Identifying Gendered Language

Shweta Soundararajan and Sarah Jane Delany

School of Computer Science, Technological University Dublin

What is the aim?

- To automatically identify examples of written content that can be unconsciously associated with gender using machine learning and natural language processing techniques.

What is Gendered Language?

Language that explicitly or implicitly indicates the gender of an entity [1][2].

- Use of words, phrases, or grammatical structures that reflect societal gender norms and expectations.
- May contain and prompt gender-based biases, stereotypes, or preferences.

Examples

Ambitious boy **Beautiful** girl

Muscular man **Pretty** woman

Why is it important to identify it?

For promoting fairness, inclusivity, and effective communication, to avoid bias and stereotyping, improve communication and meet legal requirements.

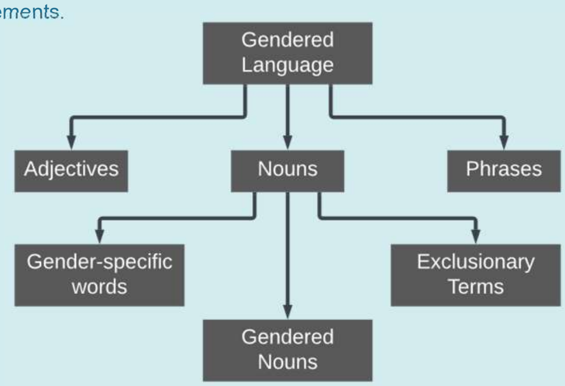

Fig. 1. Gendered Language

Challenges

- Gendered language can perpetuate gender stereotypes and bias and can exclude people with non-binary genders.
- A lack of labelled datasets to use in modelling using supervised machine learning.

Methodology

- Lexicons identify gender-coded words, both masculine and feminine. Some include a gender score.
- Used two lexicons: Gaucher et al., 2011 [3], Cryan et al., 2020 [4]

Lexicon	No. of Masculine words	No. of Feminine words	Total
1 [3]	42 (51.21%)	40 (48.78%)	82
2 [4]	702 (53.71%)	605 (46.28%)	1307

- Used ChatGPT to generate labelled datasets of gendered and non-gendered sentences based on lexicon words.

Gendered sentences
- Sentences about males with masculine terms (**MM**)
- Sentences about females with feminine terms (**FF**)

Non-gendered sentences
- Sentences about males with feminine terms (**MF**)
- Sentences about females with masculine terms (**FM**)

- Evaluated whether the sentences can be classified as gendered or non-gendered using supervised machine learning.
- Used stratified 5-fold cross validation to measure the classification performance on the generated datasets.
- Generated multiple datasets from the same lexicon to measure stability of performance.

Data Generation

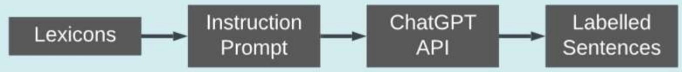

Fig. 2. Pipeline for the data generation

Instruction Prompt

Generate [x] sentences in the style of context from newspapers, magazines, children's books, job advertisements, story books, etc. using all the words listed below. Simple sentences, compound sentences, complex sentences, and compound-complex sentences must all be included in the list of sentences. Each and every sentence must necessarily be about a female, females, woman, women, girl, or girls. Any tenses and any parts of speech can be used in the sentences. The sentences can use pronouns, nouns, the name of a person, etc. to refer to the female, females, woman, women, girl, or girls being discussed in the sentences. All of the sentences must use one or more of the words mentioned below as an adjective or noun to depict the characteristic or traits of the female, females, woman, women, girl, or girls being discussed in the sentence. More than 15 words must be used in each sentence.
word_1, word_2, word_3, ….word_n

Dataset	No. of Gendered samples			No. of Non-gendered samples			Total no. of samples
	MM	FF	Total	MF	FM	Total	
1 [3]	42	40	82	40	42	82	164
2 [4]	702	605	1307	605	702	1307	2614

Sample sentences

MM: The **macho** man flexed his muscles and let out a powerful roar.

FF: Her **childlike** innocence made her oblivious to the sepia photographs in the antique shop.

MF: The **vivacious** boy who was always bouncing around the room won the race.

FM: The **manly** lady took on the traditionally male-dominated job with confidence.

Evaluation and Results

Dataset	Overall Accuracy (%)	Recall of Gendered Samples (%)	Recall of Non-gendered samples (%)	Bias in Gendered samples (%)	Bias in Non-gendered samples (%)
1 [3]	67.9 ± 6.5	67.1	68.7	10.7	17.1
2 [4]	77.2 ± 2.7	76.2	78.1	7.3	8.7

- Gender Bias of the datasets was measured using the equation:

$$\mathbf{Bias} = |\,\mathrm{Recall}_{male} - \mathrm{Recall}_{female}\,|$$

- Results show the dataset from lexicon 2 generates good examples of English natural language sentences that capture the characteristics of gendered and non-gendered language whereas lexicon 1 is more limited.

Acknowledgement

This research was funded by Technological University Dublin under the TU Dublin Scholarship – Presidents Award.

Have any feedback?

References:

[1] Bigler et al., (2015), Gendered language: Psychological principles, evolving practices, and inclusive policies, Policy Insights from the Behavioural and Brain Sciences, pp. 2(1) 187-194.
[2] Hamidi et al., (2018), Gender recognition or gender reductionism? The social implications of embedded gender recognition systems, CHI (2018), pp. 1-13.
[3] Gaucher et al., (2019), Evidence that gendered wording in job advertisements exists and sustains gender inequality, Journal of personality and social psychology, pp. 101(1):109.
[4] Cryan et al., (2020), Detecting gender stereotypes: lexicon vs. supervised learning methods, CHI (2020), pp. 1–11.