Berra Karayel

0054477

CSSM 530 / Ali Hürriyetoğlu

DOES SEXISM HOLD WOMEN BACK? :

SEXIST LANGUAGE AS AN OBSTACLE ON WOMEN'S ECONOMIC EMPOWERMENT AND ENTRANCE TO PUBLIC SPHERE

Domain Expert: David G. Carlson

What holds women back is a really crucial question to consider today. Many women are still seen as inferior in many societies and as a result of this perception, they remain their secondary status in society and many women are being more oppressed as getting closer to today. Since the perception of women's empowerment has been changing positively, still many women are not be able to participate in labor organizations and education. This issue has not been taking enough attention, although it is a central issue for governmental and nongovernmental organizations. Besides, it is really crucial for the development and sustainability of societies.

Still, many women are underrepresented in male-dominated areas like natural sciences, business, engineering and so on. According to many studies, there are various factors which affects these disparities like stereotypes, perception, and beliefs. Fewer of these factors are associated with institutional factors which perpetuate the gender inequalities within the social structure itself such as law, public policy, and so on. As people's perception is one of the biggest obstacles on women's empowerment, these institutional factors are the ones which influence people's attitude as perception towards women in those fields in addition to individual factors.

In addition to people's perception and attitude towards women 'employment, the perception of women on occupations would be another obstacle in their lower participation rate. As a part of the institutional mechanism, the reason why there is fewer women in traditionally male- dominated jobs would be the sexist language that is used in job advertisements. Employees and institutions do not give explicitly the name of the sexes that they are looking for, however, they use some gendered words which gives some subtle clues to people who want to apply to those jobs. They use some words which are associated with certain sexes. For instances, while some words like leader, dominant, active, and challenging are associated with male stereotypes, there are some words like understanding, caring, gentle, and collaborative which are identified as feminine words. This gendered wording in job advertisements is eliminating women during the first phase of their job seeking process and influence their perception towards certain jobs which have more masculinity related words. Women may find those advertisements less appealing, even though, they might be a suitable employer for the position and limit the applicant pool for these gendered advertised jobs.

Research Question & Analytical Framework

In this research, I will analyze one of the obstacles which holds women back from entering public sphere to try to answer my research question of whether sexist language that is used in job advertisements affect women's labor force participation and job choice and prevents women applying to jobs which have more masculine wording. I will analyze job advertisement in Turkey and in Turkish language. The sexist language and words that are used in adds may affect women's labor force participation in Turkey. As the job options for women are very limited due to the bias in society towards women's place, these words targeting potential male employees may decrease the number of jobs that women can apply. Even though, choice for sex is not explicitly stated in the job advertisements, there are still some words which target institutions 'sex preferences using stereotypes or some traits that are associated particularly to certain sexes. In addition to keeping women back to enter public sphere, the fact that there is less women in traditional male-dominated jobs might be the sexist language that is used during job recruitment process.

In this sense, the literature part of this research will include the history of women's labor force participation and economic empowerment, particularly focusing on women's situation in Turkey by giving some official statistics from OECD[1] and TUIK/TURKSTAT[2] (Turkish Statistical Institute). In the beginning of the literature part, the definition of empowerment, women's unpaid work/second shift will be introduced. There will be some parts such as educational attainment of women, their preferences, and wives 'income will presented, since they are very related with women's labor force participation as being obstacles. This investigation will also shed light on possible improvements to increase women's labor force participation by reducing the sexist wording in job advertisements which is one of the crucial obstacles on women's participation and create more job opportunities for them. This study eventually points out specific directions for actionable policy recommendations. To draw a broad picture, experiences of women, particularly in Turkey and different case studies will be searched for the literature part to understand the issue one step ahead.

Research Questions

- Do Turkish job advertisements include words that target certain sexes?
- Do Turkish job advertisements include masculine wording more than feminine ones?
- Does sexist language that is used in job advertisements affect women's labor force participation and job choice and prevents women applying to jobs which have more masculine wording?

---

[1] https://www.oecd.org/turkey/
[2] https://www.tuik.gov.tr/Home/Index

Contribution to the Literature

This study will be a great contribution to the literature since it is the first study to analyze gendered language in Turkish job advertisement. There are many research focusing on obstacles in front of women in their way through the entrance of public sphere, however, this study will be the first one to analyze gendered wording in job advertisements. It is a treat for also educated women, since they eliminate women before applying the jobs by making them less appealing to women. The methodology will be used in this study is also different from the similar studies in the literature. Besides, there are limited studies based on the Turkish natural language processing, since Turkish is a difficult language to work with due to its morphology. NLP starts with morphological processing, and during that process researchers might face with many challenges for language and speech processing in Turkish due to its complex morphology and the interaction between syntax and morphology (Oflazer, & Saraçlar, 2018). This study will be a great extension of the literature of women's economic empowerment and labor force participation and essential to suggest policy to make conditions better for women employers.

