

# Using Word Embeddings for Immigrant and Refugee Stereotype Quantification in a Diachronic and Multilingual Setting

Danielly Sorato<sup>1</sup>

<sup>1</sup>*Research and Expertise Centre for Survey Methodology, Universitat Pompeu Fabra, Barcelona, Spain*

## Abstract

Languages are complex and systematic instruments of communication that reflect the culture of a given population. Amongst the many phenomena that can be observed by studying language, there are the social biases, such as stereotypes. The use of stereotypical framing in discourse can be very detrimental, especially when used by media and politicians, which are often responsible for distortions regarding the outgroup's (e.g., immigrants, refugees) image inside the country. Such distortions can foster fear and encourage hate-motivated attitudes, leading to problematic outcomes. This paper describes our framework to quantify stereotypical associations concerning immigrants and refugees in public discourse, using a multilingual and diachronic setting. We present our research design and methodology concerning an experiment with a multilingual corpus of parliament texts, for the period of 1996 to 2018.

## Keywords

Word embeddings, Diachronic analysis, Multilingual analysis, Computational sociolinguistics

## 1. Introduction

Stereotype is 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], which can be observed by studying language. However, like society, languages are not static, by analyzing language over time, it is possible to gain insights into the dynamics of social, cultural, and political phenomena reflected in texts [3], such as negative stereotypes of immigrant groups.

Alongside the growing levels of immigration inflows experienced in European countries in recent decades, the increasing negative framing of immigrants and refugees in public discourse have become a major concern [4, 5, 6, 7, 8, 9]. The media and politicians or key social actors are often responsible for distortions regarding the ingroup's perceptions and attitudes towards outgroups inside the countries [10, 11, 5, 12]. Such distortions can foster fear and encourage anti-immigration attitudes, leading to problematic outcomes. The misperceptions concerning immigrant populations is especially timely and relevant, having played a major role in important political events, such as the Brexit and in the increase of support of extreme right-wing political parties and rising nationalism in Europe [11, 13, 14, 15].

---

*Doctoral Symposium on Natural Language Processing from the PLN.net network 2022 (RED2018-102418-T), 21-23 September 2022, A Coruña, Spain.*

✉ danielly.sorato@upf.edu (D. Sorato)

ORCID 0000-0002-4691-7231 (D. Sorato)



© 2022 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).



CEUR Workshop Proceedings (CEUR-WS.org)

Nonetheless, manually analyzing texts spanning several years of public discourse is unfeasible due to the large amount of data involved. As such, computational methods for diachronic linguistic analysis are playing a crucial role, and ongoing research shows that word embeddings models are helpful tools to this end, since they contain machine-learned biases in their geometry that closely depict social stereotypes [16, 17, 18, 5, 19]. Although such models should be carefully tested for biases and not blindly applied to downstream computational applications due to ethically concerning outcomes [20, 21, 22], they can be a valuable tool for sociolinguistic analysis on large volumes of textual data.

In past work conducted in this PhD, we analyzed the dynamics of stereotypical associations towards seven of the most prominent ethnic groups living in Spain (British, Colombian, Ecuadorian, German, Italian, Moroccan, and Romanian) in the period of 2007 to 2018 using word embedding models trained with news items from the Spanish newspaper *20 Minutos* [23]. We investigated biases concerning concepts related to crimes, drugs, poverty, and prostitution, exploring the relation between the stereotypical associations and sociopolitical variables (e.g., GDP per capita (PPP) of the groups' countries of origin, unemployment rates). The interpretation of main effects and interactions with sociopolitical predictors in our multilevel modelling approach pointed that the texts exhibit stereotypical associations, especially for the Colombian, Ecuadorian, Moroccan and Romanian groups.

In our ongoing research, we extend our study to a multilingual setting and a different domain: political discourse. Our goal is to quantify and compare the strength of stereotypical associations towards immigrants and refugees in the period of 1998 to 2018 concerning concepts such as crimes, poverty, and trafficking in the discourse of British, Danish, Dutch, and Spanish parliaments. Moreover, other than analyzing the stereotypes through the geometries of vector spaces, we aim to examine the effects of sociopolitical variables (e.g., immigration inflows, criminality rates) in the stereotypical association time-series using a Bayesian multilevel modelling approach. Finally, we aim to understand the different ways that the bias manifests itself in the vector spaces of different types of embeddings, such as static versus contextual embeddings, and the use of words versus sentence embeddings.

