A Linguistic Analysis on Gender Equality in The UN Assembly Speeches

Braakman, N.[15615162], Dowling, O.[15797708], Konstantinidou, C.[15469433], and Perdue, C.[15559777]

University of Amsterdam, Science Park, Amsterdam

Abstract. This study employs feature engineering techniques to analyze speeches delivered at the United Nations General Assembly from 2010 to 2023. This research aims to examine potential correlations between the linguistic patterns of these speeches using keywords for gender equality and various socioeconomic indices, including measures of life satisfaction, economic development, and social progress. We hypothesize that countries that use gender equality words in speeches have higher life happiness scores. Our methodology involved processing speech transcripts, extracting keywords, testing for correlation and using classification models. Contrary to the initial hypothesis, our findings reveal that the language used in UN Assembly speeches does not serve as a reliable predictor of a country's socioeconomic status or well-being indices. The model performed reasonably well and was able to predict a trend across different periods and geographic regions.

Keywords: Text Analysis · United Nations Speeches · Gender Equality · Data Manipulation · Machine Learning.

1 Introduction

Gender equality is a basic human right and crucial for building peaceful societies and sustainable development. The United Nations (UN) recognizes that increasing gender equality can boost economic growth and increase productivity. Despite this, significant inequalities remain, with issues like gender-based violence, unequal access to education, jobs, and leadership positions still posing challenges worldwide. Through programs like UN Women and the Sustainable Development Goals (SDGs), the UN aims to reduce these gaps and ensure equal rights and opportunities for everyone, which is important for creating stronger economies and a healthier planet[8].

Values play an important role in shaping the content of UN speeches, as they represent the goals that guide individuals and nations in evaluating actions, policies, people, and events[9]. These speeches are key reflections of a country's priorities and policy stance. According to Baturo et al.[1], UN speeches also provide valuable insights into the perspectives and preferences of governments.
By analyzing the speeches of UN member states, we can better understand their commitment to gender equality and how it connects to their broader socioeconomic situations. This research will focus on two main questions:

– Is there a link between how often gender equality is mentioned in political speeches and socioeconomic factors such as Gross Domestic Product (GDP), happiness, social support, and freedom to make life choices? (Exploratory question, SDG 5). The hypothesis that goes along with this is that countries that use more words in their UN speeches relating to gender equality have a higher score in happiness and socioeconomic factors.
– Can we predict a trend in future mentions of gender equality in political speeches based on patterns from previous years? (Predictive question, SDG 5)

Focusing on gender equality has been shown to bring social and economic advantages. Research shows that gender equality is associated with stronger economic growth, higher productivity, and more social stability. This suggests that countries that emphasize gender equality in their UN speeches may also experience higher levels of development and well-being[10].

The findings from this research will help us understand how international discussions on gender equality reflect global and national priorities, and whether countries that focus more on gender equality in their international speeches also show stronger socioeconomic outcomes at home.

2 Methodology

This research uses three datasets to answer these questions. The first dataset is focused on collecting speeches from the UN General Assembly. This dataset lists each speech, from each participating country, from each assembly dating back to 1946. For this particular project, the speeches from 2010 - 2023 are used. The second dataset used is from the Happiness World Report (2024)[3] from which we used the chosen socioeconomic factors, log GDP per capita, Life Happiness, Social Support and Freedom to make life choices. The Happiness dataset is merged with the UN speeches in an attempt to investigate the relationship between word frequencies related to gender equality and these socioeconomic factors. The model dataset mapped the speeches with their corresponding country and the year that the speech was made. The initial UN speech data frame referred to each assembly in session numbers (1 - 78), using data from the third dataset 'UNSD Methodology' to link each speech to their corresponding country and year.

2.1 Data Manipulation

The pre-processing stage begins by importing the files and converting them to data frames using the Pandas library. According to McKinney, Pandas' sophisticated index functionality makes it easy to reshape, slice and dice, perform aggregations, and select subsets of data. Since data manipulation, preparation and cleaning are important for analysis, Pandas is necessary for this research[5]. As mentioned above, each data frame was merged to create one data frame for exploratory data analysis and another for statistical learning techniques. To do this, each file of data must be examined for categorical and numerical data, missing data, redundant information, etc. When dealing with missing data, replacing the information was done by using mean values where applicable. Any rows where the mean value could not be used were dropped as this would lead to issues later when trying to fit models to the data. There were various columns in all three datasets that either repeated information or had redundant information, these columns were removed during the pre-processing phase.

