Language Patterns of Outgroup Prejudice

Abstract

Although explicit verbal expression of prejudice and stereotypes have become less common due to the recent rise of social norms against prejudice, prejudice in language still persists in more subtle forms. Leveraging a natural language corpus of 1.8 million newspaper articles, the present study examined patterns and biases underlying the language associated with 60 U.S. minority groups. We found that human perception of social distance has its linguistic footprint in language production: Groups perceived as socially distant (vs. close) are also more likely to be mentioned in abstract (vs. concrete) language. There was also a strong positive correlation between valence and concreteness unique to language concerning minority groups, suggesting a strong bias for more socially distant groups to be represented in negative contexts. We also investigated the content of outgroup prejudice by applying a topic model on language referencing minority groups in the context of immigration, which highlights their outgroup identity. We identified 15 immigrant-related topics (e.g., politics, arts, crime, illegal workers), the strength of their association with each minority group, and their relation with perceived sentiment towards minority groups. Our approach to prejudice provides a practical and ecologically valid method for comparing prejudice towards a large number of minority groups in both degree and content, supports and elucidates prior theories of outgroup prejudice, and offers a way forward for research in this area.

Keywords: outgroup prejudice, natural language processing, social distance

Language Patterns of Outgroup Prejudice

1. Introduction

Language plays a central role in prejudice. In his classic book *The Nature of Prejudice*, Gordon Allport (1954) noted that ethnic labels often attract more negative attributes than they should. Subsequent studies have shown that language not only reflects explicit and implicit prejudice but also influences how recipients perceive and judge outgroup members (see Collins & Clément, 2012, for a review). Today, despite the increase in antiprejudice norms and corresponding decreases of explicit expression of prejudice, prejudice in language still persists in more subtle forms (Augoustinos & Every, 2007; Maass et al., 1989). In the present article, we analyze patterns and biases in language underlying a prejudiced description of the 60 most common ethnic and religious minority groups in the United States. In particular, we address a number of questions related to how minority groups were represented on a leading U.S. Newspaper (*The New York Times*) at various degrees of perceived social distance and in relation to a variety of outgroup-related topics. Before we introduce these questions, we first introduce the theories that motivates them.

1.1. Cognitive Accounts of Prejudice

Outgroup negativity is difficult to eradicate because it is deeply rooted in the basic human propensity for social categorical thinking (Allport, 1954; Brewer, 1979; Tajfel, 1982). Immigrants, as natural outgroups, are often perceived as untrustworthy outsiders (Alexander et al., 1999; Cuddy et al., 2007; Cuddy et al., 2009; Peabody, 1985; Poppe, 2001) despite bringing innovation, skilled labor, investment, and cultural diversity to their host countries (Borjas, 1990; Carens, 2013; Skeldon, 1997). Outgroup negativity is partly maintained by ultimate attribution error, which is the propensity to explain others' negative behaviors as resulting from dispositional properties of their categorically defined outgroup, but their positive behaviors as the result of idiosyncratic situational factors (Pettigrew, 1979). Remarkably, outgroup status and sentiment is flexible. Laboratory analogs of group formation—often called "minimal group paradigms"—have demonstrated that the minimum condition for intergroup bias is categorization into a group, but the criteria for that categorization can be as arbitrary as a preference for Kandinsky over Klee (Tajfel et al., 1971). Furthermore, situational factors can influence group boundaries. In Sherif et al.'s (1961) Robbers' Cave experiment, boys at a camp were assigned to groups at random.

Increasing levels of prejudice and hostility toward outgroup members were observed over a period of weeks. When the groups worked collectively toward a common goal, however, the boundaries between groups rapidly broke down.

One solution known to mitigate prejudice is direct interaction with outgroups (intergroup contact theory; Allport, 1954). A meta-analysis of more than 500 studies found that increased intergroup contact that included prosocial qualities such as equal status and cooperation reduced prejudice in 94% of independent samples (Pettigrew & Tropp, 2006; also see a more recent meta-analysis: Paluck et al., 2019). Correspondingly, social distance fuels dispositional inference and prejudice (Jones & Nisbett, 1987). For example, intergroup contact plays a substantial role in explaining the rural–urban divide in perceptions about immigrants, whereby rural populations with little contact to immigrants tend to have more negative attitudes toward immigration than do urban populations who interact with immigrants regularly (Fennelly & Federico, 2008).