This paper is organized as follows. In Section 2 we discuss related works. Subsequently, in Section 3 we state our research questions, in Section 4 we present metrics, data, model training, and evaluation. Finally, in Section 5 we present our proposed discussion points.

## 2. Related Work

Human generated data is full of both intentional and non-intentional stereotypes. However, there are certain types of stereotypes that impose special difficulties, since they can be subtle and often do not rely on personality traits (e.g., honest, empathetic), such as the case of stereotypes about immigrants [1]. In this context, word embeddings showed as a valuable tool, by means of enabling efficient methods for analyzing and quantifying linguistic and social phenomena in natural language.

Overall, most works concerning the study of machine learned biases have English as target language, or approach exclusively gender bias [18, 24, 17, 16, 25, 26, 27, 20]. Nonetheless, biases can exist in all human languages, as well as in many shapes and forms, which calls for the

conduction of research using other target languages and biases.

Wevers quantified gender biases in 40 years of news published in six Dutch newspapers. Tripodi et al. investigated the antisemitism in public discourse in France, by using diachronic word embeddings trained on a large corpus of French books and periodicals containing keywords related to Jews. Sánchez-Junquera et al. detected stereotypes towards immigrants in political discourse by focusing on the frames used by political actors. They created their own taxonomy to capture immigrant stereotype dimensions and produced an annotated dataset with sentences that Spanish politicians have stated in the Congress of Deputies, which was then used to train classifiers to detect stereotypes. Kroon et al. quantified the dynamics of stereotypical associations concerning several outgroups in 11 years of Dutch news data, focusing on the difference of such associations regarding the group membership (ingroup vs outgroups). Lauscher et al. conducted an analysis about racism and sexism related biases in Arabic word embeddings across different types of embedding models and texts (e.g., user-generated content, news), dialects, and time.

The literature concerning bias detection in multilingual settings is still scarce and recent, as such scenario imposes greater challenges than monolingual ones. Câmara et al. quantified gender, racial, ethnic, and intersectional social biases across five models trained on sentiment analysis tasks in English, Spanish, and Arabic. Ahn and Oh verified the existence of ethnic biases in monolingual BERT models for English, German, Spanish, Korean, Turkish, and Chinese, while proposing a new multi-class bias measure to quantify the degree of ethnic bias in such language models. Further, they proposed two bias mitigation methods using multilingual and word alignment approaches. Névéal et al. contributed to the analysis of multilingual stereotypes by creating an English and French dataset<sup>1</sup> that enables the comparison across such languages, while also characterizing biases that are specific to each country (United States and France) and language. Their dataset includes biases types such as ethnic, gender, sexual orientation, nationality, age, among others. Such dataset was then used to verify stereotypes in three French and one multilingual language models.

Our study distinguishes itself from the aforementioned studies by (i) the interdisciplinarity with social survey research, as the selected survey questions measure attitudes of the ingroup towards immigrants and can be interpreted as a proxy for cultural/economic threat perception; (ii) our choice of multilevel modeling to combine types of phenomena (linguistic and social) and account for group effects; and (iii) the use of fine-grained lists to investigate stereotypical portrayals (e.g., concepts related to poverty, drugs, human trafficking, prostitution). Additionally, we contribute to the scarce literature on stereotypical bias analysis with non-English data sources (Danish, Dutch, and Spanish) and multilingual settings.

### 3. Research Questions

For the multilingual setting of this research, our main objective is to quantify and compare the strength of association between immigrants and refugees and stereotypical concepts (e.g., crimes, unemployment, poverty) in the discourse of Danish, Dutch, British, and Spanish parliaments across time (1996-2018). We intent to analyze the vector spaces of different embedding techniques, e.g., static versus contextual embeddings, words versus sentence embeddings. Fi-

---

<sup>1</sup><https://gitlab.inria.fr/french-crows-pairs/acl-2022-paper-data-and-code>

nally, we aim to examine the effect of sociopolitical indicators that are relevant to the context of attitudes towards immigrants on our trends, with the objective of verifying if demographic trends correlate with our reported stereotypical associations.

