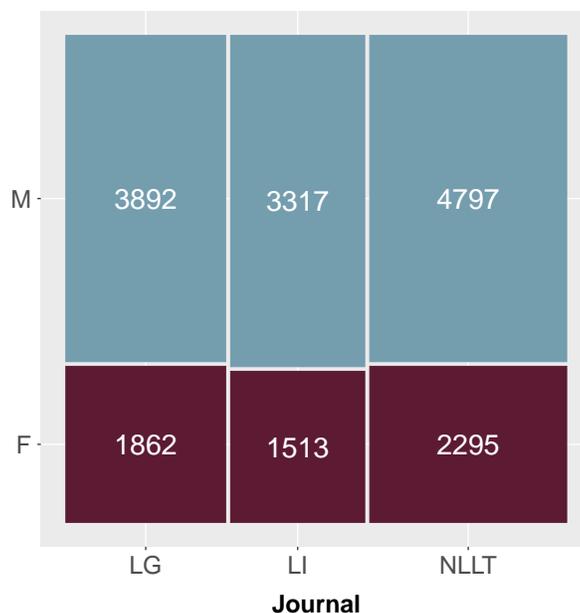


Figure 2: Gender distribution of arguments in all example sentences by journal. Specifically, we find that the ratio of female-gendered arguments in the data to the total number of arguments is 32% for papers published in *Natural Language & Linguistic Theory*, 31% for paper published in *Linguistic Inquiry*, and 32% for papers published in *Language*. These gender ratios were not found to be significantly different across journals; therefore, we show combined results for all three journals in the sections that follow.

**3.2.2. GRAMMATICAL FUNCTION AND THEMATIC RELATIONS.** Next, we consider the distribution of arguments with respect to syntactic and semantic roles. We find that 83% of male arguments are subjects (8,976 of 10,757), while only 78% of female arguments are subjects (3,769 of 4,814). That is, female-gendered arguments are less likely to occur as subjects and more likely to occur in non-subject roles (namely, direct and indirect objects) as compared to male-gendered arguments. Similarly, we observe a skew in thematic relations: Female arguments represent 30% of agents (2,754 of 9,090), 31% of experiencers (1,160 of 3,732), 35% of patients (1,159 of 3,354), and 43% of recipients (339 of 797). Since female arguments comprise 32% of the sample overall, this means that they are over-represented among patients and recipients. These findings are summarized in Figure 3.

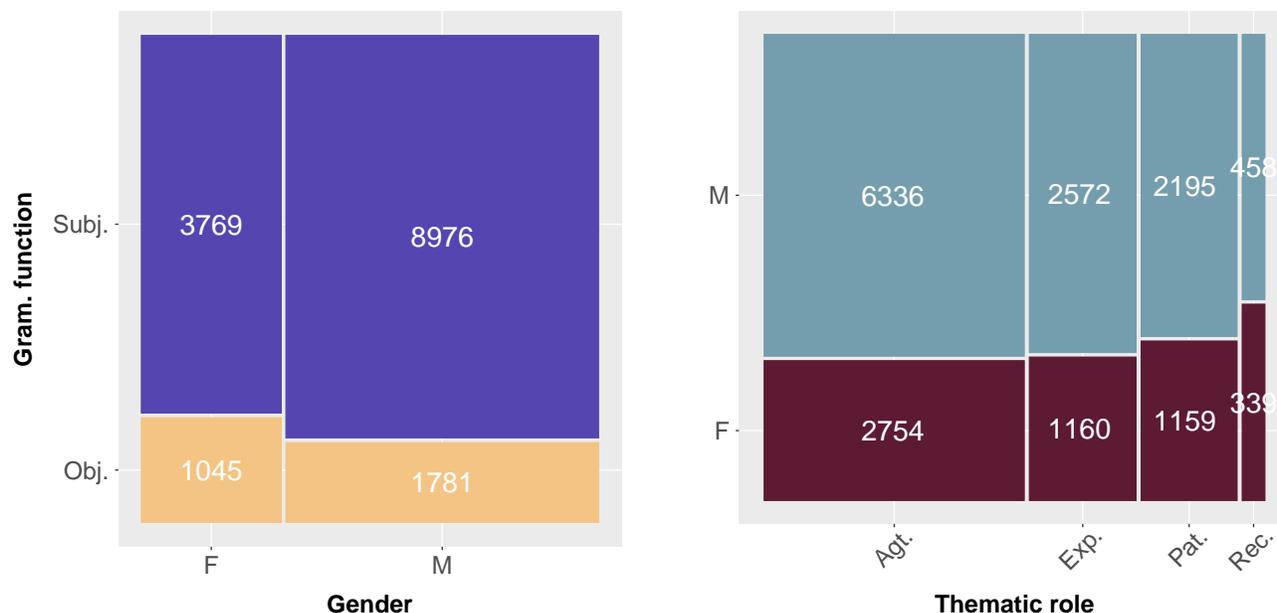


Figure 3: Distribution of arguments by grammatical function (left) and thematic relation (right)

That is, women are described often as passive observers in a male-dominated world. They are not initiators of actions but rather more likely on the receiving end. They lack independent agency.

**3.2.3. PRONOUNS AND PROPER NAMES.** We examine next the distribution of pronouns and proper names in our sample, to explore whether some of the skew in our sample could be specifically attributed to these categories. In short, we find a small skew with respect to pronouns and no skew where proper names are concerned. We conclude that the skews we observe are more fundamental and not attributable to one class of examples alone.

Turning first to pronouns, we find that male-gendered pronouns are significantly more common than female-gendered ones. Male-gendered pronouns make up 29% of all male-gendered arguments, while female-gendered pronouns make up only 23% of female arguments. We tentatively suggest that this may be a side-effect of the overall prevalence of male-gendered arguments, especially subjects, in the data. If a pronoun were to refer back to an argument, that argument is more likely to be male than female.<sup>9</sup>