One psychological impact of quality intergroup contact could be reduced social distance, a concept popularized by Emory Bogardus that refers to the degree with which, psychologically speaking, a person wants to accept or remain separate from members of different social groups (Bogardus, 1927). The Bogardus scale has nearness, intimacy, and familiarity at one end, and farness, difference, and unfamiliarity at the other. Subsequent replications of Bogardus's original 1927 study show that over the past 80 years, Americans have perceived decreasing levels of social distance towards all minority groups (e.g., Bogardus, 1958; Parrillo & Donoghue, 2005).

1.2. Linguistic Bias Underlying Prejudice

Construal level theory offers a theoretical foundation to extract perceived social distance towards outgroup members from text: The more psychologically distant an object is from the egocentric self (in terms of time, space, social relations, or hypotheticality), the more abstract the mental representation of that object (Trope & Libermann, 2010). It follows from this perspective that people who lack direct experience with an outgroup will have a more abstract construal of its members. Both laboratory and natural experiments support this prediction. For example, in their analysis of around 700,000 Twitter feeds, Snefjella and Kuperman (2015) found that, in general, language became more abstract (referring to less concrete, tangible, and imageable information) as people moved from describing family to friends to neighbors to coworkers to foreigners. It has also been shown that people use more concrete language when writing from a first-person than from a third-person perspective

(Pronin & Ross, 2006), indicating that concrete language is more likely to reflect social proximity. A recent study on dehumanization of immigrants found that in a task judging punishment for illegal activity, people who would give immigrants a longer jail sentence also describe immigrants in more impersonal pronouns (e.g., "it," "who"; Markowitz & Slovic, 2020).

Another line of research with direct focus on implicit verbal expression of prejudice shows that abstract language may also be the result of prejudice. Although they may not be aware of it, prejudice can influence the words people choose to use. For example, people tend to use more abstract language when describing stereotype-consistent behaviors than when describing stereotype-inconsistent behaviors (linguistic expectancy bias; Wigboldus et al., 2000). This is because abstract expression, as defined by the linguistic category model (Semin & Fiedler, 1988), implies the observed behavior is expected or typical. For example, according to the linguistic category model, the adjective *aggressive* is more abstract than the verb *shout at* because it concerns dispositions rather than referring to a specific object, situation, or behavior. Therefore, "John is *aggressive*" is more abstract than "John *shouted at* me," and implies that aggression is expected and typical of John's disposition. This line of research is logically consistent with what construal level theory suggests: People are more likely to use abstract language when describing socially distant outgroups and the stereotypes associated with them.

1.3. Content of Prejudice in Natural Language

The academic interest in using language to identify ethnic and racial prejudice and stereotypes dates back at least as far as Katz and Braly's (1933) classic work that asked participants to rate national and ethnic groups on a trait checklist. Perhaps responding to rising norms against prejudice, 55 years later Greenwald et al. (1988) developed the implicit association task, a commonly used measure for implicit prejudice that examines the strength of mental association between social groups (e.g., "male") and valenced attributes (e.g., "logical"). Both approaches to prejudice have been productive and inspired thousands of follow-up studies.[1] However, most of these studies were held in laboratory settings; little is known about how people express prejudice and stereotypes towards outgroups in natural environments.

[1] Note that the implicit association task, despite its popularity, has been criticized for its low validity and reliability (see Oswald et al., 2013, for a meta-analysis).

The recent rise in digitalized text has made it possible to quantitatively study large amounts of language produced outside the laboratory. Research from computer science has shown word associations embedded in written texts mirror those learned by humans (Bolukbasi et al., 2016). Some associations are morally neutral (e.g., an association between *flower* and *pleasantness* or *insect* and *unpleasantness*); others that concern gender and race often reflect stereotypes and prejudice. These machine-learned human-like biases are correlated not only with implicit measures of prejudice such as the Implicit Association Test (Bhatia, 2017; Caliskan et al., 2017), but also with historical socio-economic indicators such as employment rate (Garg et al., 2018). Thus, there are strong indicators that large-scale text analysis can be used to reveal both general and detailed perceptions of outgroups, which is our focus here.