We achieve the aforementioned objectives by seeking the answers to the hereby stated research questions:

- Can we track stereotypes about immigrants and refugees in political data across time using different embedding techniques?
- How can we systematically compare biases in the vector spaces of different embedding techniques?
- Can we compare and find patterns in the stereotypical association time-series for different languages?
- Can we inspect effect of country-specific sociopolitical variables (e.g., immigration inflows, public opinion measured by survey, criminality rates) on computed time-series?

## 4. Methodology

This thesis revolves around the study the dynamics of the stereotypical associations concerning outgroups in European public discourse (e.g., news, political speech) over time using embedding models. In previous work, we studied the stereotypical associations towards British, Colombian, Ecuadorian, German, Italian, Moroccan and Romanian nationalities using static embeddings (Fasttext implementation [32]) trained in the news domain, considering the years 2007 up to 2018 in our analysis.

In our current setup, we adopt a multilingual perspective, a different domain and time span: parliamentary speeches covering the years 1996 up to 2018. In addition, we analyze stereotypical associations towards immigrant and refugees, rather than specific nationalities. In order to compute the association trends over time in this new setting, we start by training static language-specific skip-gram embedding models using our target corpora. To answer our research questions, we adopt the following data, metrics, and models.

### 4.1. Data

For our monolingual case study, we compiled the Corpus of Spanish news *20 Minutos* [33]. The corpus contains news articles written in Spanish from Spain that were web-scraped from the newspaper’s website *20 Minutos*<sup>2</sup>. Such dataset was split by year, allowing us to train 12 yearly word embedding models.

To train embedding models in our multilingual setup, we combine the Danish, Dutch, English and Spanish portions of the following parliamentary corpora:

- *Europarl* [34] (release 7);
- *Parlspeech V2* [35];
- *ParlaMint* [36];

---

<sup>2</sup><https://www.20minutos.es/>

- *IM-PRESS/PRESS, Written Question, Written Question Answer, Oral Question and Questions for Question Time* portions<sup>3</sup> of the *Digital Corpus of the European Parliament (DCEP)* [37].

Like it was done in the monolingual study, we split our final language-specific datasets by year to then train the embedding models (4 languages  $\times$  23 years = 92 models).

#### 4.1.1. Sociopolitical data

The sociopolitical variables for our monolingual study were taken from the Instituto Nacional de Estadística (INE)<sup>4</sup> and the European Social Survey (ESS)<sup>5</sup>. We used as indicators the number of foreign population by nationality residing in Spain, the rate of the population receiving unemployment social benefits, the public opinion about immigration using survey questions from the ESS, and number of committed offenses.

In our ongoing research, we will use country-specific sociopolitical indicators from the Eurostat<sup>6</sup> (e.g., immigration influx, criminality rates, population by citizenship and labour status) and questions from the ESS. Additionally, we are studying the feasibility of including measurements of outgroup integration, and acceptance of immigrant and asylum policies.

#### 4.1.2. Defining Multilingual lists

It is crucial to ensure that concepts lists are balanced across languages and closely depict our intended domain. Our initial word list based on the multilingual European Migration Network (EMN) glossary of asylum and migration terms<sup>7</sup>. Such glossary contains approximately 500 terms and concepts reflecting the most recent European policy on migration and asylum.

Then, we consulted with native speakers and a migration studies specialist to increase the selected initial subset derived from the EMN glossary in order to expand and identify other concepts of interest, e.g. human trafficking. Finally, we prompted our dataset and models to verify the frequency of such words, excluding those with low frequency, and add missing words pointed as similar by the models. The lists were again revised by the domain specialist. During the aforementioned process, we verified and ensured that our group and concepts vector representations had low variance across the years and languages, as a way to ensure that our findings could not be attributed to instabilities in our vector representations.

