A Framework for Understanding Computational Corpus Studies of Gender in Language

Nicholas Turk

May 1, 2024

1 Introduction

Gender is a foundational aspect of human expression, but one that often goes improperly interrogated. Specifically in the related fields of Natural Language Processing (NLP) and Computational Linguistics, gender is often treated as an immutable demographic fact, rather than a socially constructed category (Cao and Daumé III, 2021). Our endeavor is to review the literature on computational linguistic approaches to gender, seeing how various studies construct and deconstruct gendered language, while also reflecting on how NLP tools can be used to promote a more thorough understanding of sociolinguistic gender in corpus studies.

Broadly, we know that social categories like gender manifest both within the text and metadata of a corpus (Nguyen et al., 2020), and the way that researchers approach gender will vary depending on the kind of data they're observing.

2 Textual Analysis

The text of a corpus consists of the actual speech acts which the corpus records, whether that's a set of books or a collection of Reddit posts. These speech acts of course record gendered language, and the way that the corpus' text expresses gender can explain to us how gender is performed and perceived within the context of that corpus. We can analyze gendered language perception in these corpora most easily through the lens of overtly gendered language.

2.1 Overtly Gendered Language

Many languages have certain lexical indicators of gender. In English, for example, nouns like "woman" encode gender, and so do pronouns like "she".

These simple encodings of gender allow for what essentially amounts to first-wave linguistic study of gender, where it is treated as a broad demographic pattern. By analyzing

the context in which gendered words appear, we can observe the extent to which gender roles are normalized within our language (Bailey et al., 2022). In addition, using NLP tools like Named Entity Recognition, we can determine how people are described based on their gender, giving us quantitative findings on long-held understanding of gender roles, like how women were historically described as more emotional than men in English-language fiction (Underwood et al., 2018).

These analyses offer useful quantitative analyses of broad social trends, but they often find themselves disregarding the fluidity of gendered performance due to their treatment of gender as a demographic trend rather than a sustained act. They also often find themselves limited in their study of gender demographics by their corpora, which rarely represent nonbinary genders at all due to the skew of language use towards the social norm of binary gender (Lauscher et al., 2022).

There are, however, other ways to view gender within textual analysis. Not all use of gender in language is overt.

2.2 Covertly Gendered Language

Covertly gendered language sees either the application of complex gender constructions or normative gender roles. On the side of complex gender construction, we may see something like the sentence "At the Halloween party, the cowgirl left his lasso in the kitchen" (Ackerman, 2019), which acts as a complex example of gender structures, where a male referent is playing the female role of cowgirl. This struggle to understand the potential complexity of gender when referring to one individual has led to struggles in the NLP domain of coreference, where the complexities of actual gender identity and performance often find themselves inadequately expressed in the systems which are designed to classify them lexically (Cao and Daumé III, 2021).

We can also see gender become a feature of language when gender roles interact with neutral-gendered language. Due to the prevalence of gender roles in training data for NLP tools, the tools often express a bias where different terms will be associated with different genders, even if the terms themselves are ungendered (Stanczak and Augenstein, 2021).

Prior discussions of these topics seem to assume that the existence of gender bias in neutral-gendered language is a representational issue of NLP tools (Stanczak and Augenstein, 2021), but the truth may be a little more difficult. Remember that gender is a performance, and while normative gender roles are harmful at the demographic level, they are vital to the very existence of that demographic perception of gender. Individuals perform their gender through the roles that they play in society. Language, then, reflects the expectations that this society produces back to us. We see here an open, yet largely undiscussed question: to what extent should our language reflect the biases of societal gender roles?

2.3 Textual Gender: Wrap Up

We see from this section that despite the broad and well-established applications of textual gender analysis, these analyses leave out many of the topic's most vital complexities. Luckily, we can try to tackle these complexities with metatextual analysis.

3 Metatextual Analysis

Corpora don't appear from thin air. The text is gathered from somewhere, and from someone.

The speakers in a corpus and the social context that the corpus' speech acts take place in offer indispensable context for understanding gender. We know, for instance, that social media gender performance is associated with the gender identity of the performer's social media contacts (Bamman et al., 2014): information that can only be extracted by knowing the structure of the social media platform under study.

We also see that this additional information can inform a more in-depth, essentially second-wave sociolinguistic approach to gender in corpora. By studying corpora originating from specific social groups, we can see how those social groups performed gender, but also how those groups constructed gender roles in their language (Underwood et al., 2018). These studies often find themselves with the same issues as before, where nonbinary genders are underrepresented, but they have an easier solution: draw on speech from communities with many nonbinary people and study those communities in more detail (Lauscher et al., 2022).

4 Conclusion

Computational studies of gender in corpus linguistics have made strides towards investigating gender in all its complexity, but the biases inherent to many NLP tools and the broader societal struggles of gender discrimination makes this field of study difficult. While first-wave and second-wave sociolinguistics have found purchase within computational corpus studies, third-wave studies seem scarce to nonexistent, despite the fact that corpus linguistics often relies on data from social media, where users will often denote the social context of their actions in very clear fashion (posting to a specific subreddit, for example). The field will undoubtedly grow and change from here, but it will need to reevaluate its own view on the subject at hand. The current treatment of gender as a demographic fact can only go so far, and we're running up against that limit.

References

Ackerman, L. (2019). Syntactic and cognitive issues in investigating gendered coreference. *Glossa: a journal of general linguistics*, *4*(1).

Bailey, A. H., Williams, A., & Cimpian, A. (2022). Based on billions of words on the internet, people= men. *Science Advances*, *8*(13), eabm2463.

Bamman, D., Eisenstein, J., & Schnoebelen, T. (2014). Gender identity and lexical variation in social media. *Journal of Sociolinguistics*, *18*(2), 135–160.

Cao, Y. T., & Daumé III, H. (2021). Toward gender-inclusive coreference resolution: An analysis of gender and bias throughout the machine learning lifecycle. *Computational Linguistics*, *47*(3), 615–661.

Lauscher, A., Crowley, A., & Hovy, D. (2022). Welcome to the modern world of pronouns: Identity-inclusive natural language processing beyond gender. *arXiv preprint arXiv:2202.11923*.

Nguyen, D., Liakata, M., DeDeo, S., Eisenstein, J., Mimno, D., Tromble, R., & Winters, J. (2020). How we do things with words: Analyzing text as social and cultural data. *Frontiers in Artificial Intelligence*, *3*, 62.

Stanczak, K., & Augenstein, I. (2021). A survey on gender bias in natural language processing. *arXiv preprint arXiv:2112.14168*.

Underwood, T., Bamman, D., & Lee, S. (2018). The transformation of gender in english-language fiction.