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

Danielly Sorato
Universitat Pompeu Fabra
Barcelona, Spain
danielly.sorato@upf.edu

Diana Zavala-Rojas
European Social Survey ERIC
Universitat Pompeu Fabra
Barcelona, Spain
diana.zavala@upf.edu

Carme Colominas Ventura
Universitat Pompeu Fabra
Barcelona, Spain
carme.colominas@upf.edu

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.

1 Introduction

Languages are complex and systematic instruments of communication that reflect the culture of a given population. By studying language, it is 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 (Sánchez-Junquera et al., 2021; Tajfel et al., 1964). As such, language analysis is a good way to depict, understand, and demonstrate stereotypes (Garg et al., 2018; Basow, 1992; Wetherell and Potter, 1993; Bonilla-Silva and Forman, 2000). Nonetheless, like society, languages are not static. Variations in lexical systems can be observed over time due to a myriad of intra- and extra-linguistic factors. By analyzing extra-linguistic aspects, it is possible to gain insights into the dynamics of social, cultural, and political phenomena reflected in texts (Marakasova and Neidhardt, 2020).

Efficient methods for performing diachronic analysis are crucial, as manually evaluating several years of text collections is unfeasible due to the large amount of data involved. As such, computational methods for diachronic linguistic analysis are of utmost importance, and ongoing research shows that word embeddings models are helpful tools to this end (Garg et al., 2018; Kroon et al., 2020; Hamilton et al., 2016; Kutuzov et al., 2018; Lauscher et al., 2020).

Word embeddings are powerful representations of language, that allow for the quantification of relationships between words through efficient numerical operations inside the vector space. In this context, previous works demonstrated that such models contain machine-learned biases in their geometry that closely depict societal stereotypes (Bolukbasi et al., 2016b; Gonen and Goldberg, 2019; Garg et al., 2018; Kroon et al., 2020), which is not surprising since stereotypes are massively present in texts used to train computational models (Sánchez-Junquera et al., 2021; Nadeem et al., 2020). Although such language models should be carefully tested for biases and not blindly applied to widely computational applications due to ethically concerning outcomes (Papakyriakopoulos et al., 2020; Brandon, 2021; Bender et al., 2021), they can be a valuable tool for enabling sociolinguistic analysis on large volumes of textual data. This topic establishes a collaboration between computer science, social sciences, and linguistics, as hypotheses about social phenomena can be tested on language using computational methods.

In this study, we analyze the dynamics of stereotypical associations with seven nationalities, in the period of 2007 to 2018. We train our word embedding models using 1,757,331 news articles published in the Spanish newspaper *20 Minutos*, for the aforementioned time span. We adopt a culturally diverse perspective by taking into account some of the most representative foreign nationalities that lived in Spain in the aforementioned period according to the Instituto Nacional de Estadística (INE)[1].

[1] "National institute of Statistics" https://www.ine.es/

Namely, British, Colombian, Ecuadorian, German, Italian, Moroccan, and Romanian are included in this study.

We conduct a fine-grained analysis, studying the association of such nationalities with drug use, prostitution, crimes, and poverty concepts. Then, we compare our findings with sociopolitical variables, such as survey items from the European Social Survey (ESS), number of residents by nationality living in Spain, the rate of the population receiving unemployment benefits from the Spanish government, and the 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)[2]. To account for both group effects and error correlation, we use multilevel Random Effects (RE) models in our analysis.

This paper is organized as follows. In Section 2 we discuss related works. Subsequently, in Section 3 we state our research questions, present metrics, data, model training, and evaluation. Section 4 comprises the findings and discussion about results derived from this study. Finally, in Section 5 we present our conclusions, limitations, and future work.

2 Related Work

Word embeddings showed as a valuable tool, by means of enabling efficient methods for analyzing and quantifying linguistic and social phenomena in natural language. In the context of model stereotypical bias analysis, which is the focus of this paper, the first disseminated studies concern gender bias (Bolukbasi et al., 2016a,b; Zhao et al., 2018; Gonen and Goldberg, 2019; Park et al., 2018; Zhou et al., 2019). Nonetheless, biases can exist in many shapes and forms, which can lead to unfairness in subsequent downstream tasks (Mehrabi et al., 2019).

Garg et al., used both pre-trained models and models trained with the New York Times Annotated Corpus to quantify gender and ethnic stereotypes in 100 years of data for the English language. The reported bias series showed strong correlations with census data and demographic changes in the United States for gender and ethnic stereotypes. Similarly, Kozlowski et al. analyzed English embedding models, but focusing on social class biases.

Most works concerning the study of machine learned biases have English as target language, since there is more availability of linguistic resources that favors such analysis. Here we cite four relevant works conducted on non-English target languages. Wevers quantified gender biases in 40 years of Dutch newspapers categorized ideologically as liberal, social-democratic, neutral/conservative, Protestant, and Catholic. The results depict differences in gender bias and changes within and between newspapers over time. 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. Using the changes over time and embedding projections, they tracked the dynamics of antisemitic bias in the religious, economic, sociopolitical, racial, ethnic and conspiratorial domains. Sánchez-Junquera et al. detected stereotypes towards immigrants in political discourse by focusing in the narrative scenarios, i.e. the frames, used by political actors. They propose a taxonomy to capture immigrant stereotype dimensions and produced an annotated dataset with sentences that Spanish politicians have stated in the Congress of Deputies. Such dataset was used to train classifiers that detect and distinguish between stereotype categories.

More similar to ours, is the work of Kroon et al. In their study, the authors quantify the dynamics of stereotypical associations with different outgroups concerning low-status and high-threat concepts in 11 years of Dutch news data. The authors investigate both time invariant and time variant hypotheses, focusing on the difference of associations regarding the group membership (ingroup vs outgroups).

Our study distinguishes itself from the aforementioned studies by (i) the interdisciplinarity with social survey research, as the selected survey questions measure attitudes of Spanish people (the ingroup) towards immigrants (the outgroups) and can be interpreted as a proxy for cultural/economic threat perception; (ii) our choice of multilevel modeling (RE model), to combine types of phenomena (linguistic and social) and account for group effects; and (iii) the use of fine-grained lists representing crimes, drugs, poverty and prostitution concepts to investigate stereotypical portrayals. Additionally,

[2]According to the Data World Bank https://databank.worldbank.org

we contribute to the scarce literature on stereotypical bias analysis with non-English data sources by using Spanish from Spain as a target language.