The goal of this research is to identify links between the frequency of gender equality words in UN speeches and the socioeconomic factors: GDP, happiness, social support and freedom to make life choices. The first step is to identify which keywords should be searched for within the speeches. Pulling inspiration from the University of Toronto - Sustainable Development Goals (SDGs), keywords such as "gender", "women", "transgender", "lgbtq" were selected [7]. The speeches for every country between the years 2010 - 2023 were scanned for these keywords and a count was made of every instance per speech. This data was then attached to the exploratory data analysis (EDA) data frame under the column heading "Frequency".

The next step involves calculating the annual frequency difference for each country within the dataset. This difference is added to a new column labeled "Frequency Difference" facilitating the identification of trends over time. For the first year represented in the dataset, the value is set to 0, as there is no prior year to compare against, meaning any resulting NaN values are filled with 0. Following this, another new column is added named "Trend" which is assigned a value of 1 for the years with a positive frequency difference (indicating an upward trend) and 0 for the years with a negative frequency difference (indicating a downward trend). This "Trend" column serves as the target variable for our prediction.

2.2 Machine Learning

During the machine learning phase of the project, it is vital to separate the dataset into training and test sets, with the data from recent years reserved as the validation set. It is important that validation data is preserved and not fed into the model during the training period. The model that is used must be able to take in data that it has never been exposed to and make a prediction. The model used in this project was fitted to data from 2010 - 2020 and data from 2023 was preserved as a validation set. It should be noted that the years in which the COVID-19 pandemic took place were not used as the data within the speeches were suspected of being dominated by this topic. Therefore, the years 2021 and 2022 were not included.

After establishing the initial split, the training data is further divided into features and the target variable. Specifically, feature set X is created by dropping the columns "Country name" "year" and "Trend" while the target variable Y is defined as the "Trend" column. The training data was split into training and testing subsets using a 70-30 split, allowing fine-tuning of the model before evaluating its predictive accuracy on the reserved test set from 2023.

After analyzing the relationship between "Frequency" and socioeconomic factors, the output suggested a low predictive power or weak correlation. Consequently, we shifted from our initial regression approach to a classification method to more effectively predict trends in frequency over time.

We expanded our approach by creating a function to find the most suitable model for our data. This function uses the models and parameters for each model defined and automates the process of finding the best combination of the two by running a grid search for each model with the specified parameter grids. It splits the data, fits the models, and records the best-performing parameters and metrics for each.

To optimize model selection, we tested various classification models, including Logistic Regression, SGD Classifier, AdaBoost Classifier, Random Forest Classifier, and Decision Tree Classifier. For each model, we performed hyperparameter tuning using grid search with cross-validation to identify the best configuration. This process involved training the models on the training dataset and evaluating their performance based on key classification metrics such as accuracy, F1 score, precision, and recall.

3 Results

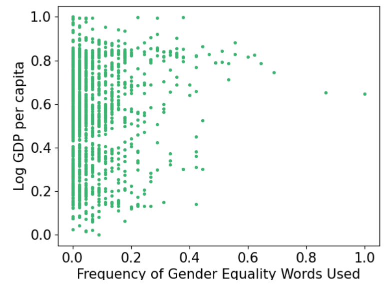
3.1 Results for Exploratory Analysis

To determine whether or not there is a correlation between the gender equality word frequencies and the socioeconomic factors: Life Ladder, log GDP per capita, social support, and freedom to make life choices, we created four scatter plots as seen in Figure 1.

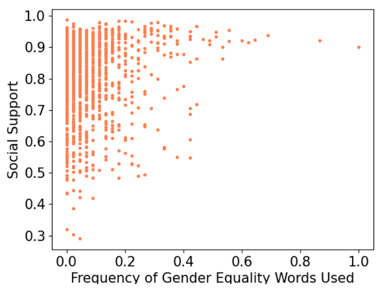
Across all of these socioeconomic factors, the scatter plots suggest weak or negligible correlations with the frequency of gender equality words used. In all plots, most data points are clustered around lower word frequencies (around 0-0.2), with a wide variation in the socioeconomic outcomes. This indicates that, based on this dataset, the frequency of gender equality discourse in UN General Assembly speeches is not a strong predictor of these socioeconomic factors.

The results from these plots can be backed up by Spearman's Rank Correlation. The Spearman correlation assesses how well the relationship between two variables can be described by a monotonic function. A monotonic association is one where, as the value of one variable increases, so also does the value of the other, or as the value of one variable increases the other variable decreases[6]. For the results of the Spearman test the values for the correlation coefficient ranged from 0.11 - 0.15 with extremely low p-values which can be seen in Table 1.