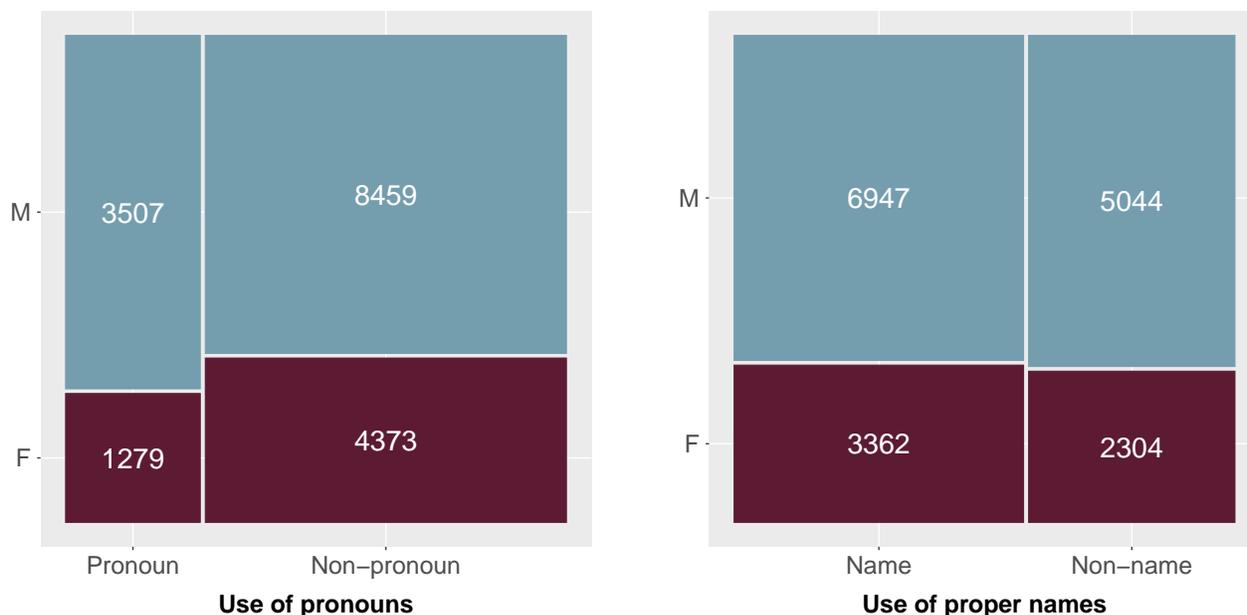


Figure 4: Pronouns (left) and proper names (right) in the data

Concentrating more closely on names, we observed a total of 10,737 names in our study. Names were coded for gender one of two ways. If a name was referred to using a gendered pronoun, e.g. “Taylor looked at a picture of herself”, it was gendered accordingly. Names without referential pronouns were coded for gender based on coders’ judgments of that name’s stereotypically perceived gender. Of the personal names identified in the data, 428 were identified as referring to non-gendered or ambiguously-gendered arguments. The remaining 10,309 names were coded as either male or female, as represented in Figure 4. Female-gendered arguments comprised 33% of names (3,362 of 10,309) and 31% of non-name arguments (2,304 of 7,348), similar ratios to the 32% of female-gendered arguments in the dataset as a whole. Likewise, the proportion of gendered arguments which were proper names was similar for male and female arguments, at 58% and 59%, respectively, a difference that was not found to be significant.

To achieve maximal accuracy in this section, the paper’s authors manually examined all names not familiar to our undergraduate student coders. We did not use an automatic gender classifier because we did not trust the ones currently available: these classifiers tend to have a bias toward Western names, and they don’t typically do well with ambiguous names. They also won’t be able to deal with the co-occurrence of a pronoun in the sentence which might disambiguate the gender

of a name, nor with information available in the gloss or translation of a non-English example.

We coded a name as male or female if (a) we were familiar with the name and language, (b) the gloss or translation explicitly stated that the name was gendered, (c) the gloss indicated gendered agreement, (d) there was a gendered pronoun corresponding to the name in the example, (e) an online search indicated that the name was clearly used only for one gender, or (f) a community of linguist experts who we consulted indicated their familiarity with the language and name.<sup>10</sup> The top five names in each gender category we coded is presented here:

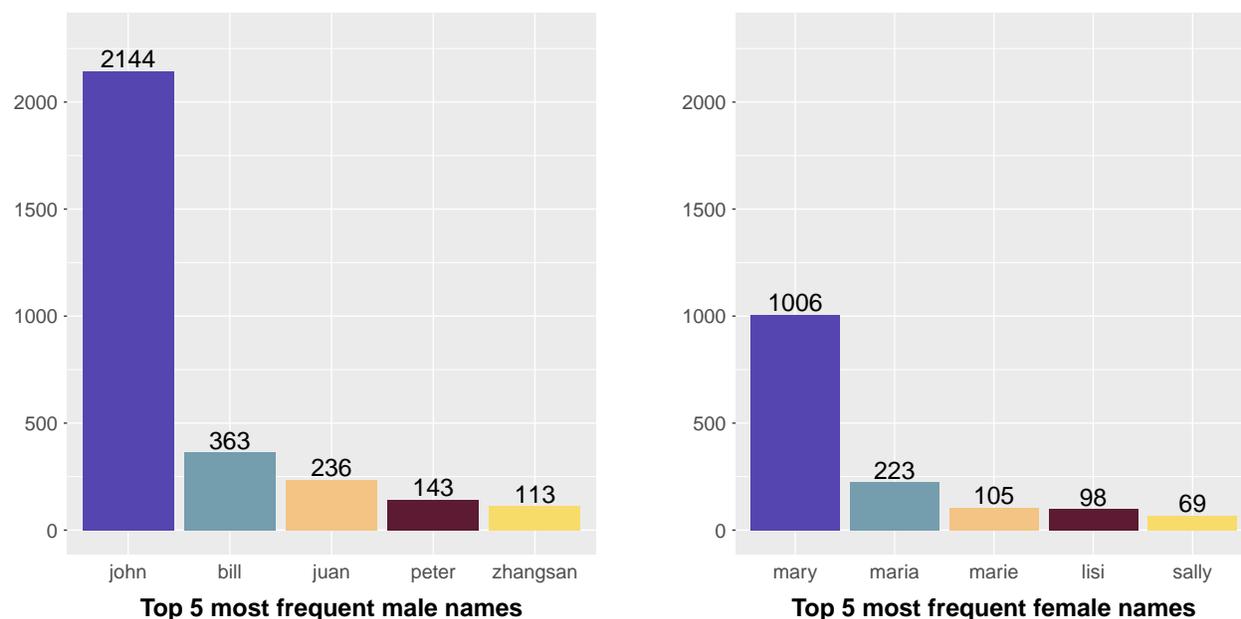


Figure 5: Top five most common male (left) and female (right) names

Among the top names, we find that 31% of all male names are *John*. Concomitantly, 30% of all female names are *Mary*. In fact, 2 of the top 5 male names are “John” variants: *John* and *Juan*, whereas 3 of the top 5 female names are “Mary” variants: *Mary*, *Maria*, and *Marie*. That is, we observe a strong Western/Christian bias in authors’ choices, with little variability. This applies to name choices regardless of gender.

**3.2.4. LEXICAL CHOICES.** We also consider various lexical choices about the example sentences in our study. We start with examples that refer to an argument’s occupation. We find that male-gendered arguments are over-represented in such examples. While male arguments

outnumber female arguments 2.1:1 overall, the ratio is close to 3:1 among arguments described as having professions (i.e. 73% are male).

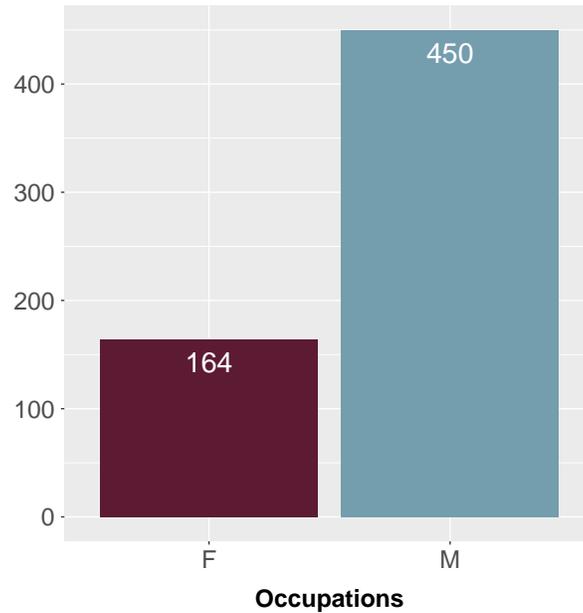


Figure 6: Arguments whose occupation is discussed in the example

Next we examine examples which involve violent events of some kind. Observing first the overall distribution of arguments, we find that male-gendered arguments are massively over-represented in such examples: 83% of the arguments in these examples are male (462 of 559, a ratio higher than 4:1). Within each gender category, however, we find a similar proportion of subjects and non-subjects in sentences describing violence. Subjects comprise 72% of all male-gendered arguments (331 of 462) and 69% of all female-gendered arguments (67 of 97).