Data Sets

- **Keyword corpus:** For sentiment analysis, a keyword corpus will be formed based on the words which identified either as feminine or masculine. For instances, "aktif", "hırslı", "atletik", "kendine güvenen", "karar verme", "güven", "cesaret" are the examples for words identified as masculine. Besides, there are some words identified as feminine such as "anlayışlı", "takım", "kendini adayan", "şefkatli", "bağımlı", "empati", "duygusal", "hassas", "dürüst", "sadık" and so on. For creating a words database which includes words that are considered to be sexed, the example studies will be examined and translated to Turkish in a proper way like the database of Roberts and Utych (2019) to analyze the link between gender language, and partisanship based on gendered words. The amount of the keywords sill be determined based on the precision and recall rate.

- **LinkedIn[3] API data set:** LinkedIn API enables researchers to work with job titles, advertisements, colleagues, and LinkedIn connections. This study will be mining advertisement data on LinkedIn based on traditional male and female dominated jobs.

Methodology

For this thesis project, Python-based data analysis pipeline will be employed. Sentiment analysis with natural language processing techniques as well as supervised machine learning methods will be performed. In the sentiment analysis, there are two essential approaches which are lexicon-based&linguistic and statistics&supervised machine learning (Turney, 2002). This study will focus on supervised machine learning method rather lexicon based approach, since SML is more successful in terms of its performance of learning the sample documents to predict the larger text documents (Gezici & Yanıkoğlu, 2018). Lexicon based approach is, an unsupervised machine learning method, used to measure opinion and subjectivity in a text document which extracts polarity and subjectivity from a text (Taboada et al., 2011).

Natural Language Processing (NLP) is a way to understand and manipulate text documents by using computational methods (Chowdhury, 2005). NLP aims to understand and produce human language content by employing computational techniques (Hirschberg et al., 2015). NLP will be used in this study, since it is possible for computers to understand text and determine the important parts. It is essential for this research, since the aim of this study is to create a classifier which can be used to detect sexist language in job advertisements in job recruitment websites such as LinkedIn [4].

[3] a social networking site focused on professional and business relationships: https://www.linkedin.com/feed/

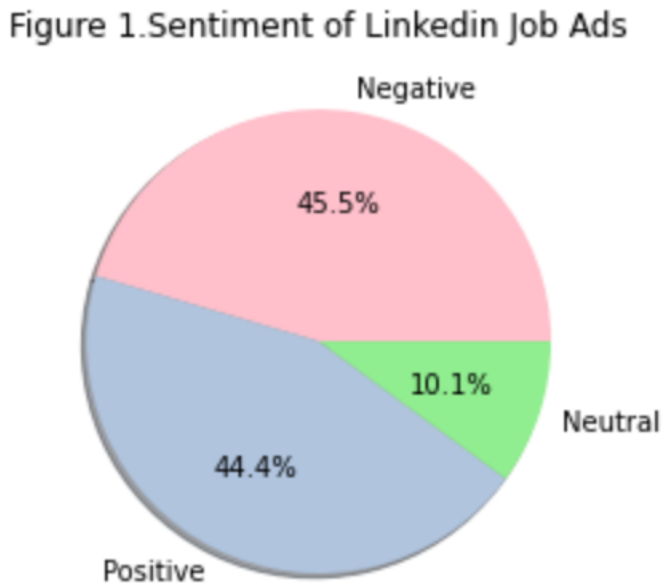
In this study, supervised machine learning method will be performed rather than unsupervised machine learning method. Machine learning helps to automate the analysis of the data by learning from the data and predict the outcomes. The difference between SML and UML methods is that SML observes an output for the each input and output creates a target for the algorithm for prediction, whereas UML only observes the input without a guidance (Molina & Garip, 2019). To be able to implement SML, the data set will be divided into three data sets which are training, test, and validation data. To fit the model, training data will be used, and to be able to decide which model to used in the study, validation data will be analyzed. For the purpose of preventing selected model from generalization error, test data will be used.