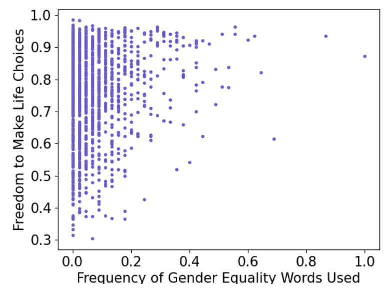
(a) Correlation with the Life Ladder



(b) Correlation with log GDP per capita



(c) Correlation with Social Support



(d) Correlation with Freedom of Choice

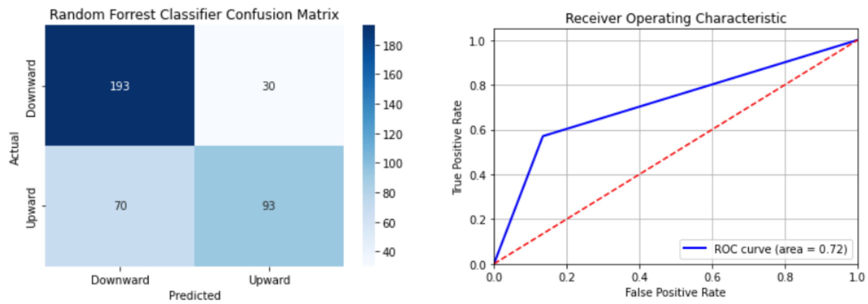
Fig. 1: Scatterplots showing the correlation between gender equality word frequencies and various socioeconomic factors

Socioeconomic factor	**Correlation Coefficiant**	**p-value**
Life Ladder	0.14	5.8e-10
log GDP per capita	0.11	7.3e-07
Social Support	0.11	2.0e-06
Freedom to make life choices	0.15	6.4e-12

Table 1: Spearman's correlation between the different socioeconomic factors and the gender equality word frequencies.

3.2 Results for Predictive Analysis

Once the models had been fitted with training data, getting scores from the models based on the validation set was imperative to understand the issues with the data. Using the "Trend" column, the data was also used to predict a categorical fit. The best model calculated from this was a RandomForestClassifier with parameters max depth=3, n estimators=100, and random state=42. Model accuracy, F1 score, precision and recall were all calculated at 74.9, 69.2, 71.7 and 66.9 percent respectively.



(a) RFC Confusion Matrix (b) AUC score and plot

Fig. 2: Confusion matrix, AUC score and plot, Model selection results

To further evaluate the model, the confusion matrix in Figure 2(a) was created based on predictions made for the validation dataset. This matrix revealed the distribution of true positives (correctly predicted upward trends), true negatives (correctly predicted downward trends), as well as false positives and false negatives. It highlighted that while the model made fewer false-positive errors (high precision), it occasionally missed some true positive trends, reflected in the lower recall. These insights allowed us to assess the model's performance on unseen data and better understand the trade-offs between precision and recall, leading to a more thorough evaluation of its predictive power[4].

Figure 2(b) illustrates the Receiver Operating Characteristic (ROC) curve and the Area Under the Curve (AUC) score, which are critical for evaluating the performance of any classification model. The blue line represents our model's performance, while the red dashed line shows a random classifier (AUC = 0.5) for comparison. Our model's AUC score indicates a 0.72 chance of correctly distinguishing between positive and negative instances, a performance better than random guessing. Visualizing this curve was essential for our analysis as it provided a clear understanding of the model's ability to differentiate between classes.

4 Discussion

4.1 Interpretation of the findings

From both the visual analysis of the scatter plots in Figure 1 and the results from the Spearman's Rank Correlation, we can conclude that there is a significantly weak positive correlation. This is confirmed since the p-values for each of the socioeconomic factors and the gender equality word frequencies are extremely small. This indicates that the frequency of gender equality words in a UN speech is not predictive of a country's overall happiness.

Ultimately, adopting a classification approach provided a more effective solution for addressing the research question. Instead of attempting to predict the exact frequency, an approach that proved unreliable, the classification method enabled us to forecast whether the trend would increase or decrease in future years. This shift in methodology offered clearer and more actionable insights, which were meaningful given the nature of the dataset.

The analysis of the confusion matrix revealed that the model achieved high precision, correctly predicting upward trends in many cases, while occasionally missing true positive trends (lower recall). The AUC score of 0.72, as shown in the ROC curve, indicates that the model performs significantly better than random guessing, with a 72% probability of correctly distinguishing between positive and negative trends. While the model demonstrates decent predictive capability, there is room for improvement, particularly in balancing precision and recall to capture a higher number of true positive trends.

These results suggest that, while it is possible to predict future trends in gender equality mentions with some accuracy, additional refinement of the model is necessary to enhance its reliability and predictive power for broader use.

4.2 Implications, Limitations and Future work

During the data analysis, we encountered a challenge where the initial models we tested did not perform as expected. The question this project aimed to answer was considered to be linear as the output of the model would be continuous data. However, upon visual inspection of the scatter plots in figure 1, it was clear that the frequency of words had little to no correlation with the socioeconomic factors. Despite the results from the scatter plots, the data was passed through a linear regression model. Unfortunately, this still led to poor results which showed that the "Frequency" column did not have much predictive power on a country's happiness score. To address this, we switched to a categorical approach and applied a grid search technique to identify the model that best fits our data.