3 Method

In this work, we aim to study the dynamics of the stereotypical portrayals of British, Colombian, Ecuadorian, German, Italian, Moroccan and Romanian nationalities with drugs, prostitution, crimes, and poverty concepts, which are some of the stereotypical frames associated to immigrants in the literature (Neyland, 2019; Kroon et al., 2020; Warner, 2005; Igartua et al., 2005; Light and Young, 2009). We investigate the effect that the Gross Domestic Product per capita (PPP) of the outgroup's country of origin has in the strength of stereotypical association. Namely, our hypothesis is that outgroups coming from countries with lower PPP than the host country (Spain), are more strongly associated with such concepts, due to posing a greater economic threat to the ingroup (Meuleman, 2011; Manevska and Achterberg, 2013)[3].

Then, we evaluate to what extent our findings can be explained by (i) the number of residents per nationality in Spain (i.e, the size of outgroup); (ii) rates of population receiving unemployment benefits; (iii) the number of offenses committed in the Spanish territory by outgroup background and; (iii) public opinion. In order to investigate such hypothesis, we adopt the following metrics, procedures and data.

3.1 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, although as shown by Garg et al., one could apply the Euclidean distance interchangeably.

To quantify social stereotypes in the trained word embedding models, we used a metric referred throughout this paper as *bias score*, which is the same metric used in Garg et al.. Such metric has been specifically chosen because 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. Hence, when we state that a word is biased toward a group, it is in the context of the bias score metric. The bias score equation is computed as in Equation 1, where S is a set of word vectors that represent a concept of interest (e.g., crimes), v_1 and v_2 are the averaged group vectors for word vectors in group one and two, respectively. An averaged group vector is computed by simply averaging the word vectors that compose a given group. 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.

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

To refer to the representation of the outgroups inside of the context of the embedding model and the bias score metric throughout this paper, we will use the name of the nationality in italics (e.g., *Spanish*, *Moroccan*).

We compare the similarity of concepts (i.e., word lists) related to drugs, prostitution, crimes and poverty to the concepts that represent the ingroup and the outgroups. For instance, if the word vector that represents the adjective $\overrightarrow{delincuente}$ ("delinquent") is more strongly associated with the word vector $\overrightarrow{rumano}$ ("Romanian") than with the word vector $\overrightarrow{español}$ ("Spanish"), that suggests there is bias in the model. It is not the similarity between $\overrightarrow{delincuente}$ and $\overrightarrow{rumano}$ that determines the presence of bias, but the fact that the distances between $\overrightarrow{rumano}$ and $\overrightarrow{español}$ are not equal regarding the adjective $\overrightarrow{delincuente}$.

3.2 Corpus

We compiled the Corpus of Spanish news *20 Minutos* (Razgovorov et al., 2019). The corpus contains 14 years of articles written in Spanish from Spain, comprising 711.840.945 distinct words, that were web-scraped from the newspaper *20 Minutos*[4] website in JSON format. Due to the limited availability of data measuring the sociopolitical indicators of interest (stated in the next subsection), we consider the years 2007 up to 2018 in our analysis.

According to a survey made in 2017 by Cardenal et al., about 40% of the consulted experts in the areas of political science and information science in

[3] The PPP of the Italian outgroup for the 2007-2018 period is only slightly higher while it is considerably higher for the British and German nationalities

[4] https://www.20minutos.es/

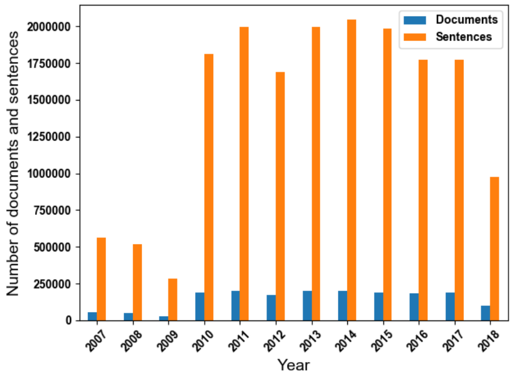

Figure 1: Number of documents and sentences per year in the *20 Minutos* data included in the analysis.

Spain consider *20 Minutos* is a neutral paper. The Figure 1 shows the number of articles and sentences per year in the corpus. Noticeably, for the years 2007 up to and including 2009 there is less data than for the subsequent years. We preprocessed the corpus, lower casing words, removing punctuation and numbers. Then, we filtered the data to create a dataset for each year of the corpus.

3.3 Sociopolitical variables

To build a sociopolitical indicator of ethnic threat perception, we use the mean score of three survey items from the European Social Survey (ESS) (NSD, 2020) studies (2006, 2008, 2010, 2012, 2014, 2016 and 2018). We used the Spanish respondent's answers (applying sample weights provided by ESS) of 11-point scales to the following questions: (i) "Is [country] made a worse or a better place to live by people coming to live here from other countries?"; (ii) "Would you say that [country]'s cultural life is generally undermined or enriched by people coming to live here from other countries?" and; (iii) "Would you say it is generally bad or good for [country]'s economy that people come to live here from other countries?". Missing data points for these time series were imputed using last observation carried forward (LOCF) strategy, which can be applied since the attitudes towards immigration tends to be stable from one year to another. Each survey was responded by at least 1500 people. The indicator of ethnic threat perception has the role of representing attitudinal data in the analysis, or in other words, identifying if the reported bias is somehow a reflection of the ingroup perceptions of these outgroups.

In addition, we use as indicators the number of foreign population by nationality residing in Spain[5], the rate of the population receiving unemployment social benefits (foreigners from the EU excluding Spain and foreigners from outside the EU)[6] and committed offenses by background, which can be countries from the EU excluding Spain (British, Germans, Italians and Romanians), America (Colombians and Ecuadorian), and Africa (Moroccans)[7]. Such datasets are publicly available and can be found in the INE database.

3.4 Word Embeddings Training and Evaluation

Using the datasets filtered by year, we trained skip-gram embedding models using the Fasttext implementation (Bojanowski et al., 2017). Since Spanish is a morphologically rich language, this model is a suitable choice as it takes into account the words' morphological structure. Due to the difference in the number of documents in the corpus across the years, we adopt a grid search strategy to define the optimal hyper-parameters of the models and favor embedding quality (see yearly hyper-parameters in Appendix). Only words that appeared at least 15 times in each yearly dataset were taken into account in the training phase. The resulting word vectors were L_2 normalized.