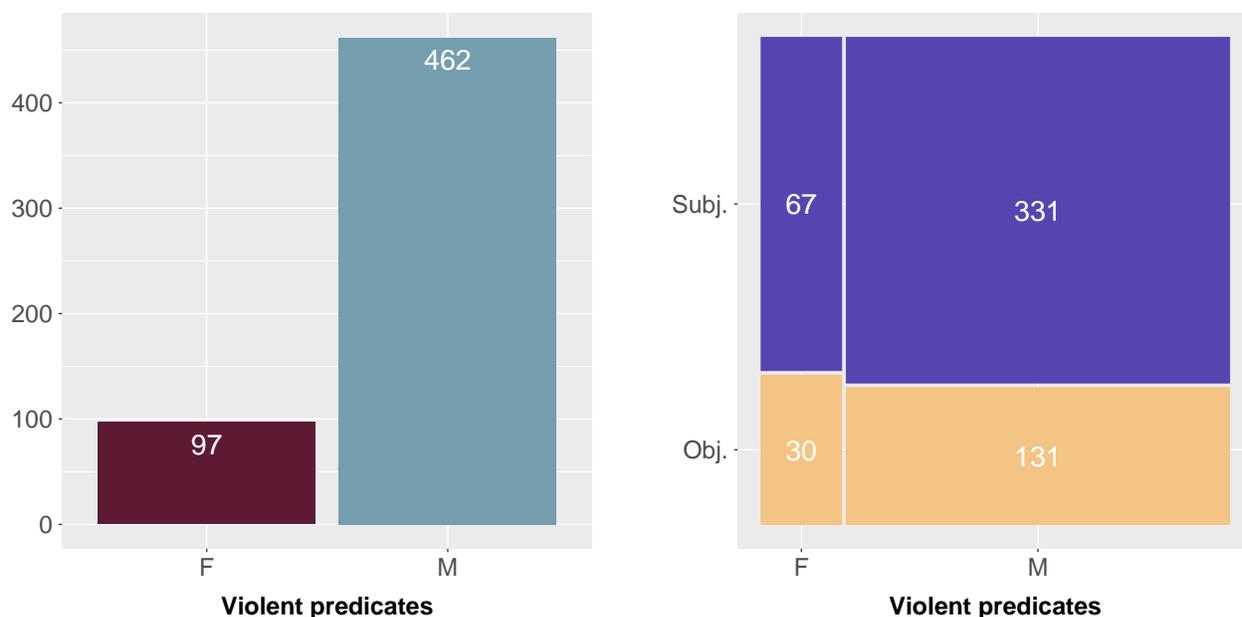


Figure 7: Violent events by gender (left) and by gender×syntactic position (right)

Turning to examples involving the expression of romantic or sexually suggestive content, we now find that the gender distribution in examples is remarkably different. Here, female-gendered arguments are over-represented, comprising 50% of all arguments (201 of 403). Taking into account that male-gendered arguments are generally over-represented at a 2.1:1 ratio compared to female-gendered arguments, this constitutes a remarkable skew. Importantly, this difference interacts with grammatical function. The over-representation of female arguments in these sentences is also seen in their greater over-representation as non-subjects. Only 59% of female-gendered arguments are subjects (118 of 201), whereas 76% of male-gendered arguments are subjects (153 of 202). To state it another way, 44% of subjects (118 of 271) and 63% of non-subjects (83 of 132) are female. That is, female arguments occur more frequently in sentences with romantic/sexual content than in the dataset as a whole, and also appear more often in those sentences as non-subjects.

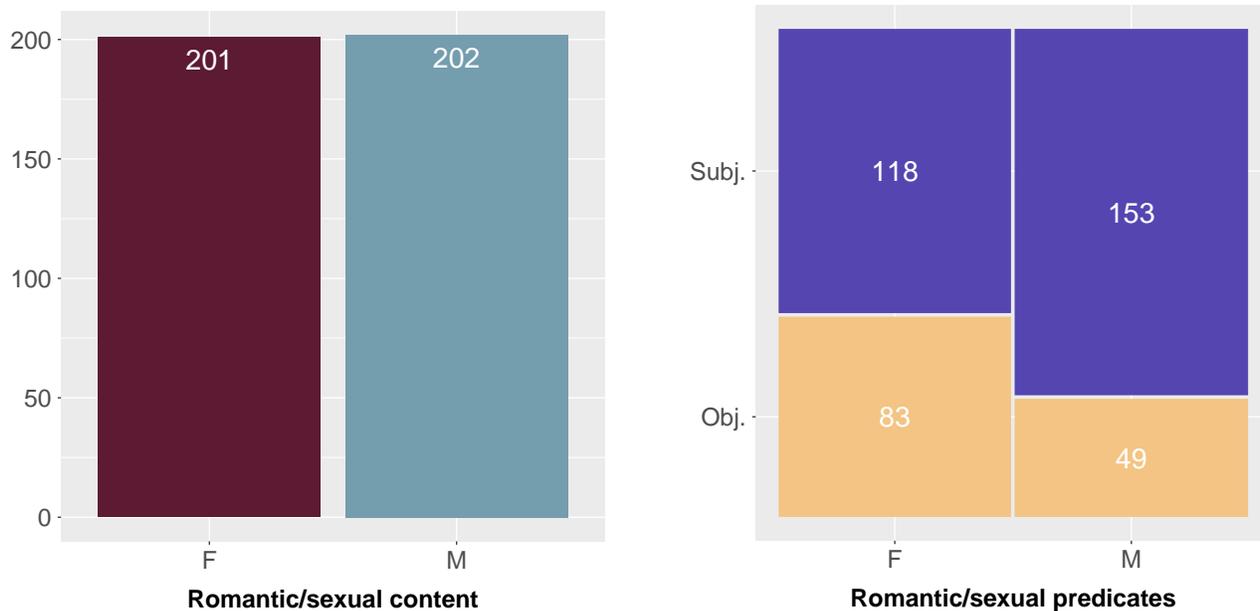


Figure 8: “Romantic” events by gender (left) and by gender×grammatical function (right)  
 Finally, female-gendered arguments are massively over-represented among those referred to by kinship terms: 57% of all such arguments are female-gendered (410 of 720). Considering again the overall 2.1:1 male skew in the data, this is a particularly striking finding.

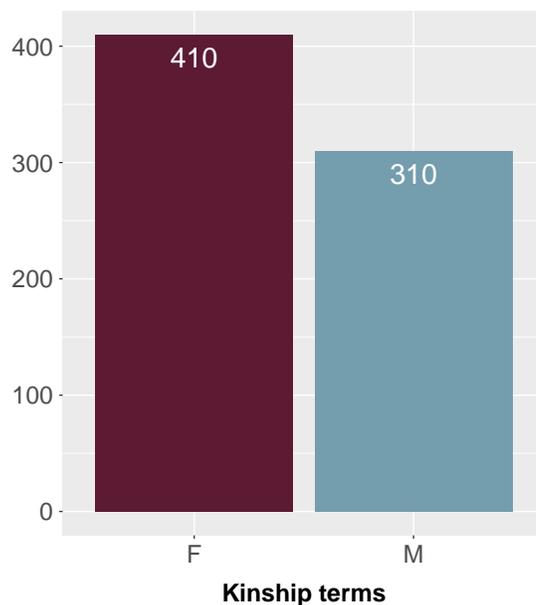


Figure 9: Use of kinship terms by gender

We also coded for example sentences relating to books and other intellectual activities. There

were 1,003 such examples, and these were associated with 330 female-gendered arguments and 673 male-gendered ones. As this distribution does not differ from the overall gender ratio, we do not comment further on these results.<sup>11</sup>