## 4.2. Models

As it was done in our Spanish case study, using the datasets filtered by year, we trained skip-gram embedding models using the Fasttext implementation. After training Only words that appeared at least 10 times in each yearly dataset were taken into account in the training phase, and the resulting word vectors were  $L_2$  normalized. We evaluate the quality of our models

<sup>3</sup>Details about the corpus portions are available on [https://joint-research-centre.ec.europa.eu/language-technology-resources/dcep-digital-corpus-european-parliament\\_en](https://joint-research-centre.ec.europa.eu/language-technology-resources/dcep-digital-corpus-european-parliament_en)

<sup>4</sup>“National institute of Statistics” <https://www.ine.es/>

<sup>5</sup><https://www.europeansocialsurvey.org/>

<sup>6</sup><https://ec.europa.eu/eurostat>

<sup>7</sup>[https://ec.europa.eu/home-affairs/networks/european-migration-network-emn/emn-asylum-and-migration-glossary\\_en](https://ec.europa.eu/home-affairs/networks/european-migration-network-emn/emn-asylum-and-migration-glossary_en)

using generic word similarity benchmarks originally in English and then extended to other languages, such as the *RG-65* and the *MC-30* benchmarks.

To test the effect of sociopolitical variables in our time-series, we adopt a multilevel modelling approach. A multilevel model is an extension of a regression, in which data is structured in groups and coefficients can vary by group [38]. Concerning the inspection of patterns in the computed stereotype time-series, Autoregressive Integrated Moving Average (ARIMA) models could be applied.

### 4.3. Metrics

Distributional semantic models maintain the properties of vector spaces and adopt the hypothesis that meaning of a word is conveyed in its co-occurrences. Therefore, in order to measure the similarity between two given words represented by the vectors  $v_1$  and  $v_2$  we can apply the  $L_2$  normalized cosine similarity.

In our published study, to quantify social stereotypes in the trained word embedding models, we used the *bias score*, as defined by Garg et al., since it has been externally validated by the authors through correlations with census data. The bias score captures the strength of the association of a given set of words  $S$  with respect to two groups  $v_1$  and  $v_2$  as shown in Equation 1. The more negative that the bias score is, the more associated  $S$  is toward group two whereas the more positive, the more associated  $S$  is towards group one.

$$bias\ score = \sum_{v_s \in S} cosine(v_s, v_1) - cosine(v_s, v_2) \quad (1)$$

As for testing for biases in the sentence and contextualized embeddings, we start our investigation by using sentence templates and principal component analysis (PCA) [39, 40].

## 5. Discussion

Aiming to improve our work by discussing it with the Natural Language Processing community, we bring the following research elements for deliberation in this doctoral symposium:

1. The creation of automatic, or semiautomatic procedures for extracting and balancing word lists that represent concepts across languages;
2. Validating the quality of embeddings trained in a specific domain, i.e. parliamentary speeches;
3. Applying the multilingual setup to contextualized embedding models (e.g., BERT, RoBERTA).

The first point refers to the time-consuming and iterative process of creating the multilingual word lists and then balancing them across languages. Although we believe that the lists should be revised by domain specialists, automatic procedures for extracting initial word lists and verifying meaning equivalence across languages would be very beneficial to reduce the time spent in this step. Exploring the use of external resources with semantic information as an automatic method for creating the lists could facilitate the process.

The second point concerns the verification of embedding quality with an approach that allows us to see if it correctly represents the intended domain, in this case are parliamentary speeches, rather than using generic word similarity benchmarks.

Lastly, we wish to apply our framework to masked language contextualized embedding models such as BERT, but Therefore, we would like to discuss model architectures suitable for smaller datasets, or the use of pre-trained models like the Spanish RoBERTA [41] model.