We evaluate our models using two Spanish word similarity benchmarks, namely *RG-65* (Camacho-Collados et al., 2015) and *MC-30* (Hassan and Mihalcea, 2009). The yearly models achieved an average of 0.72 and 0.70 Pearson correlation coefficient values in the *RG-65* and *MC-30* benchmarks for evaluating word similarity, respectively (variance $RG-65 = 0.0003$ and variance $MC-30 = 0.0011$). The evaluation results by year are shown in Appendix. In addition, we compute the average group vector for the ingroup and each of the outgroup nationalities and observe that, although some fluctuations can be observed for the *German* and *Spanish*, the variance is not significant. Therefore, our findings cannot be explained by the group vector variance.

3.5 Word lists

Here, we describe the process for selecting words that represent the crimes, drugs, poverty and prostitution concepts, as well as the ingroup and out-

[5] "Estadística del Padrón continuo. Población extranjera por Nacionalidad, provincias, Sexo y Año"

[6] "Tasa de paro por nacionalidad y periodo"

[7] "Estadística de condenados: Adultos. Condenados según número de delitos, nacionalidad y sexo"

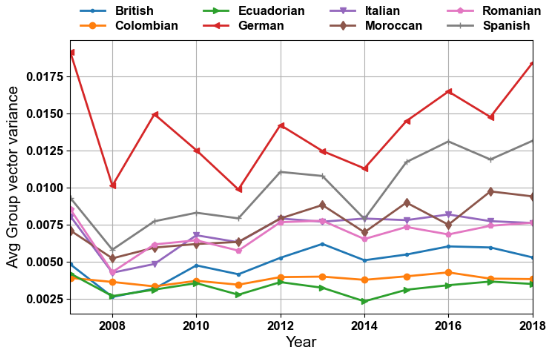

Figure 2: Average group vector variance.

groups. The word lists used for creating the vector representations of the ingroup and the outgroups were defined according to a simple rule: the nationality in masculine singular and plural form (e.g., Español, Españoles). The total frequencies per year for words that compose such lists are shown in the Appendix.

In order to identify words that represent crimes, drugs, poverty and prostitution categories, we start by fitting the high-treat and low-status words used by Kroon et al. in the aforementioned concepts[8]. Then, using an embedding model trained with the whole content of the corpus instead of the yearly slices, for each of the words in the initial list we retrieve the 20 most similar words in the vector space. Afterwards, the lists increased in the step described above were revised and updated again by the authors, excluding words that fall out of the desired concept category. We exclude feminine word inflections to favor lower group vector variances since the analyzed dataset is not very large. The lists of words for used each category of concepts are shown in the Appendix.

3.6 Panel Data

Due to the pooled structure of the data, i.e., yearly bias score measurements for each of the outgroups, we build a panel with $N = 84$ observations (12 years x 7 outgroups). The stationary behaviour of the panel was verified by applying the Levin–Lin–Chu test, which is equivalent to a pooled unit root test. The non-stationary hypothesis was rejected, meaning that the panel data series altogether is unaffected by changes in time. This same test was applied to test the panel data stationary behaviour in Kroon et al.. Additionally, we performed a careful analysis of the model residuals to ensure that there were no correlation patterns.

3.7 Random Effects model

To investigate the dependent series, we impose a Random Effects (RE) multilevel model for panel data. A multilevel model is an extension of a regression, in which data is structured in groups and coefficients can vary by group (Gelman and Hill, 2006). We consider the RE model an appropriate choice for this analysis, as we have pooled structured data and allows accounting for both group effects and error correlation. The following variables were used as predictors:

Year trend: the years from 2007 to 2018, treated as a categorical variable.

N Residents: size of outgroup residing in Spain, described in subsection 3.3.

Unemployment benefits: rate of population receiving unemployment benefits, described in subsection 3.3.

Perception: ingroup's perception of the outgroups, described in subsection 3.3.

Offenses number of offenses committed in the Spanish territory, described in subsection 3.3.

Lower PPP : dummy variable that indicates if the outgroups' country of origin has a *Lower PPP* than Spain. According to the Data World Bank[9], the countries with PPP lower than Spain for the period of analysis are Colombia, Ecuador, Morocco, and Romania ($LowerPPP = 1$). The countries with higher PPP are Germany, Italy and United Kingdom ($LowerPPP = 0$).

Analytical models should also be parsimonious, as fitting models with many random effects quickly multiplies the number of parameters to be estimated, particularly since random slopes are generally given covariances as well as variances (Bell et al., 2019; Matuschek et al., 2017). Hence, the chosen aforementioned indicators are the ones that, to the best of our knowledge, are most appropriated (both regarding data availability and purpose) to test our hypothesis.

4 Results and Discussion

In this section we discuss the findings and limitations of the present research. We analyse the dynamics of stereotypical associations comprised

[8] Excluding the words related to the police, terrorism and lack of intelligence, which do not suit the purposes of this work.

[9] Series named "GDP per capita, PPP (current international $)" available in the World Development Indicators series.

in 12 years (2007-2018) of Spanish local news published in the newspaper *20 Minutos*, comprising 1,757,331 news items, by training and analyzing yearly word embedding language models. Our objective is to quantify stereotypes in such items towards the aforementioned outgroups, taking into account a cultural dimension by studying seven of the most prominent foreign outgroups living in Spain considering the aforementioned period of analysis. We explore the hypothesis that outgroups coming from countries which have a *Lower PPP* than the host country (Spain), have stronger stereotypical associations with concepts related to crimes, drugs, poverty and prostitution, as a consequence of representing a greater social threat to the ingroup.

The yearly average bias scores concerning concepts related to crimes and drugs are depicted in Figures 3 and 4. The trends in Figure 3 show that, most of the outgroups are more strongly associated with the crimes concepts than the *Spanish* ingroup. The *Colombian* and the *Romanian* are the outgroups more strongly associated with crimes concepts, while the *German* and the *British* are the two outgroups less associated. In fact, for most years, the bias score values are negative for the *German* and the *British* outgroups. In contrast, for the *Colombian*, *Ecuadorian*, *Morrocan*, and *Romanian* outgroups, bias score values are always positive. A similar pattern can be observed in Figure 4, in the case of stereotypes concerning drugs.

The results of the Random effects model for the aforementioned series are presented in Table 1, and the main effects of the predictors are shown in the Model 1. In accordance to our expectations, the *Lower PPP* variable affects the bias significantly in both series. The positive coefficients indicate that the *Colombian*, the *Ecuadorian*, the *Moroccan* and the *Romanian* outgroups have higher stereotypical association with crimes and drugs concepts than the *German*, the *British* and the *Italian* outgroups. The year trend does have a significant effect, except for years 2009 and 2011 for crimes series, and years 2010 and 2011 for the drugs series. The positive coefficients indicate that the bias score for such years was higher than for the basis year, 2007.