**3.2.5. SENTIMENT ANALYSIS.** Finally, we use the R function ‘get\_sentiments’ in the ‘tidytext’ package (Silge & Robinson 2016) to run sentiment analysis on the data. We find that male arguments are over-represented in sentences conveying negative emotions such as *anger* and *fear*, while female arguments are over-represented in sentences conveying positive emotions such as *joy* and *trust*.<sup>12</sup>

Predicates were counted once for each gender-coded argument associated with them, i.e. a predicate with gendered subject and object arguments is counted twice. Sentiments were determined using both ‘Bing’ (Liu 2012) and ‘NRC’ (Mohammad & Turney 2013) methods of categorization, with corresponding lexicons. The Bing method bins predicates into binary *positive* and *negative* sentiment categories, while the NRC method categorizes predicates into ten distinct groups: *anger*, *anticipation*, *disgust*, *fear*, *joy*, *sadness*, *surprise*, *trust*, and *negative* and *positive* categories that function as an ‘elsewhere’ case, as each predicate is only assigned to one of these ten categories. If a predicate is not included in the lexicons of the method in use, the ‘get\_sentiments’ function excludes it. Specifically, out of the 17,676 predicates in the data, only 2,171 predicates were categorized by the Bing method, and 10,323 predicates were categorized by the NRC method. The rest of the predicates are excluded from the analysis in this section.

Results from the Bing categorization of sentiments reveal a slight skew in the genders of arguments in positive versus negative sentences. While the overall male-to-female ratio of arguments is 2.1:1, the gender ratio for negative sentiments is 2.3:1, indicating a skew toward male arguments in these sentences. Conversely, female arguments are slightly over-represented in positive sentences; the gender ratio for positive sentiments is 1.7:1.

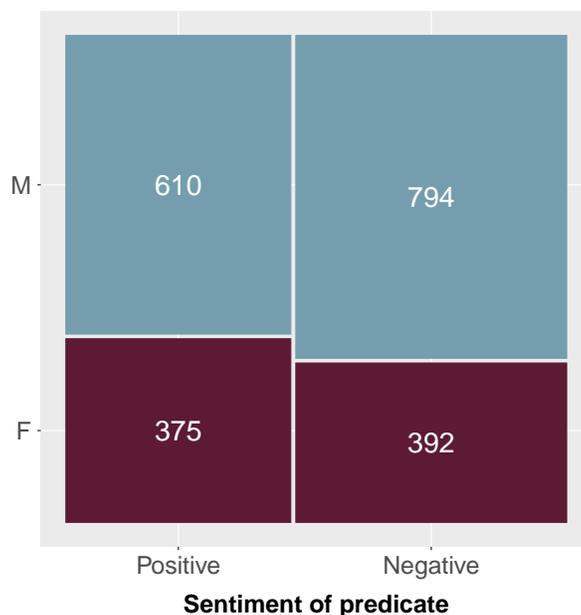


Figure 10: Positive/negative sentiments categorized by Bing method, by gender

This trend continues to be borne out in the results of the NRC method of analysis. The figure below shows the male-to-female argument ratios for each of the 10 sentiments identified with this method, with the black line showing the overall gender ratio of 2.1:1 for reference. The negative sentiments — *fear*, *anger*, *negative*, *sadness*, and *disgust* — have higher ratios (to the right of the black line), indicating an over-representation of male arguments in these sentences. The ratio is as high as 3 male arguments for every female argument, for *fear* and *anger*. The clearly positive sentiments — *positive*, *trust*, and *joy* — show the opposite trend, with lower-than-average ratios that indicate an over-representation of female arguments. This also holds for *anticipation*, which is not clearly positive or negative, and *surprise* is about at the average ratio given the general 2.1:1 distribution of male-to-female arguments..

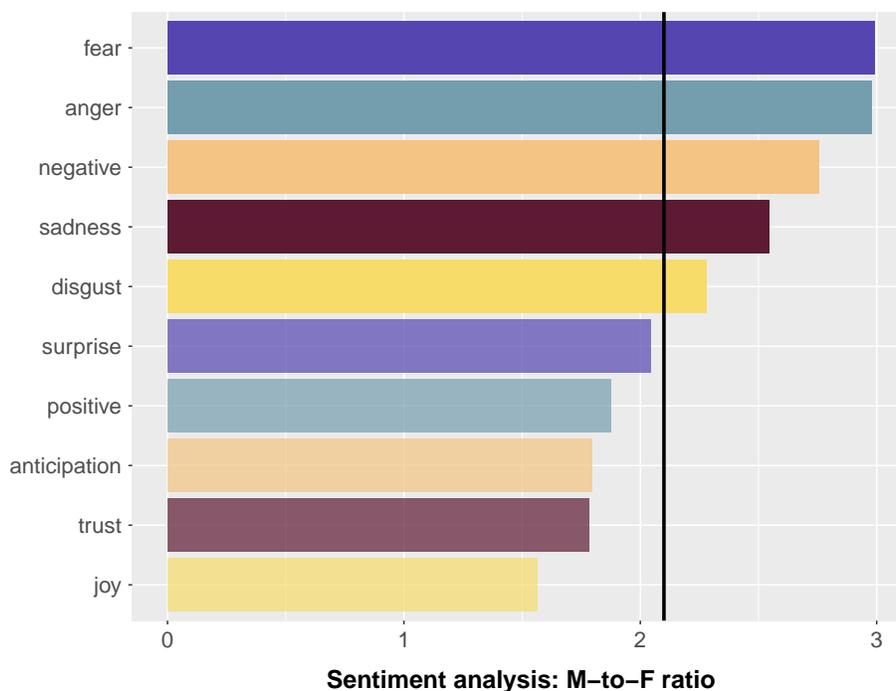


Figure 11: Male-to-female argument ratios by sentiment, categorized with NRC method; black line indicates overall M-to-F ratio of 2.1:1

These results are in keeping with previous results in our study, showing an over-representation of male arguments in violent sentences, and over-representation of female arguments in romantic/sexual sentences. Female arguments are more often used in sentences concerning positive emotions, while male arguments are more often used in sentences conveying negative, and especially violent and angry, sentiments.

**3.3. SOME ILLUSTRATIVE EXAMPLES.** Like in Pabst et al. 2018 and unlike in the earlier Macaulay & Brice 1997 study, we do not find many sexually explicit or suggestive language. Nonetheless, we find that stereotypes of both genders are commonly used in linguistic example sentences. In the following example we provide a sample of such sentences.

Most of the examples below reflect more than one type of bias or stereotype. For example, examples 12a–c show stereotypical choices involving the object of the verb ‘wash’, as well as stereotypical gender roles. Examples 12d–e describe a woman’s mental state as unstable. Examples 12f–j suggest that women are not as intellectually capable as men: men are Nobel Prize

winners, professors, students, geniuses, and — remarkably — linguists. Examples 12j–l illustrate stereotypical uses of kinship terms. Examples 12l–n exemplify various references to violence. Finally, the remaining examples illustrate other stereotypical lexical choices. These examples also attempt to roughly reproduce the overall 2:1 skew of male to female subjects in our sample.

