Using Word Embeddings to Quantify Ethnic Stereotypes in 12 years of Spanish News

Universitat Pompeu Fabra Barcelona

Danielly Sorato, Diana Zavala-Rojas, Carme Colominas Ventura

Universitat Pompeu Fabra

Abstract

The current study provides a diachronic analysis of the stereotypical portrayals concerning seven of the most prominent foreign nationalities living in Spain in a Spanish news outlet. We use 12 years (2007–2018) of news articles to train word embedding models to quantify the association of such outgroups with drug use, prostitution, crimes, and poverty concepts. Then, we investigate the effects of sociopolitical variables on the computed bias series, such as the outgroup size in the host country and the rate of the population receiving unemployment benefits. Our findings indicate that the texts exhibit bias against foreign-born people, especially in the case of outgroups for which the country of origin has a lower Gross Domestic Product per capita (PPP) than Spain.

Introduction

- Languages are complex and systematic instruments of communication that reflect the culture of a given population
- By studying language, it is also possible to observe **stereotypes**
  - a type of **social bias** that is present when discourse about a given group **overlooks the diversity of its members and focuses only on a small set of features** [1, 2]
- Like society, **languages are not static**
  - Extra-linguistic factors can give insights into the dynamics of **social, cultural, and political phenomena reflected in texts** [3]
- Studying language over time implies on vast amounts of data to analyze
  - For instance, 1,757,331 news articles covering 12 years of Spanish news
  - **Efficient computational methods** for performing diachronic analysis play a crucial role
  - Literature shows that **word embeddings** models are helpful tools to this end

- In this work, we analyze the **dynamics of stereotypical associations concerning seven outgroup nationalities**
  - Some of the **most representative foreign nationalities** living in Spain
  - Namely, **British, Colombian, Ecuadorian, German, Italian, Moroccan, and Romanian**
- Word embedding models trained with news articles published in the Spanish newspaper *20 Minutos* for the period of **2007 to 2018**
- Fine-grained analysis studying the association of such nationalities with **drug use, prostitution, crimes, and poverty concepts**
- Measured associations are compared with sociopolitical variables
  - **Survey items from the European Social Survey (ESS)**
  - **Number of residents** by nationality living in Spain
  - Rate of the population receiving **unemployment benefits** from the Spanish government
  - **Number of offenses** committed in Spain by outgroup background
  - Additionally, we investigate the effect of the outgroups' countries of origin having a **lower Gross Domestic Product per capita (PPP) than the host country (Spain)**
- To account for both group effects and error correlation, we use multilevel **Random Effects (RE) models**

Methodology

- Yearly Fasttext models trained with the *Corpus 20 minutos* data from **2007–2018**, comprising **1,757,331 news articles**
- Prior to model training, we lowercased words, removed punctuation and numbers
- Only words that appeared at least 15 times in each yearly dataset were taken into account in the training phase
- Resulting vectors were L2 normalized
- Model quality evaluated using two (translated to Spanish) word similarity benchmarks, namely RG-65 and MC-30
- Associations between words can be measured using cosine similarity (or euclidean distance)
- Word lists were defined to represent the ingroup, the outgroups (i.e. the seven nationalities), and drug use, prostitution, crimes, and poverty concepts
  - The vector representations of those words were then averaged

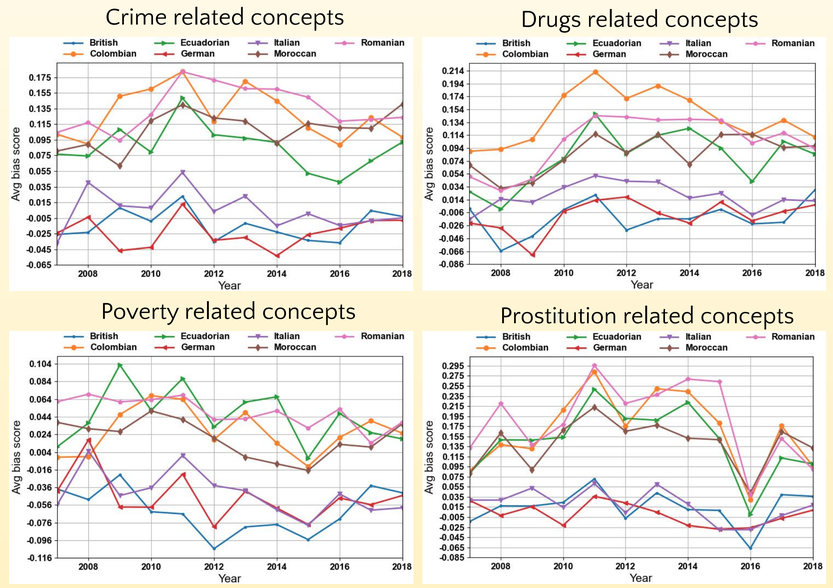
- The strength of association between the outgroups and the ingroups was measured using the following metric [4]