To further inspect the effects of the *Lower PPP* variable, we add interaction terms in Model 2. For both series, there is a strongly significant relationship between *Lower PPP* and *Unemployment benefits*, such that when the rate of population receiving unemployment benefits increases, the stereotype association for *Colombian*, *Ecuadorian*, *Moroccan* and *Romanian* ($LowerPPP = 1$) also increases, but decreases for *German*, *British* and *Italian* outgroups. Similarly, the interaction with the number of committed offenses in the drugs series reveals that an increase in the offenses lead to stronger stereotypical associations for the first outgroups, but not for the latter. For the series concerning crimes concepts, it is also possible to observe that the public opinion threat perception decreases as stereotypical associations increases.

The yearly average bias scores for concepts related to poverty and prostitution are depicted in Figures 5 and 6. For poverty related concepts, *German*, *Italian*, and *British* bias score values are negative for most years, meaning that poverty concepts are actually more associated with the *Spanish* ingroup when compared to such outgroups. The same is not true for *Colombian*, *Ecuadorian*, *Moroccan*, and *Romanian* outgroups. Again, in Figure 6 it is possible to observe that same division between outgroups. The descriptive analysis show that, overall, outgroups in the *Lower PPP* classification exhibit stronger association with concepts related to prostitution and poverty.

The Table 2 shows the results of the Random Effects model for the aforementioned bias series. Consistently, for the two dependent series a strong effect regarding the *Lower PPP* variable can be observed meaning that again the *British*, the *German*, and the *Italian* are appreciably less associated with poverty and prostitution concepts than the *Colombian*, the *Ecuadorian*, the *Moroccan*, and the *Romanian* outgroups.

Concerning time effects, only the years 2009 and 2011 affect significantly the poverty series, while the year trend is not significant for the prostitution stereotypical associations. Comparably to the findings described for the crimes and drugs concepts, the *Unemployment benefits* predictor has a significant involvement with the dependent series, indicating discrepancy between lower and higher PPP groups. Aside from the interaction with the unemployment benefits predictor, which has the same pattern described above for the crimes and drugs series, no other predictor interacts significantly with the *Lower PPP* group.

The strong effect of the *Lower PPP* predictor on our analysis that news discourse emphasises the ethnicity of certain outgroups more than others. Furthermore, the interpretation of main effects and

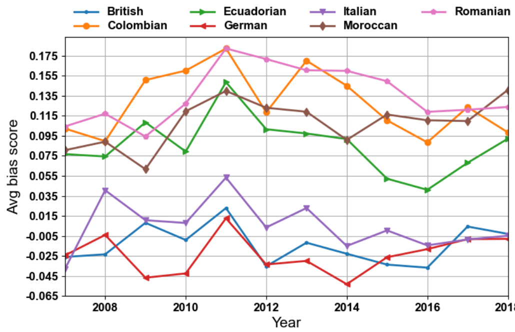

Figure 3: Average bias score for crimes concepts.

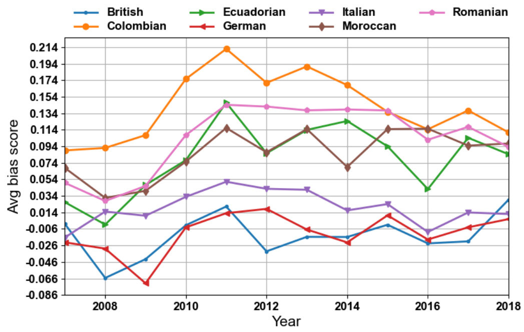

Figure 4: Average bias score for drugs concepts.

	Crimes		Drugs	
Predictors	**Model 1**	**Model 2**	**Model 1**	**Model 2**
Year.2008	0.0297 (0.0150)	0.0508** (0.0164)	-0.0047 (0.0219)	-0.0166 (0.0211)
Year.2009	0.0408* (0.0197)	0.0881*** (0.0217)	0.0139 (0.0269)	0.0440 (0.0294)
Year.2010	0.0306 (0.0264)	0.0731* (0.0327)	0.0508** (0.0351)	0.1314** (0.0393)
Year.2011	0.0753** (0.0303)	0.1232** (0.0376)	0.0868* (0.0400)	0.1786*** (0.0453)
Year.2012	0.0406 (0.0347)	0.0958* (0.0366)	0.0636 (0.0424)	0.1641*** (0.0470)
Year.2013	0.0551 (0.0325)	0.1118** (0.0394)	0.0736 (0.0423)	0.1750*** (0.0466)
Year.2014	0.0378 (0.0316)	0.0904* (0.0366)	0.0577 (0.0392)	0.1516*** (0.0425)
Year.2015	0.0292 (0.0252)	0.0689* (0.0294)	0.0581 (0.0319)	0.1321*** (0.0339)
Year.2016	0.0054 (0.0247)	0.0340 (0.0296)	0.0224 (0.0340)	0.0865* (0.0374)
Year.2017	0.0185 (0223)	0.0393 (0.0268)	0.0364 (0.0288)	0.0883* (0.0305)
Year.2018	0.0068 (0.0249)	0.0162 (0.0321)	0.0259 (0.0344)	0.0902 (0.0405)
Lower PPP	0.1207*** (0.0102)	0.2263*** (0.0637)	0.1186*** (0.0131)	0.0508 (0.0827)
N Residents	3.428e-05 (1.796e-05)	2.281e-05 (2.121e-05)	-3.799e-05 (2.239e-05)	-6.058e-05 (2.559e-05)
Unemployment benefits	-0.0013 (0.0012)	-0.0054** (0.0018)	-0.0009 (0.0015)	-0.0077*** (0.0021)
Offenses	2.842e-06 (1.953e-06)	4.621e-06 (2.465e-06)	1.543e-06 (2.396e-06)	-1.391e-06 (3.221e-06)
Perception	-0.0004 (0.0002)	-0.0002 (0.0003)	-0.0002 (0.0003)	0.0007 (0.0004)
Unemployment x Lower PPP	-	0.0023** (0.0008)	-	0.0040*** (0.0010)
Offenses x Lower PPP	-	-1.821e-06 (1.672e-06)	-	2.072e-06* (2.352e-06)
Perception x Lower PPP	-	-0.0007* (0.0003)	-	-0.0003 (0.0003)
N	84	84	84	84
Residual	0.000354	0.000292	0.000426	0.000342
R-squared	0.93	0.95	0.90	0.92