## References

- [1] J. Sánchez-Junquera, B. Chulvi, P. Rosso, S. P. Ponzetto, How do you speak about immigrants? taxonomy and stereoimmigrants dataset for identifying stereotypes about immigrants, *Applied Sciences* 11 (2021) 3610.
- [2] H. Tajfel, A. A. Sheikh, R. C. Gardner, Content of stereotypes and the inference of similarity between members of stereotyped groups., *Acta Psychologica* (1964).
- [3] A. Marakasova, J. Neidhardt, Short-term semantic shifts and their relation to frequency change, in: *Proceedings of the Probability and Meaning Conference (PaM 2020)*, 2020, pp. 146–153.
- [4] M. J. Creighton, P. Schmidt, D. Zavala-Rojas, Race, wealth and the masking of opposition to immigrants in the netherlands, *International Migration* 57 (2019) 245–263.
- [5] A. C. Kroon, D. Trilling, T. Raats, Guilty by association: Using word embeddings to measure ethnic stereotypes in news coverage, *Journalism & Mass Communication Quarterly* (2020) 1077699020932304.
- [6] P. M. Sniderman, L. Hagendoorn, M. Prior, Predisposing factors and situational triggers: Exclusionary reactions to immigrant minorities, *American political science review* (2004) 35–49.
- [7] P. Sniderman, L. Hagendoorn, *Multiculturalism and its discontents in the netherlands: When ways of life collide*, 2007.
- [8] G. Lahav, et al., *Immigration and politics in the new Europe: Reinventing borders*, Cambridge University Press, 2004.
- [9] L. McLaren, H. Boomgaarden, R. Vliegthart, News coverage and public concern about immigration in britain, *International Journal of Public Opinion Research* 30 (2018) 173–193.
- [10] R. Zapata-Barrero, Perceptions and realities of moroccan immigration flows and spanish policies, *Journal of Immigrant & Refugee Studies* 6 (2008) 382–396.
- [11] A. Gorodzeisky, M. Semyonov, Perceptions and misperceptions: actual size, perceived size and opposition to immigration in european societies, *Journal of Ethnic and Migration Studies* 46 (2020) 612–630.
- [12] R. Tripodi, M. Warglien, S. L. Sullam, D. Paci, Tracing antisemitic language through diachronic embedding projections: France 1789-1914, in: *Proceedings of the 1st International Workshop on Computational Approaches to Historical Language Change*, 2019, pp. 115–125.
- [13] D. Herda, Too many immigrants? examining alternative forms of immigrant population innumeracy, *Sociological Perspectives* 56 (2013) 213–240.
- [14] Y. Pottie-Sherman, R. Wilkes, Does size really matter? on the relationship between



- immigrant group size and anti-immigrant prejudice, *International Migration Review* 51 (2017) 218–250.
- [15] E. Schlueter, P. Scheepers, The relationship between outgroup size and anti-outgroup attitudes: A theoretical synthesis and empirical test of group threat-and intergroup contact theory, *Social Science Research* 39 (2010) 285–295.
- [16] T. Bolukbasi, K.-W. Chang, J. Y. Zou, V. Saligrama, A. T. Kalai, Man is to computer programmer as woman is to homemaker? debiasing word embeddings, in: *Advances in neural information processing systems*, 2016, pp. 4349–4357.
- [17] H. Gonen, Y. Goldberg, Lipstick on a pig: Debiasing methods cover up systematic gender biases in word embeddings but do not remove them, in: *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, 2019, pp. 609–614.
- [18] N. Garg, L. Schiebinger, D. Jurafsky, J. Zou, Word embeddings quantify 100 years of gender and ethnic stereotypes, *Proceedings of the National Academy of Sciences* 115 (2018) E3635–E3644.
- [19] A. Lauscher, R. Takiyeddin, S. P. Ponzetto, G. Glavaš, AraWEAT: Multidimensional analysis of biases in Arabic word embeddings, in: *Proceedings of the Fifth Arabic Natural Language Processing Workshop, Association for Computational Linguistics, Barcelona, Spain (Online)*, 2020, pp. 192–199.
- [20] O. Papakyriakopoulos, S. Hegelich, J. C. M. Serrano, F. Marco, Bias in word embeddings, in: *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 2020, pp. 446–457.
- [21] J. Brandon, Using unethical data to build a more ethical world, *AI and Ethics* 1 (2021) 101–108.
- [22] E. M. Bender, T. Gebru, A. McMillan-Major, S. Shmitchell, On the dangers of stochastic parrots: Can language models be too big?, in: *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 2021, pp. 610–623.
- [23] D. Sorato, D. Zavala-Rojas, M. d. C. C. Ventura, Using word embeddings to quantify ethnic stereotypes in 12 years of spanish news, in: *Proceedings of the The 19th Annual Workshop of the Australasian Language Technology Association*, 2021, pp. 34–46.
- [24] A. C. Kozlowski, M. Taddy, J. A. Evans, The geometry of culture: Analyzing the meanings of class through word embeddings, *American Sociological Review* 84 (2019) 905–949.
- [25] K. Kurita, N. Vyas, A. Pareek, A. W. Black, Y. Tsvetkov, Measuring bias in contextualized word representations, in: *Proceedings of the First Workshop on Gender Bias in Natural Language Processing, Association for Computational Linguistics, Florence, Italy*, 2019, pp. 166–172. doi:10.18653/v1/W19-3823.
- [26] T. Manzini, L. Yao Chong, A. W. Black, Y. Tsvetkov, Black is to criminal as caucasian is to police: Detecting and removing multiclass bias in word embeddings, in: *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, Association for Computational Linguistics, Minneapolis, Minnesota, 2019, pp. 615–621. doi:10.18653/v1/N19-1062.
- [27] M.-E. Brunet, C. Alkalay-Houlihan, A. Anderson, R. Zemel, Understanding the origins of bias in word embeddings, in: *International conference on machine learning*, PMLR, 2019,



