

On Genitalia, Reproduction and Pleasure: Biases in the Representation of Sexes

Sadhi Vornberger

Independent Researcher

sadhi@sadhivornberger.de

Peter Bourgonje

Universität Potsdam

bourgonje@uni-potsdam.de

1 Introduction

Research on cultural and social norms in textual and other representations of human sexuality and gender has well documented that bodies and people defined by the respective sources as “male” are often associated with sexual pleasure, whereas bodies and people defined as “female” are placed in a context of reproduction (Moore and Clarke, 1995; Vornberger, 2015; Bourdieu, 2002; Young et al., 2019). This bias is problematic since it leads to the discrimination, exclusion, and impossibility to express identities fully and authentically for many individuals and undermines equality, equal access to resources and even the respectful and dignified framing of people in the discussion of medical issues.

In this contribution, we attempt to investigate this bias that links reproduction to bodies defined as “female”, and sexual pleasure to bodies defined as “male”. More specifically, we suspect that there is a stronger association with reproduction where medical terms for “female” genitalia are concerned, and a stronger association with sexual pleasure where medical terms for “male” genitalia are concerned.

We apply this principle to the NLP domain by using state-of-the-art text representation methods, and investigate this bias in both general domain and domain-specific resources.

2 Data & Method

We set out to assess to what extent our assumption about “(fe)male” genitalia being associated to reproduction or sexual pleasure holds, by extracting cosine similarity scores for term-pairs from contextualised word embeddings. We select terms related to genitalia, and to reproduction and sexual pleasure.

2.1 Genitalia Terms

We assume the categories of both sex and gender to be socially constructed (Sgier, 1994; Fausto-Sterling, 1992; Vornberger, 2011), neither exhaustive nor applicable to every individual (body), and the criteria used to categorise are problematic, controversial, ambiguous, and rarely clearly defined, even in research claiming to be scientific (Fausto-Sterling, 2000; Vornberger, 2011, 2015). People and bodies that are categorised as either “female” or “other” usually are underrepresented in comparison to those categorised as “male” (ibid., Krüger (2019)). In order to research this bias and expose where it exists, we temporarily need to recognise sex as binary category, even though to us, biologically-empirically, it appears more reasonable and beneficial to use non-binary ways of representing sex, for example, in a continuum. For this reason, we consistently use quotation marks for the terms “female” and “male”, and adopt the binary categorisation just for the purpose of investigating it in popular NLP representation methods.

We compile a list of sexually defined genitalia terms (Table 1 in the Appendix), based on the catalogue of examination topics (subject human anatomy) of the German National Institute for State Examinations in Medicine, Pharmacy and Psychotherapy (IMPP¹). We are aware of the potential for bias we introduce by this procedure, since the catalogue itself suffers from an underrepresentation of the clitoris in comparison to the penis (the former is not listed as a genital organ in its own right). Our past research has shown that anatomical information on the clitoris is generally harder to find than on the penis, for example in regards to their erectile tissues (Vornberger, 2015; Krüger, 2019). In order to avoid our data just representing this bias, we run the experiment with the original IMPP list, as well as with the IMPP list supple-

¹<https://www.impp.de>

mented by the term "clitoris" (indicated in grey).

To get some idea of frequency for the terms we selected, we checked their frequencies in the English sentences of the UFAL Medical Corpus² (223 million words).

2.2 Reproduction and Sexual Pleasure Terms

For lack of a more authoritative source, our terms related to sexual pleasure are the result of the authors' creativity, consulting thesauri³ and Wikipedia. See Table 2 for the result (again augmented with frequencies in our reference corpus).

2.3 Term Pair Associations

To investigate the suspected bias, we calculate cosine similarity scores for term pair combinations. We average the scores of all 5 "female" genitalia terms paired with the 2 reproduction terms (10 pairs in total) and 5 sexual pleasure terms (25 pairs in total), and repeat this process for the terms related to male genitalia. We attempt to investigate this bias both in the general domain and use the `bert-base-cased` model (Devlin et al., 2018) (trained on English books (Zhu et al., 2015) and the English Wikipedia), and in the clinical domain and use the embeddings specifically trained on biomedical texts (electronic health records from ICU patients) from Alsentzer et al. (2019). We encode the input, average the model output if the tokenizer splits up the input into more than one token⁴, get the representation of the last hidden state and use that as input for cosine similarity calculation. The last four Tables in the Appendix include the results.

3 Results & Discussion

With regard to frequency, "male" genital organs are overrepresented in comparison to "female" genitalia. For both the general and clinical domains, our bias is confirmed that "female" genitalia terms are more often associated to reproduction than to sexual pleasure. When running the experiment adding the term "clitoris" to the list of "female" genitalia, the effect is comparable, but not as strong (Tables 5 and 6). The "male" column (general domain) shows the complementary bias, with a stronger association to sexual pleasure than to reproduction,

²https://ufal.mff.cuni.cz/ufal_medical_corpus

³<https://www.linguee.com/>, <https://dict.leo.org/german-english/>

⁴Either because the input is a multi-word term, or because of BPE subword tokenization.

although the bias is stronger for the "female" genitalia column ($0.08 > 0.02$). Looking at the rows however, "male" terms have significantly stronger associations to both reproduction and sexual pleasure. The former rejects our hypothesis: we expected "female" genitalia terms to have a stronger association to reproduction than "male" genitalia terms. A possible explanation could be that masculinity is often described as active, and femininity as passive (Keller, 1995): "male" genitalia are represented as more relevant to (and actively generating) both reproduction and sexual pleasure, and "female" bodies merely as passive recipients.