Table 1: Random Effects model predictions of bias scores for concepts related to crimes and drugs. **p* < .05, ***p* < .01, ****p* < .001. Standard errors for each coefficient shown in parenthesis.

interactions with sociopolitical variables indicates that stereotypical portrayals seem to be dissociated from real demographic trends. 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 ideologies (Van Dijk, 2000). Our findings go in line with frames described in other studies made with European newspapers, which indicate the semantic link between foreigners, prostitution, criminality and degeneracy (Neyland, 2019; Stenvoll, 2002; Light and Young, 2009; Igartua et al., 2005; Rancu, 2011), especially for Eastern European and Latin American backgrounds. We join previous studies pointing that media coverage can be stereotypical, associating ethnic outgroups with stigmatized attributes, and therefore having serious negative effects both on individuals and society, as news are powerful sources of the discursive demoralization of marginalised groups (Hamborg et al., 2018; Zilber and Niven, 2000; Angermeyer and Schulze, 2001; Sui and Paul, 2017; Kroon et al., 2020; Farris and Silber Mohamed, 2018; Milioni et al., 2015; Abrajano et al., 2017; Saiz de Lobado García et al., 2018; Neyland, 2019).

We cite the following limitations of our findings. The present analysis considers only one data source, therefore our conclusions cannot be generalized to other Spanish media outlets. Although the unavailability of other diachronic corpora for Spanish from Spain limits our conclusion to a single news outlet, we argue that this study is a valuable contribution to stereotype analysis in media discourse using a non-English target language.

Further, we acknowledge that by excluding gender inflected words, stereotypes about women that could be informative were left out. We do wish to explore gender inflected words in future work

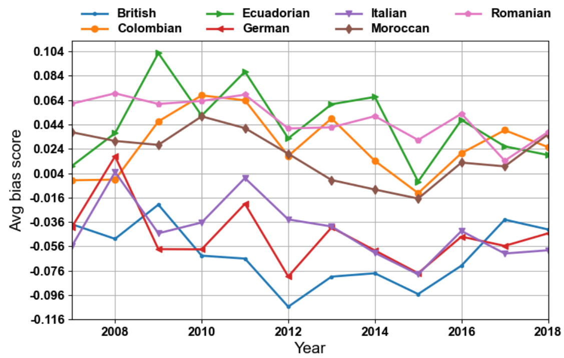

Figure 5: Average bias score for poverty concepts.

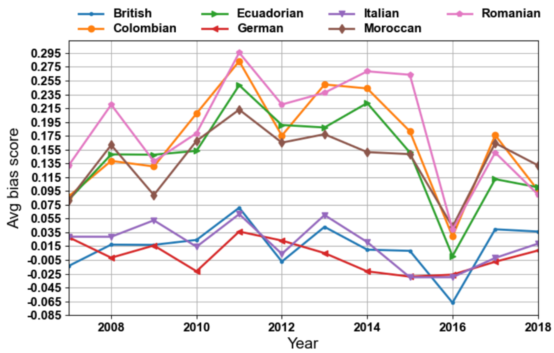

Figure 6: Average bias score for prostitution concepts.

	Poverty		Prostitution	
Predictors	**Model 1**	**Model 2**	**Model 1**	**Model 2**
Year.2008	0.0409 (0.0206)	0.0388* (0.0180)	0.0529 (0.0268)	0.0606* (0.0289)
Year.2009	0.0595** (0.0177)	0.0899*** (0.0202)	0.0397 (0.0387)	0.1230** (0.0377)
Year.2010	0.0429 (0.0253)	0.1036** (0.0328)	0.0350 (0.0446)	0.1720*** (0.0479)
Year.2011	0.0611* (0.0278)	0.1232** (0.0376)	0.1043 (0.0525)	0.2576*** (0.0550)
Year.2012	0.0270 (0.0316)	0.1027* (0.0389)	0.0487 (0.0551)	0.2184*** (0.0586)
Year.2013	0.0427 (0.0285)	0.1191** (0.0384)	0.0792 (0.0558)	0.2503*** (0.0597)
Year.2014	0.0302 (0.0270)	0.1008** (0.0352)	0.0736 (0.0563)	0.2305*** (0.0551)
Year.2015	-0.0033 (0.0229)	0.0516 (0.0291)	0.0425 (0.0507)	0.1627** (0.0504)
Year.2016	0.0197 (0.0219)	0.0656 (0.0296)	-0.0726 (0.0414)	0.0239 (0.0428)
Year.2017	0.0095 (0.0197)	0.0460 (0.0256)	0.0166 (0.0355)	0.0920* (0.0355)
Year.2018	0.0023 (0.0230)	0.0440 (0.0311)	-0.0233 (0.0380)	0.0565 (0.0445)
Lower PPP	0.0991*** (0.0108)	0.0821 (0.0767)	0.1399*** (0.0173)	0.1622 (0.1083)
N Residents	-1.664e-05 (1.549e-05)	-3.534e-05 (1.798e-05)	3.574e-05 (2.41e-05)	-1.492e-05 (2.731e-05)
Unemployment benefits	-0.0018 (0.0012)	-0.0070*** (0.0018)	-0.0007 (0.0021)	-0.0125*** (0.0029)
Offenses	1.004e-06 (1.708e-06)	1.084e-07 (2.227e-06)	4.065e-06 (2.168e-06)	5.893e-06 (3.789e-06)
Perception	-0.0003 (0.0002)	0.0003 (0.0003)	-0.0005 (0.0003)	0.0004 (0.0005)
Unemployment x Lower PPP	-	0.0031** (0.0009)	-	0.0070*** (0.0013)
Offenses x Lower PPP	-	-1.806e-08 (2.227e-06)	-	-5.458e-06 (3.336e-06)
Perception x Lower PPP	-	-0.0002 (0.0003)	-	-0.0003 (0.0004)
N	84	84	84	84
Residual	0.000366	0.000334	0.00118	0.000769
R-squared	0.84	0.87	0.93	0.96

Table 2: Random Effects model predictions of bias scores for concepts related to poverty and prostitution. $*p < .05. **p < .01$, $***p < .001$. Standard errors for each coefficient shown in parenthesis.

with a more suitable dataset. Lastly, we would like to point that all these nationalities have intricate and deep political relationships with Spain which certainly go beyond having a higher or lower GPD per capita.