Data Collection

For the data collection part, I have started with collecting 100 job ads from LinkedIn. I have decided on 10 areas and collecting job ads for each of these branches based on the total jobs ads in LinkedIn for those jobs. I have created two groups, one is representing traditionally male-dominated jobs and the other one is representing traditional female-dominated jobs in Turkey. I have decided on 5 branches for each of them and collected the data based on these job titles. For traditionally male dominated jobs, I have generally focused on managerial jobs, since there are more men in leadership positioning jobs (Parlaktuna, 2010). For this reason, I have created 5 branches which are Engineer, Manager/Director/Head, Manager/Director/Head Assistant, Sales Representative/Specialist, and Controller/Inspector/Supervisor. I have searched for 20 job ads for Engineering, 10 for Manager/Director/Head, 5 for their assistants, 10 for Sales Representative/Specialist, and 5 for Controller/Inspector/Supervisor. On the other hand, I have chosen more subsidiary and secondary jobs for other categories, since women are generally concentrated on those jobs (Parlaktuna, 2010). For traditionally female dominated jobs, I have created branches which are Assistant, Secretary, Receptionist, Teacher/Teaching, and Flight Attendant/Ticket Sales Specialist. I have searched for 15 job ads for Secretary, 10 for Assistant, 10 for Teacher/Teaching, 10 for Receptionist, and 5 for Flight Attendant/Ticket Sales Specialist. With this way, I will be using this data set as a training data set for machine learning. To expand my data set, I have also created a code for web scrapping jobs ads from LinkedIn to be used as a test data set.

Data Labeling

To be able to create my data set, I have started by analyzing those job advertisements to discover the sentiment hidden within them. My data set includes 100 job advertisements collected from LinkedIn and also 3 types of classification which are positive, negative, and neutral based on their usage of sexist language. These adds tagged manually by making pair wise comparisons, if a job ad is sexist, it would be tagged as positive. After completing those processes, data set will be ready for the sentiment analysis part.



Preprocessing

Before starting the analysis, the data cleaned by droping the nan values. After this step, unnecessary columns removed to be able to focus on the columns that are wanted for the analysis. In the same direction, neutral tags are also removed, since they are not needed for the binary classification model. Since numerical data is needed for machine learning, labels, which are categorical, converted to numerical values using factorize() method. After this step, 0 represents positive which are male dominated ones, whereas 1 represents negative sentiment which are female dominated. For machine to be able to understand the data, text converted into array of vector embeddings. This helps to represent the relationship between words in the text. Before proceeding sentiment analysis, by using keras tokenizor, all the words have been broke down inro small parts called tokens and assigned numbers associated with words.

As a next step, all the text in category and requirements columns converted to lowercase, punctuations removed. In addition to those, by using Zemberek sentence normalization, removal of stopwords has applied. For the analysis, lemmatization and tokenization have used by using Zeyrek and implemented to the data set.

WordClouds for Traditionally Female and Male Dominated Jobs

To be able to confirm my keyword list, mostly-occurred keywords have checked in the data set. For this part, Word clouds have created for each group and certain words are analized.

Image 1. Mostly Occurred Keywords in Traditionally-Female-Dominated Job Ads

As it can be seen in Image 1., there are similar words for women applicants with the words identified as feminine in the literature. When employers are recruiting women applicants, they tend the use. Words like “güler yüzlü”, “bayan”, “özen gösteren”, “yatkın”, “iletişim”, “MS Office”, “eğitim”, “düzgün diksiyon”, “ekip çalışması”.

Image 2. Mostly Occurred Keywords in Traditionally-Male-Dominated Job Ads

In addition to those for women, similar words for men can be seen in the wordcloud as in the literature (Image 2). There are words such as “seyahat engeli olmayan”, “deneyim”, “güçlü”, “yüksek”, “sahip”, “tecrübeli”, “çalışmaya yatkın”, “iyi derecede”, “analitik”, and so on.

Sentiment Analysis

Sentiment analysis is a popular methodology which is used to analyze a piece of text to be able to discover the sentiment hidden within that text. Sentiment analysis allows scholars to examimne the expressed feeling in the text by combining natural language processing (NLP) with machine learning. In this Project, to classify the sentiment behind LinkedIn job ads, binary text classifier has been built. Various NLP preprocessing techniques has been implemented such as cleaning the data and utilizing LSTM layers to be able to build the text classifier. LSTM layers has used in the machine learning model. In order to avoid overfitting, Dropout mechanism has introduced between LSTM layers. Then, sentiment analysis model trained with the validaiton split 20%. To be able to execute the model, a function has defined which takes a job ad as an input and outputs its predicted label.

