Women in the Workplace: Analyzing the Gender Biases in Corporate Email Communications

Computational Linguistics, Natural Language Processing, Gender Bias, Organizational Discourse, Email, Inclusivity

Extended Abstract

Language both reflects society and influences the perspectives and structures within it. Notably, with the pervading gaps in the representation of women across many areas, gender biases present the potential dangers of such disparities. The pervasiveness of gender biases across language has been a prevalent focus in computational social sciences research. Research concerning gender equality has expanded across sectors to reveal the inequalities in news, media (Dacon and Liu, 2021), and literature (Babaeianjelodar et al., 2020).

With the prominence of gender disparities in the workplace, organizations are especially relevant to study. In particular, email communications play an important role in the transmission of information and the administrivia within organizations. Gender biases in the workplace have been studied in attempts to identify and address the inequality of women in corporate environments; they have been found to hinder women in job applications, recommendations, and progression to managerial positions (Strol, 2020). However, despite the scale in the use of email and its relevance in organization discourse, email is under-studied in workplace language and gender research.

In our study, we work with the Enron Corpus (Klimt and Yang, 2004) as a case study of corporate communications. The Enron Corpus is a uniquely available dataset, consisting of nearly half a million emails from Enron employees, as it is of the only open source corpus of real emails and enables the naturalistic study of real workplace discourse. The methods of our research took different angles of measuring linguistic gender biases to provide a comprehensive analysis on the role of gender in the workplace.

Our first question examines gender bias through differences in language use toward men and women. We ask whether men and women are treated differently at work over email by determining whether language use is predictive of email receivers' genders. Computational text analysis through LIWC generated linguistic and psychological features from emails' bodies and enabled the analysis of imbalances in language use. In order to identify linguistic differences based on the directed receiver gender, we developed a predictive model with a gradient-boosting based machine learning algorithm, and analyzed the results with model explainability methods to extract specific characteristics distinguishing emails to men from those to women. We determined that the linguistic and structural characteristics of the email held significant predictive power in binary classification of the gender of the email receiver. Our results indicate potential bias in the interactions of men and women in the workplace in treatment and organizational role distribution.

The second question explored gender bias through the asymmetric associations of gender with profession words as captured by work embeddings. We study the gender disparities in the content of emails to determine the imbalance in how men and women are addressed in email, identifying genderedness in ungendered profession words. We follow the methods of Bolukbasi et al. (2019) to measure direct gender bias, generating embeddings based on the

discourses of the data. From the procedure, we determine a direct gender bias of 0.08 in the emails. According to Bolukbasi et al., this value confirms occupation words to have significant components along the gender direction. The substantial associations of ungendered professions with genders in the language of workplace communications presents further evidence of bias in an important aspect of corporate structures.

Our third question aims to identify sexist phrases in workplace emails, creating a classification model to analyze the distribution of such statements in organizational discourse and propose its use as a tool for flagging problematic language during email composition in the workplace. We work with the ISEP dataset (Grosz and Conde-Cespedes, 2020), and develop a model to perform the task of classifying the statements for sexism. We fine-tune four large language models on the ISEP dataset to build a tool for predicting sexist comments common in the workplace. Once we determine the effectiveness of NLP tools for flagging sexist statements, with a ROC AUC score of 0.97, we examine the prevalence of such comments in real workplace emails, applying the model to the Enron emails. On the Enron dataset, the model identified about 10% of sentences to be sexist, revealing that such comments were quite common in real organizational communications.

In this paper, we examined the presence of gender biases in workplace email communications on multiple dimensions. Our extensive analysis reveals gender biases on multiple levels confirm the inequality faced by women in workplaces that affect women's careers. Further implications of these findings in the Enron dataset expand to potential discrimination persisting in the present day's companies. Representation of women in the C-suite and high corporate positions is scarce, and understanding the everyday gender biases that influence women provides insight into how the views of surrounding individuals may dictate such gaps. This establishes a need for ways to address prejudice and promote diversity in corporations.

References

Marzieh Babaeianjelodar, Stephen Lorenz, Josh Gordon, Jeanna Matthews, and Evan Freitag. 2020. Quantifying Gender Bias in Different Corpora. In Companion Proceedings of the Web Conference 2020, pages 752–759, New York, NY, USA, April. Association for Computing Machinery.

Tolga Bolukbasi, Kai-Wei Chang, James Zou, Venkatesh Saligrama, and Adam Kalai. 2016. Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings. July. arXiv:1607.06520 [cs, stat].

Jamell Dacon and Haochen Liu. 2021. Does Gender Matter in the News? Detecting and Examining Gender Bias in News Articles. In Companion Proceedings of the Web Conference 2021, pages 385–392, New York, NY, USA, April. Association for Computing Machinery.

Dylan Grosz and Patricia Conde-Cespedes. 2020. Automatic Detection of Sexist Statements Commonly Used at the Workplace. In Wei Lu and Kenny Q. Zhu, editors, Trends and Applications in Knowledge Discovery and Data Mining, pages 104–115, Cham. Springer International Publishing.

Bryan Klimt and Yiming Yang. 2004. Introducing the Enron Corpus. In CEAS 2004 - First Conference on Email and Anti-Spam, July 30-31, 2004, Mountain View, California, USA.

Oksana O. Strol. 2020. Gender-Biased Language of the Workplace., January.

Feature Name	Feature Definition/Example	Feature Importance	Associated Gender
BigWords	Percent words >= 7 letters	0.1502	M
achieve	Achievement (ex: better, best)	0.0670	M
reward	Reward (ex: opportun*, win)	0.0326	M
Clout	Language of leadership, status	0.0256	M
Tone	Degree of positive tone	0.0070	M
work	Work (ex: work, working)	0.0899	F
number	Numbers (ex: one, two)	0.0640	F
prosocial	Prosocial behavior (ex: care)	0.0513	F
WPS	Average words per sentence	0.0282	F
i	1st person singular (ex: I, my)	0.0078	F

Figure 1. Feature importances of select features in predicting email receiver gender based on linguistic features. Features are sorted by associated receiver gender and feature importance (FI) on the model. High FI indicates impact on the model's predictions, and the associated gender indicates the direction.

Identified Phrase
This one has no volume but be careful. Why women can't be mechanics...
all women are noisy fucks.
This must have been created by one of your fellow engineers. You guys just have a bad case of penis envy.
I do not know anything about Kristen. I prefer at least one aggressive person on the desk, Monte and Ashley are kind of shy.

Figure 2. Biased comments from emails in the Enron dataset identified with the model trained on the ISEP dataset to detect sexism.