5 Conclusion

In this work we analyzed the dynamics of stereotypical associations concerning seven of the most prominent ethnic outgroups living in Spain using language models trained with 12 years of news items from the Spanish newspaper *20 Minutos*. We investigated biases concerning concepts related to crimes, drugs poverty and prostitution, exploring the relation between the stereotypical associations and the GPD per capita (PPP) of the outgroups' countries of origin, public opinion, outgroup size, unemployment subsidy, and number of committed offenses in the Spanish territory.

Our results show that the texts exhibit stereotypical associations, especially for the Colombian, Ecuadorian, Moroccan and Romanian outgroups. We conclude that the examined news articles emphasize the nationality of certain ethnicities, which hinder the integration process of already marginalized outgroups. Moreover, these associations can be further propagated and amplified through computational algorithms if available data indiscriminately (Bolukbasi et al., 2016b; Nadeem et al., 2020), leading to concerning outcomes.

As future work, we aim to move to a multilingual perspective and compare outgroup stereotypes across different languages. Furthermore, we wish to examine stereotypes in political discourse, to inspect if patterns similar to the ones found in this work can be observed.

Acknowledgments

This work was partially supported by the Universitat Pompeu Fabra (UPF), through a PhD studentship grant and the use of Marvin Cluster[10] to train and analyse the language models.

References

Marisa A Abrajano, Zoltan Hajnal, and Hans JG Hassell. 2017. Media framing and partisan identity: The case of immigration coverage and white macropartisanship. *Journal of Race, Ethnicity and Politics*, 2(1):5.

Matthias C Angermeyer and Beate Schulze. 2001. Reinforcing stereotypes: how the focus on forensic cases in news reporting may influence public attitudes towards the mentally ill. *International Journal of Law and Psychiatry*.

Susan A Basow. 1992. *Gender: Stereotypes and roles*. Thomson Brooks/Cole Publishing Co.

Andrew Bell, Malcolm Fairbrother, and Kelvyn Jones. 2019. Fixed and random effects models: making an informed choice. *Quality & Quantity*, 53(2):1051–1074.

Emily M Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, pages 610–623.

Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov. 2017. Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics*, 5:135–146.

Tolga Bolukbasi, Kai-Wei Chang, James Zou, Venkatesh Saligrama, and Adam Kalai. 2016a. Quantifying and reducing stereotypes in word embeddings. *arXiv preprint arXiv:1606.06121*.

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.

Eduardo Bonilla-Silva and Tyrone A Forman. 2000. "i am not a racist but...": Mapping white college students' racial ideology in the usa. *Discourse & society*, 11(1):50–85.

Jamie Brandon. 2021. Using unethical data to build a more ethical world. *AI and Ethics*, 1(2):101–108.

[10]https://www.upf.edu/web/sct-sit/marvin-cluster

José Camacho-Collados, Mohammad Taher Pilehvar, and Roberto Navigli. 2015. A framework for the construction of monolingual and cross-lingual word similarity datasets. In *Proceedings of ACL (2)*, pages 1–7.

Anas Cardenal, Carol Galais, and Joaquim Moré. 2018. El reto de medir el sesgo ideológico en los medios escritos digitales. *quaderns del cac*.

Emily M Farris and Heather Silber Mohamed. 2018. Picturing immigration: How the media criminalizes immigrants. *Politics, Groups, and Identities*, 6(4):814–824.

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.

Andrew Gelman and Jennifer Hill. 2006. *Data analysis using regression and multilevel/hierarchical models*. Cambridge university press.

Hila Gonen and Yoav Goldberg. 2019. Lipstick on a pig: Debiasing methods cover up systematic gender biases in word embeddings but do not remove them. *arXiv preprint arXiv:1903.03862*.

Felix Hamborg, Norman Meuschke, and Bela Gipp. 2018. Bias-aware news analysis using matrix-based news aggregation. *International Journal on Digital Libraries*, pages 1–19.

William L Hamilton, Jure Leskovec, and Dan Jurafsky. 2016. Diachronic word embeddings reveal statistical laws of semantic change. *arXiv preprint arXiv:1605.09096*.

Samer Hassan and Rada Mihalcea. 2009. Cross-lingual semantic relatedness using encyclopedic knowledge. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, pages 1192–1201.

Juan José Igartua, Lifen Cheng, and Carlos Muňiz. 2005. Framing latin america in the spanish press: A cooled down friendship between two fraternal lands.

Austin C Kozlowski, Matt Taddy, and James A Evans. 2019. The geometry of culture: Analyzing the meanings of class through word embeddings. *American Sociological Review*, 84(5):905–949.

Anne C Kroon, Damian Trilling, and Tamara Raats. 2020. Guilty by association: Using word embeddings to measure ethnic stereotypes in news coverage. *Journalism & Mass Communication Quarterly*, page 1077699020932304.

Andrey Kutuzov, Lilja Øvrelid, Terrence Szymanski, and Erik Velldal. 2018. Diachronic word embeddings and semantic shifts: a survey. *arXiv preprint arXiv:1806.03537*.

Anne Lauscher, Rafik Takieddin, Simone Paolo Ponzetto, and Goran Glavaš. 2020. Araweat: Multidimensional analysis of biases in arabic word embeddings. *arXiv preprint arXiv:2011.01575*.

Duncan Light and Craig Young. 2009. European union enlargement, post-accession migration and imaginative geographies of the 'new europe': Media discourses in romania and the united kingdom. *Journal of Cultural Geography*, 26(3):281–303.

Katerina Manevska and Peter Achterberg. 2013. Immigration and perceived ethnic threat: Cultural capital and economic explanations. *European Sociological Review*, 29(3):437–449.

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.

Hannes Matuschek, Reinhold Kliegl, Shravan Vasishth, Harald Baayen, and Douglas Bates. 2017. Balancing type i error and power in linear mixed models. *Journal of memory and language*, 94:305–315.

Ninareh Mehrabi, Fred Morstatter, Nripsuta Saxena, Kristina Lerman, and Aram Galstyan. 2019. A survey on bias and fairness in machine learning. *arXiv preprint arXiv:1908.09635*.

Bart Meuleman. 2011. Perceived economic threat and anti-immigration attitudes: Effects of immigrant group size and economic conditions revisited. *Cross-cultural analysis: Methods and applications*, pages 281–310.

Dimitra L Milioni, Lia-Paschalia Spyridou, and Konstantinos Vadratsikas. 2015. Framing immigration in online media and television news in crisis-stricken cyprus. *The Cyprus Review*, 27(1):155–185.