Results and Discussion

This study is a first step to understand the exclusion of women from labor force participation. As a first step, it is important to understand the feminine-coded and male-coded keywords lists in the literature and confirm that whether they are really identifying certain words. As the data set has been analyzed, it can be said that these keywords in the literatüre are really being used in the job advertisements. For that reason, this study is important in terms of confirming the literature, also to be able to create a new keyword list based on the sample for future analysis and have a more accurate results.

When the results of WordClouds have been examined, it can be seen that these words are already being used in the society to identify women. It can be said that society reflects these stereotypes in the workplaces, even during the recruitment process. This becomes an obstacle on women's way to the entrance to labor force and keeps them away from applying to jobs which includes more masculine-coded words.

Strengths of the Project

This project is unique in terms of the methodology it used from the similar studies in the literature. Also, this study is the first one which analyzes sexist language in Turkish job advertisements. One strength of the data set part is the annotation part. Labeling has been done manually by a knowledgable person in the are of genders studies and women's labor force participation. This makes sentiments more accurate, since human can capture all the words and hidden sentiment behind those job ads, whereas machines could miss come important points.It is important to anotate training data manually, particularly in this case, for increasing the accuracy of machine learning models and this study overall.

Challenges/Pitfalls and Ethical Concerns

During the data collection part, random selection process, some of the job ads in LinkedIn were in English language. That's why, these job ads are translated into Turkish by annotater. This could create a problem since word choice is really important for this study and translation could change the sentiment of the job ads regarding the change in word choice. One weakness of this study is that it needs futrher elaboration since the data set used in this research is really small to test the hypothesis. In addition to those, job givers in LinkedIn are more aware of the importance of language usage in job ads so that sexist language in those job ads are not explicit or targeting any certain gender. During the data collection part, since there are more magaerial and leading jobs in LinkedIn, it was hard to find job categories and job ads for women applicants. More men are concentrated in those jobs and tehre are very limited job ads for women in LinkedIn.

All the data used in this study does not create a problem concerning ethical issues since all the variables used are publicly available and do not contain any personal information. However, the biggest ethical concern of this study is the binary classification of women and men. It excludes non-binary people so that this may create a problem for some groups. However, this need to be first step to be taken to be able to understand the problem and suggest policies and regulations in a broader context in the future.

Future Implications

For the future studies, kariyer.net can also be used to extract data since it could give more insight about the recruitment environment in Turkey. Besides, the data set should be larger to be able to understand the problem as a whole and suggest regulations and policies to solve the issue. Finally, mixed methodology, perticularly, survey can be implemented to be able to understand women experiences during job searching and recruitment process better.

Bibliography

- Becker, G. S. (2009). Human capital: A theoretical and empirical analysis with special reference to education (3rd ed.). Chicago, IL: The University of Chicago Press.

- Chowdhury, G. G. (2005, January 31). *Natural language processing*. Association for Information Science & Technology. Retrieved January 18, 2022, from https:// asistdl.onlinelibrary.wiley.com/doi/full/10.1002/aris.1440370103? casa_token=aZe1Tffv2cAAAAAA%3AjxpFq4DEJ0oTyUO6fkSbUR7Ur0pecHi0d4dJ7uS2 lren ufwhNC8kqAKFgup0FVtexh8qxL1B9nxMrFk

- Crompton, R., & Lyonette, C. (n.d.). Mothers 'Employment, Work–Life Conflict, Careers and Class. *Women and Employment*. doi: 10.4337/9781848442931.00018

- Crompton, R., & Harris, F. (1999). Employment, careers, and families: The significance of choice and constraint in women's lives. In R. Crompton (Ed.), *Restructuring gender relations and employment: The decline of the male breadwinner* (pp. 128-149). New York: Oxford University Press.

- DiMaggio, P., Mohr, J., 1985. "Cultural Capital, Educational Attainment, and Marital Selection". The American Journal of Sociology, Vol. 90, No. 6, pp. 1231-1261. American Sociological Association

- Envision2030: 17 goals to transform the world for persons with disabilities Enable. (n.d.). Retrieved from https://www.un.org/development/desa/disabilities/envision2030.html.