(12) *Some stereotypical examples found in our study:*

- 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 (not Peter) washed cars well
- d. John told Bill that Mary began to cry without any reason
- e. \* Kelly broke again tonight when she did the dishes
- f. Which Nobel prize winning author came in his car?
- g. The students are all the boys
- h. At least one student of every professor<sub>i</sub> is horrified at his<sub>i</sub> grading procedure
- i. No linguist<sub>1</sub> here recommended some of his<sub>1</sub> own books, but I don't know which of his<sub>1</sub> own books
- j. Ray<sub>1</sub> mother thinks he<sub>1</sub> a genius
- k. Aoyama's sister-in-law knitted a scarf
- l. An Iraqi father drowned his 17 year old daughter
- m. Rabe forced women to wash clothes
- n. to leave the maiden . . . unmolested
- o. Married him, didn't she/\*Marge/%the gold digger?
- p. Bill won't go to the bar and James to the liquor store
- q. Mary thought that it pleased John [PRO to speak his/\*her mind]
- r.
  - i. ?? John seems considered a fool
  - ii. Also, Anne Elliot seems considered a spinster by everyone, including herself. . .

Here it is important to note that we are not cherry-picking examples. These are representative of the examples flagged by our undergraduate coders. We furthermore note that the stereotypes we observe lean in both directions: all individuals are cast into stereotypical roles.

**3.4. META-ANALYSIS.** Finally, we comment on three additional aspects of the data we have collected.

**3.4.1. THE LANGUAGE OF THE EXAMPLES.** First, we note that whether the example was in English or not did not affect the results we observe. We find a total of 33% of female-gendered arguments in English examples, and 31% female-gendered arguments in non-English examples.

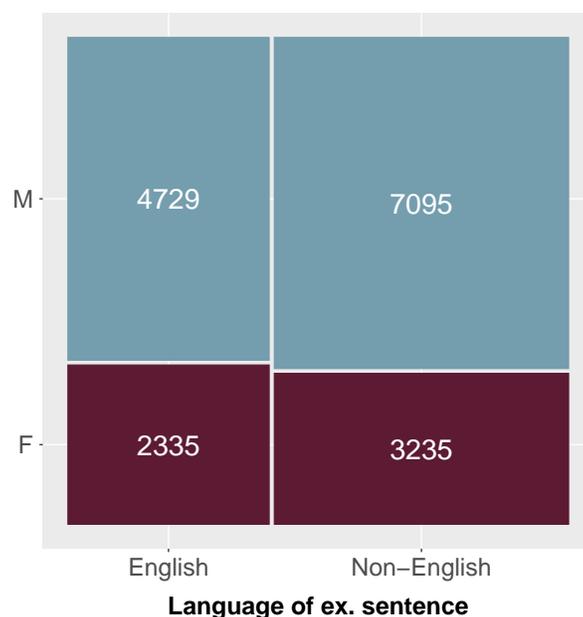


Figure 12: Gender distribution of arguments in all example sentences in this study. We believe that this, again, is suggestive of a broad issue in our field. The bias is not introduced by non-English examples, where some constraints on data collection may affect the sentences tested by field linguists in various ways. Instead, English and non-English examples show a similar bias, indicating that access to data or speakers, or familiarity with the language, do not significantly affect the gender distribution of arguments.

**3.4.2. GENDERED ARGUMENTS OVER TIME.** Next, considering the distribution of male vs. female arguments in example sentences over time, we notice a positive trend over the past 20 years. However, at no point — in no year over the 20 years of this study — were the genders at parity. On average, we have moved 5% closer to parity over the course of two decades. At this pace, all things being equal, we would not expect to reach equality in the use of male and female arguments for at least 50 more years.

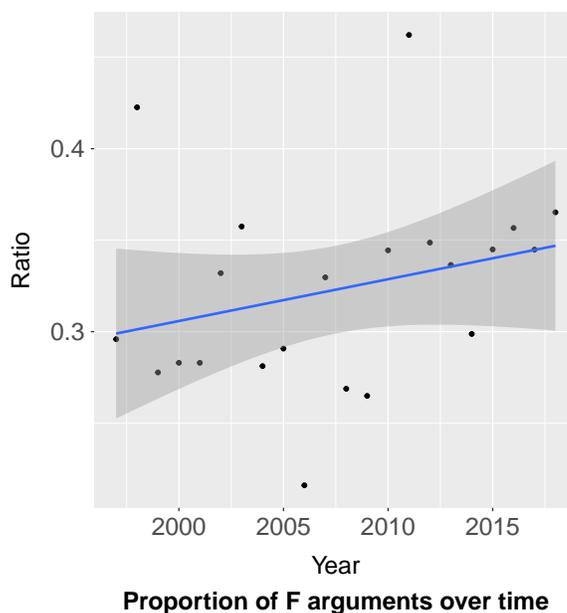


Figure 13: Gender distribution of arguments over time

Moreover, when comparing the distribution of arguments by syntactic position over time, we observe that the increase in Figure 13 may be attributed to an increase in the proportion of female objects in example sentences over time. The proportion of female subjects has remained stable and on average below 40% in the entire time period we study.

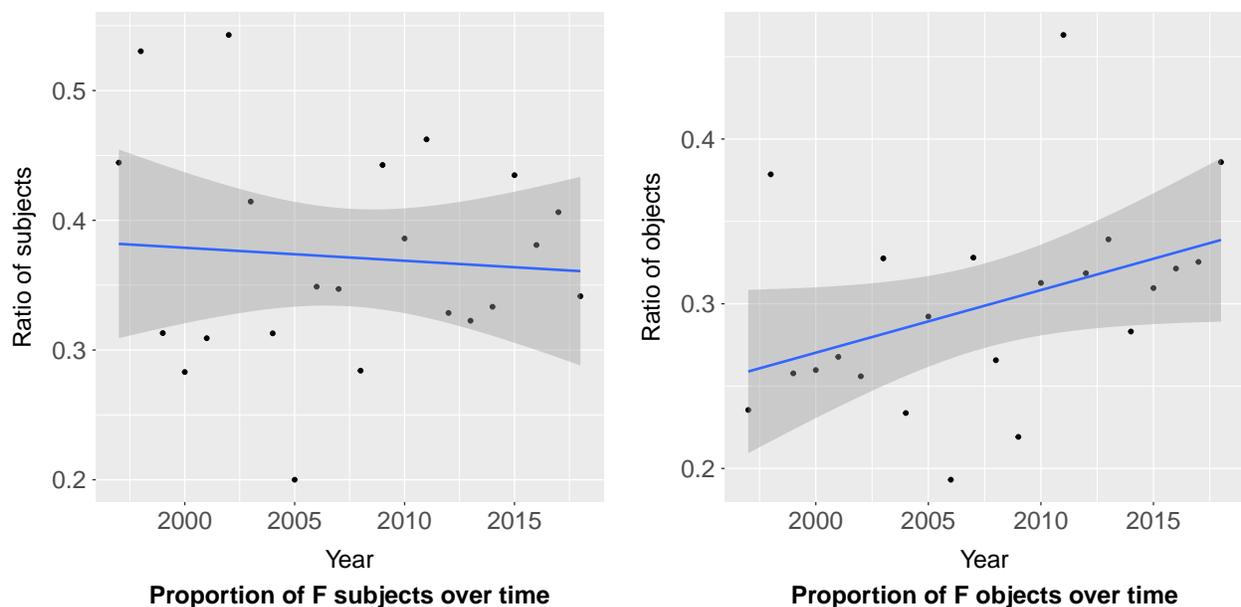


Figure 14: Gender distribution of subjects (left) and objects (right) over time

**3.4.3. THE GENDER OF AUTHORS.** Finally, we ask whether the gender of the authors of the papers we examined had an effect on the choices these authors make in their examples. Specifically, we consider whether the gender of a paper's author(s) has an effect on (a) the use of male versus female arguments, or (b) the use of gendered versus non-gendered arguments such as *the student*, *the children*, etc. We manually classified all the authors of the papers we examined into the binary conceptual gender categories of male and female based on their names at the time of publication, and additionally retained information about whether each author was the first author of the paper or not.<sup>13</sup>

In the three journals considered in this study, the proportion of female-authored papers has risen over time but has never exceeded 42% of all papers published. In fact, the graph in the next Figure suggests that the upward trend in the proportion of female authors levels out around 2005 and remains stable for the next 12 years we examined.