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

Matilda FF Neyland. 2019. *The Sexual Other: Discursive constructions of migrant sex workers in New Zealand media*. Ph.D. thesis, Victoria University of Wellington.

NSD. 2020. European social survey cumulative file, ess 1-9 (2020). *Data file edition 1.0. NSD - Norwegian Centre for Research Data, Norway - Data Archive and distributor of ESS data for ESS ERIC*.

Orestis Papakyriakopoulos, Simon Hegelich, Juan Carlos Medina Serrano, and Fabienne Marco. 2020. Bias in word embeddings. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, pages 446–457.

Ji Ho Park, Jamin Shin, and Pascale Fung. 2018. Reducing gender bias in abusive language detection. *arXiv preprint arXiv:1808.07231*.

Raluca Oana Matu Rancu. 2011. Exclusion, marginalization and prejudice: The image of the romanian woman in spanish society. *International Journal of Diversity in Organisations, Communities & Nations*, 10(5).

Pavel Razgovorov, David Tomás, et al. 2019. Creación de un corpus de noticias de gran tamano en espanol para el análisis diacrónico y diatópico del uso del lenguaje.

María Ester Saiz de Lobado García et al. 2018. Metáfora y percepción: análisis de la ideología subyacente en el discurso jurídico sobre inmigración.

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.

Dag Stenvoll. 2002. From russia with love? newspaper coverage of cross-border prostitution in northern norway, 1990—2001. *European Journal of Women's Studies*, 9(2):143–162.

Mingxiao Sui and Newly Paul. 2017. Latino portrayals in local news media: Underrepresentation, negative stereotypes, and institutional predictors of coverage. *Journal of Intercultural Communication Research*, 46(3):273–294.

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*.

Rocco Tripodi, Massimo Warglien, Simon Levis Sullam, and Deborah Paci. 2019. Tracing antisemitic language through diachronic embedding projections: France 1789-1914. *arXiv preprint arXiv:1906.01440*.

Teun A Van Dijk. 2000. *On the analysis of parliamentary debates on immigration*. Citeseer.

Judith Ann Warner. 2005. The social construction of the criminal alien in immigration law, enforcement practice and statistical enumeration: Consequences for immigrant stereotyping. *Journal of Social and Ecological Boundaries*, 1(2):56–80.

Margaret Wetherell and Jonathan Potter. 1993. *Mapping the language of racism: Discourse and the legitimation of exploitation*. Columbia University Press.

Melvin Wevers. 2019. Using word embeddings to examine gender bias in dutch newspapers, 1950-1990. *arXiv preprint arXiv:1907.08922*.

Jieyu Zhao, Yichao Zhou, Zeyu Li, Wei Wang, and Kai-Wei Chang. 2018. Learning gender-neutral word embeddings. *arXiv preprint arXiv:1809.01496*.

Pei Zhou, Weijia Shi, Jieyu Zhao, Kuan-Hao Huang, Muhao Chen, Ryan Cotterell, and Kai-Wei Chang. 2019. Examining gender bias in languages with grammatical gender. *arXiv preprint arXiv:1909.02224*.

Jeremy Zilber and David Niven. 2000. Stereotypes in the news: Media coverage of african-americans in congress. *Harvard International Journal of Press/Politics*, 5(1):32–49.

A Word lists

In the next subsections we specify the word lists that were used to represent crimes, drugs, poverty and prostitution concepts, as well as the ingroup and outgroups. Please notice that some of the words in the lists are plural inflections that have no corresponding translation in English. We identify such words by adding '(plural)' next to the singular translation.

A.1 Ingroup and outgroups

Ingroup in Spanish: Español, Españoles'.
Ingroup translation: "Spanish", "Spanish (plural)".
British outgroup in Spanish: Británico, Británicos.
British outgroup translation: "British", "British (plural)".
Colombian outgroup in Spanish: Colombiano, Colombianos.
Colombian outgroup translation: "Colombian", "Colombians".
Ecuadorian outgroup in Spanish: Ecuatoriano, Ecuatorianos.
Ecuadorian outgroup translation: "Ecuadorian", "Ecuadorians".
German outgroup in Spanish: Alemán, Alemanes.
German outgroup translation: "German", "Germans".
Italian outgroup in Spanish: Italiano, Italianos.
Italian outgroup translation: "Italian", "Italians".
Moroccan outgroup in Spanish: Marroquí, Marroquíes.
Moroccan outgroup translation: "Moroccan", "Moroccans".
Romanian outgroup in Spanish: Rumano, Rumanos.
Romanian outgroup translation: "Romanian", "Romanians".

A.2 Frequency of Ingroup and outgroup words

The table 3 shows the frequencies by year of the words that were used to create the ingroup and outgroup vector representations in our study.

A.3 Crimes

Words in Spanish: Cabecilla, cabecillas, arrestado, arrestados, detenido, detenidos, sospecho, sospechos, sospechoso, sospechosos, ilegal, ilegales, ilegalidad, clandestino, clandestinos, clandestinidad, narcotráfico, narcotraficante, narcotraficantes, traficante, traficantes, contrabando, contrabandista, contrabandistas, aprehensión, aprehensiones, incautación, incautaciones, atraco, atracos, atracador, atracadores, asalto, asaltos, asaltante, asaltantes, crimen, criminalidad, criminal, criminales, delito, delitos, agresión, agresiones, delincuencia, delincuente, delincuentes, malhechor, malhechores, robo, robos, hurto,hurtos, sustracción, sustracciones, mafia, mafias, mafioso, mafiosos, violación, violaciones, violador, violadores, pedófilo, pedófilos, asesino, asesinos, asesinato,asesinatos, homicidio, homicidios, homicida, homicidas, violencia, violento, violentos,maltrato, maltratos, maltratador, maltratadores.

Translations: "faction leader", "faction leaders", "arrested", "arrested (plural)","detained", "detained (plural)", "suspect", "suspects", "shady", "shady (plural)", "illegal", "illegal (plural)", "illegality", "clandestine", "clandestine (plural)", "underground", "drug trafficking", "drug dealer", "drug traffickers", "trafficker", "traffickers", "smuggling", "smuggler", "smugglers", "apprehension", "apprehensions", "seizure", "seizures", "robbery", "robberies", "robber", "robbers", "assault", "assaults", "burglar", "burglars", "crime", "criminality", "criminal", "criminals", "felony", "felonies", "aggression", "aggressions", "delinquency", "delinquent", "delinquents", "malefactor", "malefactors", "stealing", "stealing (plural)", "theft", "theft (plural)", "thievery", "thievery (plural)", "mafia", "mafias", "gangster", "gangsters", "rape", "rapes", "rapist", "rapists", "pedophile", "pedophiles", "murderer", "murderers", "murder", "murders", "homicide", "homicides", "killer", "killers", "violence", "violent", "violent (plural)", "maltreatment", "maltreatments", "batterer", "batterers".