$$bias\ score = \sum_{v_{word} \in WordList} cos(v_{word}, v_{outgroup}) - cos(v_{word}, v_{spanish})$$

- The more positive the bias score is, the more associated it is with the outgroup, the more negative, the more associated with the ingroup
- Random Effects models were fitted with the following predictors, where the bias score is the dependent variable
  - Number of residents per nationality in Spain
  - Rates of population receiving unemployment benefits
  - Number of offenses committed in the Spanish territory by outgroup
  - Public opinion, i.e. European Social Survey questions about immigration
  - Lower Gross Domestic Product per capita (PPP) (**Colombian, Ecuadorian, Moroccan, Romanian**) or higher PPP (**British, German, Italian**) than the host country
  - Year trend

Results

- **Colombian, Ecuadorian, Moroccan and Romanian outgroups show appreciably stronger association across all concepts**
- **Strong and significant effect ($p < .001$) of Lower PPP predictor** in all cases
- For all concepts: when the rate of people receiving unemployment benefits increases the bias also increases **but only for Colombian, Ecuadorian, Moroccan and Romanian outgroups (Lower PPP)**
- For drug use related concepts: at a lower significance level (**$p < .05$**), when the number of offences increases the bias also increases **but only Lower PPP group**



Conclusion

- Interpretation of main effects and interactions with sociopolitical variables indicates that **stereotypical portrayals seem to be dissociated from real demographic trends**
- The strong effect of the Lower PPP predictor on our analysis that news discourse emphasises the ethnicity of certain outgroups more than others
- **Discourse is one of the everyday social practices that may be used for discriminatory purposes, for instance in intra-group discourse about resident minorities or immigrants frame these "others" negatively, thus leading to the reproduction of ethnic prejudices or ideologie**
- Problem can be **further propagated and amplified** through computational algorithms if available data indiscriminately leading to concerning outcomes [6, 7]
  - Downstream tasks that use biases models underneath can lead to biased results

Acknowledgements

This work was partially supported by the Universitat Pompeu Fabra (UPF), through a PhD studentship grant and the use of Marvin Cluster to train and analyse the language models.

References

[1] Javier Sánchez-Junquera, Berta Chulvi, Paolo Rosso, and Simone Paolo Ponzetto. 2021. How do you speak about immigrants? taxonomy and stereoimmigrants dataset for identifying stereotypes about immigrants. Applied Sciences, 11(8):3610.

[2] Henri Tajfel, Anees A Sheikh, and Robert Charles Gardner. 1964. Content of stereotypes and the inference of similarity between members of stereotyped groups. Acta Psychologica.

[3] Anna Marakasova and Julia Neidhardt. 2020. Short-term semantic shifts and their relation to frequency change. In Proceedings of the Probability and Meaning Conference (PaM 2020), pages 146–153.

[4] Nikhil Garg, Londa Schiebinger, Dan Jurafsky, and James Zou. 2018. Word embeddings quantify 100 years of gender and ethnic stereotypes. Proceedings of the National Academy of Sciences, 115(16):E3635–E3644.

[5] Teun A Van Dijk. 2000. On the analysis of parliamentary debates on immigration. Citeseer.

[6] Tolga Bolukbasi, Kai-Wei Chang, James Y Zou, Venkatesh Saligrama, and Adam T Kalai. 2016b. Man is to computer programmer as woman is to homemaker? debiasing word embeddings. In Advances in neural information processing systems, pages 4349–4357

[7]Moin Nadeem, Anna Bethke, and Siva Reddy. 2020. Stereoset: Measuring stereotypical bias in pretrained language models. arXiv preprint arXiv:2004.09456

Universitat Pompeu Fabra Barcelona