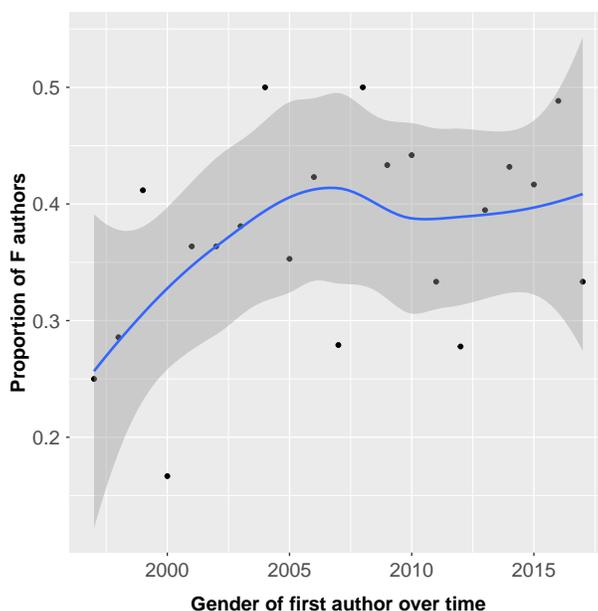


Figure 15: Distribution of argument gender by gender of author

To analyze the effect that author gender has on the types of arguments used, we constructed two logistic mixed effects models in R (R Core Team 2013) using the ‘lme4’ package (Bates et al. 2015). The first model concerns the use of male versus female gendered arguments. The dependent variable in this model was a binary factor indicating gender as ‘male’ or ‘female’; author gender<sup>14</sup> was included as a fixed effect, with a random effect of individual author to control for individual variation. The second model includes the same fixed and random effects, but the dependent variable consists of arguments categorized as ‘gendered’ (male or female) or ‘non-gendered’/ambiguous.

Figure 15 shows the distribution of male, female, and non-gendered arguments by gender of author. Male-authored papers have a higher total number of arguments than female-authored ones because of the overall disparity in publishing rates, and not because of any systematic gender differences in use of example sentences. Both male and female authors over-represent male arguments in their example sentences relative to female ones. However, while female authors write example sentences with female arguments in 35.5% of gendered arguments, male authors only include female arguments 31.7% of the time. This represents a statistically significant ( $z = |2.39|, p < 0.05$ ) difference in gendered argument ratios depending on the author’s gender.

We also consider whether an author's gender is a good predictor of the use of NON-GENDERED arguments, such as non-gendered common nouns and plurals.<sup>15</sup> Female authors use non-gendered arguments in 20.8% of their example sentences, while male authors use them 23.6% of the time. This difference was not found to be significant ( $z = |1.03|, p > 0.1$ ).

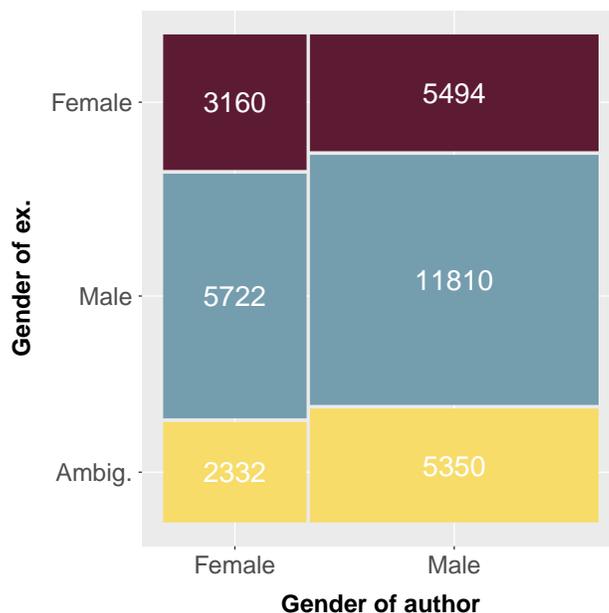


Figure 16: Distribution of argument gender, by gender of first author

**3.5. SUMMARY.** To summarize our main findings, male-gendered arguments are over-represented in the sample overall and in particular as subjects and agents. In addition, both men and women occur in many examples stereotypical of their gender. These results are consistent across journals and time, and replicate findings for textbooks in prior work. Our findings are summarized in 13–15.

(13) *Main findings: male-gendered arguments*

- a. Appear twice often as ARGUMENTS in total
- b. Appear more often as SUBJECTS AND AGENTS
- c. Engage in significantly more VIOLENCE
- d. Have significantly more OCCUPATIONS

e. Tend to exhibit more NEGATIVE EMOTIONS

(14) *Main findings: female-gendered arguments*

- a. Are over-represented as NON-SUBJECTS, especially as RECIPIENTS
- b. Are over-represented in sentences involving ROMANTIC/SEXUAL LANGUAGE
- c. Are massively over-referred to using KINSHIP terms
- d. Tend to exhibit more POSITIVE EMOTIONS

(15) *Main findings:s other trends in the data*

- a. No effect of language of example (English vs. non-English) on the results
- b. A small trend of improvement in gender ratios: from low-30% to mid-30%, attributable to an increase in the proportion of female objects over time
- c. Few or no overtly suggestive or explicit examples
- d. Persistent gendered stereotypes very much evident

**4. DISCUSSION AND CONCLUSION.** Example sentences are one of the main sources of data in theoretical linguistics. Some examples become enshrined as ‘canonical’, often divorced from their original sources. As scientists we are trained to regard such data as an impartial, empirical source of evidence in support of our arguments, whose ultimate goal is to further our understanding of the language faculty. However, we often ignore the social aspects that these examples occur in and that they exemplify. We have demonstrated here that constructed example sentences used in the linguistic literature may encode implicit biases (even at a very subtle level). These then get handed down to new generations of linguists, perpetuating a cycle.

This paper provided a comprehensive study of the distribution of conceptual gender in constructed example sentences. We chose three leading journals in theoretical linguistics: *Language*, *Linguistic Inquiry*, and *Natural Language & Linguistic Theory*, and investigated all papers published in these journals between the years 1997–2018. As the results in section 3 show at length, we find that gender bias permeates many aspects of these examples: from the fact that male-gendered arguments are over represented at a 2.1:1 ratio compared to female arguments, to a

multitude of gender stereotypes affecting both genders, to the fact that this trend is consistent and unchanged over time, across the three journals investigated, and by the language of the example.

As we noted at the outset of this paper, we take the makeup of example sentences — the arguments they use and the predicates in them — to be signals to students and researchers alike about what we take the world to be like: who is a free-thinking agent; a genius; who is likely a professor or a student; and who is a recipient of others' actions or belongings; the object of their affections; a caregiver; a spouse. While not blatant like in previous decades, the bias is nonetheless extensive and pervasive. It sends a powerful message about who is welcome in our field and who is less so. This, in turn, affects the kinds of research questions that are welcome, and the kinds of answers that we expect and ultimately adopt. It should be in everyone's interest to increase the inclusivity of our field. The more diverse we are as a field, the richer our ideas, research topics, approaches, types of data collected, and solutions.