One limitation of the study is that the World Happiness data is only available starting in 2008, whereas the UN speech dataset begins in 1946. As a result, while examining correlations between speech frequencies and socioeconomic factors, we were unable to include the years 1946-2007, potentially limiting the scope of the analysis.

Additionally, although gender equality-related words are detected in the UN speeches, the Natural Language Toolkit (NLTK) used for this analysis does not provide insight into the overall theme or sentiment of the speeches. Gender equality terms represent only a small fraction of the total word count. Additionally, since each speech covered multiple topics, we were unable to determine the specific theme related to gender equality within the broader context of the speeches.

When examining speeches from 2010 to 2020 for the frequency of gender equality terms, the results were notably low. A review of the speeches revealed that topics such as war and peace were more frequently discussed than gender equality. Had these assemblies placed greater emphasis on gender equality, more substantial data could have been gathered to support a stronger research project. The lack of attention to this issue in the assemblies hindered the research presented in this report.

For future studies, one could improve the robustness of our model using a stratified sampling approach based on the "Life Ladder" column. By categorizing the data into distinct ranges—lower, mid, and upper happiness levels—before splitting it, we can ensure a more balanced representation of each segment. This method can prevent the randomness in our data split from favoring one particular group, such as an over-representation of countries with lower happiness scores. Stratified sampling can improve model performance by ensuring that all segments of the data are adequately represented, thereby leading to more reliable results.

By assessing the ROC and AUC score in figure 2(b) one could take into account the curve's shape and proximity to the top-left corner, and pinpoint areas where the model performs well and identify opportunities for improvement [4].

5 Conclusion

For the explanatory research question we explored whether there is a link between the frequency of gender equality mentioned in UN speeches and the socioeconomic factors Gross Domestic Product (GDP), life ladder, social support, and freedom to make life choices. Our analysis indicates that while some weak correlations exist, the overall relationships are minimal. This suggests that the frequency of gender equality mentioned in speeches does not have a strong or consistent link to these socioeconomic factors, contrary to expectations based on existing literature highlighting, the positive impact of gender equality on economic and social outcomes.

Additionally, for the predictive research question we investigated whether future trends in gender equality mentioned in UN speeches can be predicted based on patterns from previous years. The results indicate that the model can moderately differentiate between upward and downward trends, showing some potential for forecasting. However, while the model demonstrated high precision in identifying trends, it occasionally missed certain upward movements, highlighting a need for improvement in recall. Overall, the predictive capability is promising but not fully reliable, suggesting that while patterns from past years provide some insight, additional refinement is needed to achieve more accurate and consistent predictions. However, this research should encourage possibilities on further investigations into the topic of gender equality within the UN Assembly.

Bibliography

[1] Baturo, A., Dasandi, N., & Mikhaylov, S. J. (2017). Understanding state preferences with text as data: Introducing the UN General Debate corpus. Research & Politics, 4(2), 205316801771282. https://doi.org/10.1177/2053168017712821

[2] European Institute for Gender Equality. (2023). Gender equality index 2023. https://eige.europa.eu/gender-equality-index/2023/ES

[3] Helliwell, J. F., Layard, R., & Sachs, J. (Eds.). (2024). *World happiness report 2024*. Sustainable Development Solutions Network. https://worldhappiness.report/ed/2024/#appendices-and-data

[4] Erickson, B. J., & Kitamura, F. (2021). Magician's corner: 9. Performance metrics for machine learning models. Radiology Artificial Intelligence, 3(3), e200126. https://doi.org/10.1148/ryai.2021200126

[5] McKinney, W. (2017). Python for data analysis: Data wrangling with Pandas, NumPy, and IPython. O'Reilly Media.

[6] Sedgwick, P. (2014). Spearman's rank correlation coefficient. BMJ, g7327. https://doi.org/10.1136/bmj.g7327

[7] The University of Toronto. (2022). Sustainability development goals (SDGs) keywords. https://sustainability.utoronto.ca/inventories/sustainable-development-goals-sdgs-keywords/

[8] United Nations. (z.d.). Gender Equality — United Nations. https://www.un.org/en/global-issues/gender-equality

[9] Waheed, M., Schuck, A., Neijens, P., & De Vreese, C. (2013). Values in UN speeches. International Communication Gazette, 75(7), 597–618. https://doi.org/10.1177/1748048513482542

[10] World Bank. (2012). World development report 2012: Gender equality and development. World Bank Group. https://doi.org/10.1596/978-0-8213-8810-5