pp. 803–811.

- [28] M. Wevers, Using word embeddings to examine gender bias in dutch newspapers, 1950-1990, in: Proceedings of the 1st International Workshop on Computational Approaches to Historical Language Change, 2019, pp. 92–97.
- [29] A. Câmara, N. Taneja, T. Azad, E. Allaway, R. Zemel, Mapping the multilingual margins: Intersectional biases of sentiment analysis systems in English, Spanish, and Arabic, in: Proceedings of the Second Workshop on Language Technology for Equality, Diversity and Inclusion, Association for Computational Linguistics, Dublin, Ireland, 2022, pp. 90–106.
- [30] J. Ahn, A. Oh, Mitigating language-dependent ethnic bias in BERT, in: Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Online and Punta Cana, Dominican Republic, 2021, pp. 533–549. doi:10.18653/v1/2021.emnlp-main.42.
- [31] A. Névél, Y. Dupont, J. Bezançon, K. Fort, French crows-pairs: Extending a challenge dataset for measuring social bias in masked language models to a language other than english, in: ACL 2022-60th Annual Meeting of the Association for Computational Linguistics, 2022.
- [32] P. Bojanowski, É. Grave, A. Joulin, T. Mikolov, Enriching word vectors with subword information, Transactions of the Association for Computational Linguistics 5 (2017) 135–146.
- [33] P. Razgovorov, D. Tomás, et al., Creación de un corpus de noticias de gran tamaño en español para el análisis diacrónico y diatópico del uso del lenguaje, Comité Editorial 62 (2019) 29–36.
- [34] P. Koehn, Europarl: A parallel corpus for statistical machine translation, in: Proceedings of machine translation summit x: papers, 2005, pp. 79–86.
- [35] C. Rauh, J. Schwalbach, The parlspch v2 data set: Full-text corpora of 6.3 million parliamentary speeches in the key legislative chambers of nine representative democracies (2020).
- [36] T. Erjavec, M. Ogrodniczuk, P. Osenova, N. Ljubešić, K. Simov, A. Pančur, M. Rudolf, M. Kopp, S. Barkarson, S. Steingrímsson, et al., The parlmint corpora of parliamentary proceedings, Language resources and evaluation (2022) 1–34.
- [37] N. Hajlaoui, D. Kolovratnik, J. Väyrynen, R. Steinberger, D. Varga, Dcep-digital corpus of the european parliament, in: Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC’14), 2014.
- [38] A. Gelman, J. Hill, Data analysis using regression and multilevel/hierarchical models, Cambridge university press, 2006.
- [39] R. A. Ch’avez Mulsa, G. Spanakis, Evaluating bias in Dutch word embeddings, in: Proceedings of the Second Workshop on Gender Bias in Natural Language Processing, Association for Computational Linguistics, Barcelona, Spain (Online), 2020, pp. 56–71.
- [40] K. Kurita, N. Vyas, A. Pareek, A. W. Black, Y. Tsvetkov, Measuring bias in contextualized word representations, in: Proceedings of the First Workshop on Gender Bias in Natural Language Processing, 2019, pp. 166–172.
- [41] A. G. Fandiño, J. A. Estapé, M. Pàmies, J. L. Palao, J. S. Ocampo, C. P. Carrino, C. A. Oller, C. R. Penagos, A. G. Agirre, M. Villegas, Maria: Spanish language models, Procesamiento del Lenguaje Natural 68 (2022). doi:10.26342/2022-68-3.