Year	British	Colombian	Ecuadorian	German	Italian	Moroccan	Romanian	Spanish
2007	340	199	226	433	411	679	472	3094
2008	338	312	172	362	273	981	457	3335
2009	190	124	93	271	167	539	171	2095
2010	1208	400	207	1927	954	2476	627	21158
2011	1294	387	165	2286	1171	1681	613	23566
2012	1240	288	122	1761	890	1738	443	18141
2013	1618	346	130	2212	905	2119	561	21183
2014	1519	357	104	2194	1154	2381	449	22082
2015	1366	286	88	1767	1051	1802	381	19123
2016	1526	206	141	1701	899	1087	287	15450
2017	1307	196	83	1518	947	1061	255	13986
2018	545	114	40	907	499	529	163	7556

Table 3: Frequency of the words that compose the ingroup and outgroup representations in the corpus *20 Minutos* by year.

A.4 Drugs

Words in Spanish: Droga, drogas, adicción, adicciones, adicto, adictos, drogadicción, drogadicto, drogadictos, estupefaciente, estupefacientes, drogodependencia, drogodependencias, drogodependiente, drogodependientes,alcohol, alcoholismo, borracho, borrachos, heroína, cocaína, papelina, papelinas, bolsita, bolsitas, hachís, marihuana, sustancia, sustancias, cannabis, metanfetamina, anfetamina, speed, éxtasis, mdma.

Translations: “drug”, “drugs”, “addiction”, “addictions”, “addict”, “addicts”,“drug addiction”, “drug addict”, “drug addicts”, “narcotic”, “narcotics”, drug addiction, drug addiction, “junkie”, “junkies”, “alcohol”, “alcoholism”, “drunk”, “drunk (plural)”, “heroin”, “cocaine”, “ “drug paper”[11], “drug papers”, “drug bag”[12], “drug bags” “hashish”, “marijuana”, “substance”, “substances”, “cannabis”, “methamphetamine”, “amphetamine”, “speed”, “ecstasy”, “mdma”.

A.5 Poverty

Words in Spanish: miseria, miserable, miserables, pobreza, pobre, pobres, empobrecimiento, empobrecido, empobrecidos, mendicidad, mendigo, mendigos, desfavorecido, desfavorecidos, necesitado, necesitados, desesperación, desesperados, desesperado, vulnerabilidad, vulnerables, vulnerable, chabola, chabolas, chabolista, chabolistas, infravivienda, infraviviendas, barriada, barriadas, vagabundo, vagabundos, marginalidad, marginal, marginales, marginación, marginado, marginados.

[11]Papelina is a piece of paper to hold small amounts of drugs.

[12]Bolsita is a small plastic bag to hold small amounts of drugs.

Translations: “misery”, “miserable”, “miserable (plural)”, “poverty”, “poor”, “poor (plural)”, “impoverishment”, “impoverished”, “impoverished (plural)”, “begging”, “beggar”, “beggars”, “disadvantaged”, “disadvantaged (plural)”, “people in need”, “people in need (plural)”, “desperation”, “desperate”, “desperate (plural)”, “vulnerability”, “vulnerable”, “vulnerable (plural)”, “shanty town”, “shanty town (plural)”, “person that lives in shanty town”, “person that lives in shanty town (plural)”, “slum”, “slums”, “poor neighborhood”, “poor neighborhoods”, “vagabond”, “vagabonds”, “marginality”, “marginal”, “marginal (plural)”, “marginalization”, “marginalized (plural)”,“ marginalized (plural)”.

A.6 Prostitution

Words in Spanish: Prostitución, prostíbulo, prostíbulos, prostituta, prostitutas, proxenetismo, proxeneta, proxenetas.

Translations: “Prostitution”,“ brothel”, “brothels”, “prostitute”, “prostitutes”, “pimping”, “pimp”, “pimps”.

B Word Embeddings

In the following subsections we show the hyper-parameters used to train the word embedding models and the yearly scores of the $RG - 65$ and $MC - 30$ semantic similarity benchmarks.

B.1 Hyper-parameters

All Fasttext skipgram models were trained with 250 dimensions, five epochs and minimum word frequency of 15 occurrences. The hyper-parameters selected by the grid-search are shown below in the Table. Default values were used for hyper-parameters that are not mentioned here [13].

Year	Window size	N-grams	Min/max
2007	7	1	4/6
2008	8	2	2/6
2009	8	4	3/6
2010	7	3	default (0/0)
2011	6	1	2/6
2012	5	1	default (0/0)
2013	5	3	default (0/0)
2014	8	1	default (0/0)
2015	5	4	default (0/0)
2016	4	4	3/6
2017	4	1	default (0/0)
2018	5	1	4/6

Table 4: Embedding training hyper-parameters. Min/max means the minimum and maximum length of char ngram.

	RG-65 Pearson coefficient	RG-65 p-value	MC-30 Pearson coefficient	MC-30 p-value
2007	0.74	4.54e-08	0.67	2.99e-04
2008	0.75	2.51e-09	0.72	7.2e-04
2009	0.75	2.43e-07	0.78	9.56e-04
2010	0.70	5.66e-09	0.71	4.2e-04
2011	0.72	6.79e-09	0.66	1.6e-0.3
2012	0.70	7.75e-09	0.68	9.49e-04
2013	0.70	5.88-09	0.69	7.96e-04
2014	0.73	1.22e-09	0.71	4.35e-04
2015	0.71	3.35e-10	0.72	2.7e-04
2016	0.73	2.17e-09	0.69	7.76e-04
2017	0.73	5.16e-09	0.66	1.89e-03
2018	0.72	1.4e-08	0.72	5.27e-04

Table 5: Yearly semantic similarity evaluation results for RG-65 and MC-30 benchmarks.

B.2 Semantic similarity evaluation

The Table 5 shows the Pearson coefficients and p-values for the $RG-65$ and $MC-30$ Spanish word similarity scores, for each of the yearly trained embedding models.

[13]https://fasttext.cc/docs/en/options.html