Tables 4 and 6 show that overall, the gap between associations between sexual pleasure on the one hand and reproduction on the other hand is smaller for the medical domain (but the bias is still noticeable), though it is larger for the "female" than for the "male" terms ($0.3-0.4 > 0.1$). More notable, in the clinical domain, it is not the case that "male" terms are more associated to sexual pleasure than to reproduction, but the reverse. This needs more investigation.

Using contextualised embeddings to represent single words in a way defeats their purpose. Using static word embeddings (e.g., Mikolov et al. (2013)) would be more appropriate for our use case. However, since we aim to investigate this bias in currently popular methods, we opted for contextualised embeddings instead.

4 Conclusion

In this ongoing work, we investigate a potential bias to associate terms related to "female" genitalia to reproduction, and terms related to "male" genitalia to sexual pleasure. While some of our suspicions are confirmed (stronger association of "female" genitalia to reproduction than to sexual pleasure), others are not (for example, our current data does not suggest a stronger association of reproduction to "female" genitalia than to "male" genitalia). We plan to continue this line of work by including more languages, and by further developing and refining the terms (e.g. including colloquial terms). Furthermore, we intend to critically assess cosine similarity as a proxy for association, the (statistical) significance of differences for term pairs, and the effects of the skewed frequency distributions of the terms related to "female" and "male" genitals, and reproduction and sexual pleasure.

References

- Emily Alsentzer, John Murphy, William Boag, Wei-Hung Weng, Di Jindi, Tristan Naumann, and Matthew McDermott. 2019. [Publicly available clinical BERT embeddings](#). In *Proceedings of the 2nd Clinical Natural Language Processing Workshop*, pages 72–78, Minneapolis, Minnesota, USA. Association for Computational Linguistics.
- Pierre Bourdieu. 2002. *La domination masculine*. Saint-Amand-Montrond: Éditions du seuil.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. [BERT: pre-training of deep bidirectional transformers for language understanding](#). *CoRR*, abs/1810.04805.
- Anne Fausto-Sterling. 1992. *Myths of Gender: Biological Theories About Women and Men*. Basic Books.
- Anne Fausto-Sterling. 2000. *Sexing the Body: Gender Politics and the Construction of Sexuality*. Sexing the Body: Gender Politics and the Construction of Sexuality. Basic Books.
- Evelyn Keller. 1995. [Gender and science: Origin, history, and politics](#). *Osiris*, 10.
- S. Krüger. 2019. Representation of the female genital area in biology schoolbooks - a contribution to emancipatory sex education. Master's thesis, Humboldt-Universität zu Berlin.
- Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. [Distributed representations of words and phrases and their compositionality](#). In *Advances in Neural Information Processing Systems*, volume 26. Curran Associates, Inc.
- Lisa Jean Moore and Adele E. Clarke. 1995. "Clitoral Conventions and Transgressions: Graphic Representations in Anatomy Texts, c1900-1991.". *Feminist Studies*, 21, no. 2.
- I. Sgier. 1994. *Aus eins mach zehn und zwei lass gehn: Zweigeschlechtlichkeit als kulturelle Konstruktion*. Schriftenreihe (Verein Feministische Wissenschaft Schweiz). eFeF-Verlag.
- S. Vornberger. 2011. Naturwissenschaftskritik und Einkörperung bei Fausto-Sterling: Wie der (geschlechtliche) Körper soziologisch gefasst werden kann. Bachelor's thesis, Universität Konstanz.
- S. Vornberger. 2015. L'oiseau et le ver. Les représentations anatomiques du clitoris et l'anatomie comme lieu de luttes de signification. Master's thesis, Université de Lausanne.
- Kate Young, Jane Fisher, and Maggie Kirkman. 2019. "do mad people get endo or does endo make you mad?": Clinicians' discursive constructions of medicine and women with endometriosis. *Feminism & Psychology*, 29(3):337–356.
- Yukun Zhu, Ryan Kiros, Rich Zemel, Ruslan Salakhutdinov, Raquel Urtasun, Antonio Torralba, and Sanja Fidler. 2015. [Aligning books and movies: Towards story-like visual explanations by watching movies and reading books](#). In *2015 IEEE International Conference on Computer Vision (ICCV)*, pages 19–27.

A Appendix

Terms Related to “Female” Genitalia	Terms Related to “Male” Genitalia
uterus (2,137)	prostate (8,016)
vagina (1,674)	penis (2,001)
ovaries (1,614)	testicles/testes (861)
vulva (289)	ejaculate (76)
fallopian tubes (149)	epididymides (63)
clitoris (127)	seminal vesicles (45)
—	spermatic cord (33)
total: 5,990	total: 11,095

Table 1: Terms related to “female” and “male” genitalia

Synonyms of Reproduction	Synonyms of Sexual Pleasure
reproduction (6,724)	arousal (386)
procreation (18)	excitement (287)
—	orgasm (258)
—	pleasure/sexual pleasure (234)
—	lust (17)
total: 6,742	total: 1,182

Table 2: Synonyms of reproduction and sexual pleasure

	“Female” Genitalia	“Male” Genitalia
Reproduction	0.58	0.63
Sexual Pleasure	0.50	0.65

Table 3: Cosine similarities on bert-base-cased without “clitoris”

	“Female” Genitalia	“Male” Genitalia
Reproduction	0.74	0.76
Sexual Pleasure	0.70	0.75

Table 4: Cosine similarities on Bio_ClinicalBERT without “clitoris”

	“Female” Genitalia	“Male” Genitalia
Reproduction	0.60	0.63
Sexual Pleasure	0.55	0.65

Table 5: Cosine similarities on bert-base-cased with “clitoris”

	“Female” Genitalia	“Male” Genitalia
Reproduction	0.74	0.76
Sexual Pleasure	0.71	0.75

Table 6: Cosine similarities on Bio_ClinicalBERT with “clitoris”