The gender biases we observed in our study do not occur in isolation: rather, they reflect biases that are prevalent across our society at large, and in the field of linguistics more broadly. For example, studies have found similar biases in English language textbooks for various dialects of English (e.g. Bergvall 1996a; Lee & Collins 2010; Lewandowski 2014; Tarrayo 2014), and recently in example sentences in French journals (Richy & Burnett 2019). The issue has garnered some attention in recent LSA presentations, panels, and workshops, as cited in the introduction. It has featured in discussions of the representation of gender more broadly, including in reference to language acquisition (e.g. Eckert & McConnell-Ginet 2013; McConnell-Ginet 2014; Leslie et al. 2015; Meyer et al. 2015; Bian et al. 2017, among others).

This trend may also be seen as a part of a broader trend sometimes referred to as the 'leaky pipeline', whereby fewer and fewer non-male individuals are represented in an academic field the higher the rank of the individual (see e.g., Valian 1998, 2005; Goulden et al. 2011, among many others). Recently, some work has shown that this trend can be observed in linguistics, too, as a part of a general trend of less favorable conditions for non-male individuals, for example in the LSA's *The State of Linguistics in Higher Education Annual Report* from 2019<sup>16</sup> and the University of Michigan's *Survey of linguists and language researchers*.<sup>17</sup> We join the many authors who

study these issues in the belief that the field of linguistics would benefit from more inclusive citation, as well as hiring and promotion practices.

Nonetheless, as scientists, we believe that we **COULD** and **SHOULD** strive to avoid perpetuating bias, even if implicit, in our work. We argue here that better constructed example sentences, using inclusive language, can send an important message to the field: inclusive language encourages participation from underrepresented groups, leading to a better community and therefore to better science, at the cost of just a little more thoughtfulness.

In the remainder of this section we list out several concrete actions that we, as researchers and teachers, can take to improve on the current state of the field. We additionally refer the reader to the section “Actions people can take” on the University of Michigan website (see footnote 17).

First and foremost, as an author and teacher, pay attention to the ratio of gendered arguments in your example sentences, and to the distribution of grammatical functions and thematic relations of those arguments. Strive for equity in the examples that you construct or choose to cite. Consider in general the nature of the world that is portrayed through your example sentences and any stereotypes or misconceptions they may inadvertently fall into. Consider inclusivity beyond binary genders and beyond hetero-normative gender roles. Your examples could additionally serve to send a message about the diversity of races and ethnicities.

Avoid the use of gendered lexical items such as *-man* and *he* where not necessary. Adopt and encourage instead the use of singular *they*, as a more inclusive pronoun when referring to (singular) nouns whose gender is unknown. Consider using singular *they* even when the argument’s gender is known, but is irrelevant to the example (see e.g. Bjorkman 2017; Ackerman 2019 Bradley et al. 2019). Use inclusive nouns such as *Congressperson* and *humankind*. Keep the LSA’s *Guidelines for Inclusive Language* in mind when writing, as well as when reviewing and editing papers.<sup>18</sup>

When using names in example sentences, consider using diverse names, paying explicit attention to the distribution of gender in your examples. We do not advise here the use of “gender-neutral” names, given the existing research that shows that such names are often not truly perceived as neutral, and furthermore that the perceived gender of a given name may change over

time (see Barry & Harper 1982; Van Fleet & Atwater 1997; Lieberson et al. 2000; Hahn & Bentley 2003, among others). Your examples could include non-Western names as an additional signal of diversity. Sources for diverse names include the database of names compiled by Sanders, Umbal, and Konnoly (2020), which provides names for every letter of the English alphabet from different languages and cultures, categorized by gender (feminine, masculine, non-binary), and Kirby Conrod's list of non-binary names.<sup>19</sup>

When citing the existing literature, notice the trends represented in your examples. Where possible, consider citing a different source for better or more balanced examples. Likewise, when possible, you may choose to paraphrase an original example to avoid stereotypes and give the citation as “following” or “minimally changed from” the original. We acknowledge that in some cases this may not be possible or it may be difficult, for example when citing literature on less studied languages with fewer resources. In such cases, look for ways to offset any imbalances introduced in your examples in the rest of the text.

We acknowledge next and discuss two concerns that frequently arise in this context: How would one navigate the issues we have raised here (a) during fieldwork elicitation, and (b) in analyzing naturalistic corpus data. Unlike in constructed example sentences, where the linguist directly composes the data, data in both scenarios is constrained by additional factors.

Concentrating first on data collected in a fluid elicitation scenario, researchers should be mindful about gathering data that does not deal with sensitive or harmful topics, especially when the subject matter under discussion does not affect one's scientific aims. Similar to fully constructed examples sentences, sentence elicitation requires presenting a collaborator with small variations of the same sentence to tease apart linguistic differences. Before going on an extended inquiry, consider if there is implicit bias or stereotyping in your baseline sentence. When conducting fieldwork in general, consider redirecting the topic of discussion to something equal in evidentiary value while avoiding perpetuating stereotypes, whenever appropriate to do so.

Considering next corpus data, such data may comprise a variety of text types, such as narratives and conversations. When working with data drawn from corpora, we recommend exercising choice in which examples you cite, wherever that is possible. If multiple examples are available

that equally illustrate the concept under study, be mindful in selecting data that avoids enshrining social biases into an analysis. We readily acknowledge that our recommendations for both fieldwork and corpus work are an ideal, and will not always be possible. However, we expect that mindfulness and explicit planning of elicitation and corpus searches that take these issues into account will help mitigate at least some observed biases in current work.

Before concluding, we return again to the limitations of our study. First, as noted at the outset, we relied on conceptual (perceived) gender and assume a gender binary. This clearly does not represent the reality of gender outside of the linguistic literature, but it is a faithful representation of assumptions made within this literature. We furthermore do not examine the representation of race, non-binary individuals, non-cisgender or heterosexual representations, or any other minoritized or marginalized representation of individuals.

We additionally restricted ourselves to published work, but did not, for example, study unpublished manuscripts, preprints, or prior versions of published papers, nor conference presentation handouts or slides, or even conference proceedings papers. Moreover, the choice to focus our attention on example sentences means that the majority of our data comes from works published in the fields of syntax and (to a lesser extent) semantics. We acknowledge this limitation of our work, but believe that the findings here are relevant to any linguist who engages with example sentences in any capacity — be it in their research or in their teaching, including, importantly, in introductory linguistics courses.

Finally, we concentrated here on journals that publish mainly constructed examples. In preliminary work leading up to this study, we considered comparing such examples to those from corpora and naturally occurring speech. In particular, we mined data from the journals *Language Documentation and Description* and *Language in Society* following the same procedures described in section 3.1. However, we found the data in both journals to (a) use far fewer example sentences in general, (b) be more heavily comprised of first and second person arguments, and (c) contain more naturalistic dialogue not easily parsed into separate sentences. On the whole we decided that it would not be possible to compare the latter journals to the other three represented in this study, and instead chose to leave such a comparison to a separate future study. However, we

acknowledge that this is a potential factor that could affect the distribution of gender of arguments, which could be interesting and important to consider.