- Ermiş, A. (2008). THE INCOMPATIBILITY OF TURKISH WOMEN'S EDUCATIONAL ATTAINMENT AND OCCUPATIONAL PARTICIPATION. *Hacettepe Universitesi*

Sosyolojik Araştırmalar, 1–16. Retrieved from http://www.sdergi.hacettepe.edu.tr/makaleler/a-ermis.pdf

- EVS/WVS (2021). European Values Study and World Values Survey: Joint EVS/WVS 2017-2021 Dataset (Joint EVS/WVS). JD Systems Institute & WVSA. Dataset Version 1.1.0, doi:10.14281/18241.14.

- Gezici G., Yanıkoğlu B. (2018) Sentiment Analysis in Turkish. In: Oflazer K., Saraçlar M. (eds) Turkish Natural Language Processing. Theory and Applications of Natural Language Processing. Springer, Cham. https://doi.org/10.1007/978-3-319-90165-7_12

- Gove W, Tudor J (1973) Adult sex roles and mental illness. Am J Sociol 78: 812–835

- Hakim, C. (2002). Lifestyle preference as determinants of women's differentiated labor market careers. Work and Occupations, 29, 428-459.

- Hirschberg, J., & Manning, C. D. (2015, July 17). *Advances in natural language processing*. Science. Retrieved January 15, 2022, from https://www.science.org/doi/abs/10.1126/science.aaa8685

- Hochschild, A. R., & Machung, A. (1997). The second shift: working parents and the revolution at home. New York, NY: Avon Books.

- İpek İlkkaracan (2012) Why so Few Women in the Labor Market in Turkey?, Feminist Economics, 18:1, 1-37, DOI: 10.1080/13545701.2011.649358

- *İstatistiklerle Kadın, 2020 - tuik.gov.tr*. (2021, March 5). Retrieved January 15, 2022, from https://data.tuik.gov.tr/Bulten/GetBultenPdf?id=37221

- Kan, M. Y. (2007). Work Orientation and Wives Employment Careers. *Work and Occupations*, *34*(4), 430–462. doi: 10.1177/0730888407307200

- Lazo, L. (1995). “Some Reflection on the Empowerment of Women”. In C. Medel-Añonuevo (Eds.pp:23- 37), Women Education and Empowerment: Pathways towards Autonomy. Hamburg: UNESCO Institute for Education.

- Maite Taboada, Julian Brooke, Milan Tofiloski, Kimberly Voll, Manfred Stede; Lexicon-Based Methods for Sentiment Analysis. *Computational Linguistics* 2011; 37 (2): 267–307. doi: https:// doi.org/10.1162/COLI_a_00049

- Molina, M., & Garip, F. (2019, May 13). *Machine Learning for Sociology*. Annual Reviews. Retrieved January 15, 2022, from https://www.annualreviews.org/doi/abs/10.1146/annurev-soc-073117-041106

- OECD (2022), Employment rate (indicator). doi: 10.1787/1de68a9b-en (Accessed on 18 January 2022)

- Oflazer K., Saraçlar M. (2018) Turkish and Its Challenges for Language and Speech Processing. In: Oflazer K., Saraçlar M. (eds) Turkish Natural Language Processing. Theory and Applications of Natural Language Processing. Springer, Cham. https://doi.org/ 10.1007/978-3-319-90165-7_1

- PARLAKTUNA, İ. (2010). Türkiye de Cinsiyete Dayalı mesleki Ayrımcılığın Analizi. *Ege Akademik Bakis (Ege Academic Review)*, *10* (4), 1217–1230. https://doi.org/10.21121/eab.2010419607

- Reskin, B. F., Hartmann, H. I.(eds.), 1986. "Women's Work, Men's Work: Sex Segregation on the Job". Committee on Women's Employment and Related Social Issues, National Research Council.

- Roberts, D. C., & Utych, S. M. (2020). Linking Gender, Language, and Partisanship: Developing a Database of Masculine and Feminine Words. Political Research Quarterly, 73(1), 40–50. https://doi.org/10.1177/1065912919874883

- Turney PD (2002) Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In: Proceedings of ACL, Philadelphia, PA, pp 417–424

- *Which jobs do men and women do? occupational breakdown by gender*. Careersmart. (n.d.). Retrieved June 11, 2022, from https://careersmart.org.uk/occupations/equality/which-jobs-do-men-and-women-do-occupational-breakdown-gender
- Yenilmez, M. I. (2017). Measuring And Analysing Women's Empowerment In Turkey. *Yönetim Ve Ekonomi Araştırmaları Dergisi*, 171–171. doi: 10.11611/yead.30681