## REFERENCES

- Ackerman, Lauren. 2019. Syntactic and cognitive issues in investigating gendered coreference. *Glossa: A Journal of General Linguistics* 4:117.
- Bergvall, Victoria. 1996a. Humpty Dumpty does syntax: Through the looking glass, and what Alice found there. *Natural Language & Linguistic Theory* 14:433–443.
- Bergvall, Victoria. 1996b. 'merely data'? gender bias in constructed linguistic examples. Presentation at the 70th Annual Meeting of the Linguistic Society of America, San Diego.
- Bjorkman, Bronwyn. 2017. Singular *they* and the syntactic representation of gender in English. *Glossa: A Journal of General Linguistics* 2:80.
- Boroditsky, Lara, Lauren Schmidt, and Webb Phillips. 2003. Sex, syntax, and semantics. In *Language in mind: Advances in the study of language and cognition*, ed. D. Gentner and S. Goldin-Meadows, 61–80. Cambridge University Press.
- Charity Hudley, Anne, and Christine Mallinson. 2019. Linguistic Society of America statement on race. URL <https://www.linguisticsociety.org/content/lisa-statement-race>, with contributions from Mary Bucholtz, Nelson Flores, Nicole Holliday, Elaine Chun, Arthur Spears, and Wesley Y. Leonard.
- Kibbey, Tyler. 2019. Transcriptivism: An ethical framework for modern linguistics. Presentation at the 93rd Annual Meeting of the Linguistic Society of America, New York City.
- Linguistic Society of America. 2019. The state of linguistics in higher education. URL [https://www.linguisticsociety.org/sites/default/files/Annual\\_Rept\\_Final\\_2019.pdf](https://www.linguisticsociety.org/sites/default/files/Annual_Rept_Final_2019.pdf), annual report: Seventh Edition.
- Linguistic Society of America Committee on the Status of Women in Linguistics. 2016. Guidelines for inclusive language. URL <https://www.linguisticsociety.org/resource/guidelines-inclusive-language>, with additional input from LSA committee chairs.
- Liu, Bing. 2012. Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies* 5:1–167.
- 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.
- Namboodiripad, Savirthry, Lynn Hou, and Corrine Occhino. 2019. A survey of linguists and language researchers: Harassment, bias, and what we can do about it. URL <https://sites.google.com/umich.edu/lingclimatesurvey/home>, plenary panel at the 93rd Annual Meeting of the Linguistic Society of America, New York City.
- Pabst, Katharina, Paola Cépeda, Hadas Kotek, and Kristen Syrett. 2018. Gender bias in linguistics textbooks: Has anything changed since Macaulay & Brice (1997)? Presentation at the 92nd Annual Meeting of the Linguistic Society of America, Salt Lake City.
- R Core Team. 2013. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.
- Sanders, Nathan, Pocholo Umbal, and Lex Konnoly. 2020. Methods for increasing equity, diversity, and inclusion in linguistics pedagogy. In *Proceedings of the 2020 Annual Conference of the Canadian Linguistic Association*.
- Valian, Virginia. 1998. *Why so slow*. Cambridge, MA: MIT Press.
- Valian, Virginia. 2005. Beyond gender schemas: Improving the advancement of women in academia. *Hypatia* 20:198–213.

## NOTES

<sup>1</sup>Revised and expanded from the LSA *Guidelines for non-sexist usage* (1997).

<sup>2</sup>In more detail, GRAMMATICAL GENDER comprises formal morphosyntactic features, namely the properties of words that allow the formal grammatical process of agreement to be carried out. This includes agreement of grammatical gender categories such as masculine, feminine, neuter, common, etc. These features are properties of the morphemes themselves, and may be independent from the real-world biosocial genders associated with the referents. See Ackerman 2019 for details.

<sup>3</sup>This is a word-play on the homonym ‘pair’, a slang term for a woman’s breasts.

<sup>4</sup>We chose to exclude Macaulay and Brice 1997, published in *Language*, to avoid skewing the results.

<sup>5</sup>This means that we did not explore any examples that appeared in the text of a paper, outside of numbered examples. Furthermore, we recognize that our automated means of extracting examples may have caused some to be missed. If our method missed data, we have no reason to believe it would skew the dataset in any meaningful way. Therefore, we believe that our results are representative and valid.

<sup>6</sup>That is, we did not consider inanimate nouns with grammatical gender in languages such as Hebrew or German, such as *ha-gesher* (the-bridge.masc, Hebrew) or *die Brücke* (the.fem bridge, German), despite the fact that there exists research showing that the way speakers refer to such nominals does differ in stereotypical ways that follow from the grammatical gender assigned to them (Boroditsky, Schmidt, and Phillips 2003). Here our focus is solely on arguments which refer to humans or anthropomorphized individuals. As anthropomorphized individuals were rare in the study, we refer to arguments as human in the remainder of the text.

<sup>7</sup>We realize that this may result in some inaccuracies in the coding in cases where the structure of the language in question varied from that of English, and the coders were not aware of the difference. For example, a reviewer provides the case of French ‘tu me manques’ vs. English ‘I miss you’ where the participants are in different syntactic positions. We believe such cases to be fairly rare. Further, the fact that our findings do not differ based on the language of the example (see section 3.4.1) leads us to believe that this choice did not greatly affect our results.

<sup>8</sup>These include arguments such as “the students,” “who,” and “everyone”, and names such as “Taylor,” which can traditionally be assigned to both male-identifying and female-identifying individuals, when no corresponding pronoun or context can disambiguate the intended referent. We return to the case of names in section 3.2.3.

<sup>9</sup>We thank Susan Fischer (p.c.) for this suggestion.

<sup>10</sup>Our spreadsheet included all names not familiar to the authors, for review by the community. This was especially important in order to minimize the number of non-Western names that were excluded from the study as unknown. See the acknowledgements footnote for the names of those who helped in this process.

<sup>11</sup>Our dataset additionally contains example sentences featuring 35 male-gendered, 15 female-gendered, and 7 non-gendered “geniuses.” Given these small numbers, we offer that data without further comment.

<sup>12</sup>Interestingly, this is the opposite finding of the one in Pabst et al. 2018, where women are over-represented in negative emotion cases. We do not offer a potential explanation here.

<sup>13</sup>We once again acknowledge here that the binary classification is imperfect in the ways described earlier in the paper, especially when discussing working linguists rather than fictional individuals in example sentences.

<sup>14</sup>The models reported in the text include all paper authors (first and non-first authors). Models were also run considering the gender of first authors only. The results from those models show the same statistical trends as for all authors, and therefore we do not report on them separately here.

<sup>15</sup>Pabst et al. (2018) report that female authors are more likely to use non-gendered arguments than male authors in their study.

<sup>16</sup>[https://www.linguisticsociety.org/sites/default/files/Annual\\_Rept\\_Final\\_2019.pdf](https://www.linguisticsociety.org/sites/default/files/Annual_Rept_Final_2019.pdf), in particular “Job Type by Gender”, pages 15–17.

<sup>17</sup><https://sites.google.com/umich.edu/lingclimatesurvey/home> (Namboodiripad, Hou, and Occhino 2019)

<sup>18</sup><https://www.linguisticsociety.org/resource/guidelines-inclusive-language>

<sup>19</sup>[https://docs.google.com/spreadsheets/d/1GF6c5qFFzTqYGukRYia8WcSam48tBHm\\_](https://docs.google.com/spreadsheets/d/1GF6c5qFFzTqYGukRYia8WcSam48tBHm_)

[R6MJB5tJPiI/edit#gid=0](#). The document is described as follows: “These are names (and pronouns, where possible) of people who responded to a tweet asking for nonbinary volunteers who’d be okay with their names being used in linguistics examples ... The purpose of the list is to provide linguists with names to use in example sentences, which historically have suffered from significant gender bias.”