2. The Current Study

we analyzed language surrounding 60 U.S. ethnic and religious minority groups using a corpus containing nearly all news articles published in the *New York Times* over a 20-year period, from 1987 to 2007 (Sandhaus, 2008). We constructed a corpus for each group by collating articles that mentioned the corresponding ethnic or religious label (e.g., Mexican, Christian). With this data set, we investigated five related questions, with the first two questions concerning the degree of outgroup prejudice and the last three on its content. First, do linguistic patterns underlie descriptions of minority groups related to social distance? Extending Snefjella and Kuperman (2015)'s work, we examined whether concrete language can reliably predict human ratings of perceived social distance towards U.S. minority groups. The human ratings of social distance were obtained from Parrillo and Donoghue's (2005) survey using the Bogardus scale. Second, is social distance (inferred from language concreteness) related to sentiment? Given that the use of abstract language is related to descriptions of both socially distant outgroups (according to the construal level theory) and stereotype-consistent behavior (linguistic expectancy bias), we hypothesized that minority groups represented in abstract language are more likely to be described negatively and that this negative association between language concreteness and sentiment is a unique feature of language describing minority groups.

Concreteness and sentiment in language make it possible to compare minority groups on two primary dimensions. The cost for such comparability is the lack of granularity into concrete content of language about minority groups. To enhance the granularity of our analysis, our three further questions focused on the specific topics that emerged in articles that highlighted a minority group's outgroup identity by referencing immigration): What are the topics associated with language with explicit reference to immigration? How are these topics distributed across the different minority groups? And, finally, how are immigrant-related topics associated with perceived pleasantness? To answer these questions, we applied Latent Dirichlet Allocation (LDA; Blei et al., 2003) to extract immigrant-related topics from all news articles that contained the word "immigrant" or its inflections. LDA is an unsupervised machine learning algorithm that uses Bayesian inference to cluster language based on underlying patterns (or topics) that best explain corpus structure. We then analyzed the associations between each topic and the 60 U.S. minority groups, as well as the underlying sentiment of each topic. This approach allowed us to tease apart the underlying social contexts that may explain positive or negative sentiment.

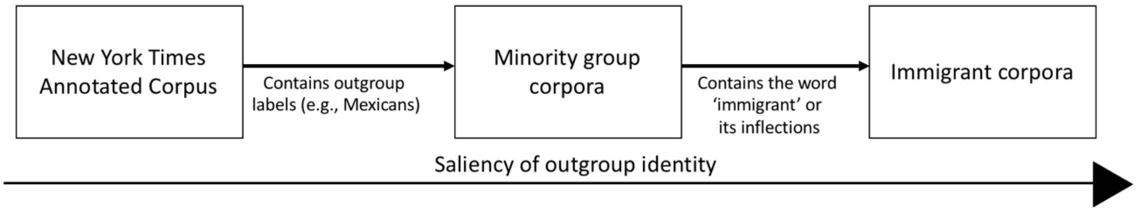
3. Materials and Methods

3.1. Subcorpora of the *New York Times* Annotated Corpus

The *New York Times* Annotated Corpus (Sandhaus, 2008) contains nearly all articles (over 1.8 million) published by the *New York Times* between January 1981 and June 2007. It can be accessed with a license through Linguistic Data Consortium (https://catalog.ldc.upenn.edu/LDC2008T19). We created two types of subcorpora in this study (Figure 1). For each minority group, we constructed a minority group corpus by collating all articles that contained the corresponding group labels (e.g., *Mexican* or *Muslim*). Next, from each minority group corpus, we created an immigrant group corpus by selecting articles that contained at least one occurrence of the word "immigrant" or its inflections.[2] Therefore, for each minority group, its immigrant corpus is a subset of its minority group corpus. The proportion of articles in a minority group corpus that were included in the immigrant corpus ranged from 4% (Australian) to 57% (Guyanese).

Figure 1 *The Two **Types of** Corpora Used in the Current Study*

[2] Inflections of *immigrant* includes *immigrants*, *immigration*, *immigrate*, *immigrated*, *immigrating*, etc.



Articles in minority group corpora may contain information that is not directly related to outgroup identity yet still impacts how the outgroup is represented. For example, news reports on the Tokyo Olympics may not bear any relevance to Japanese diasporas in the United States, but may still have a positive influence on how Japanese people in general are perceived. In contrast, articles in the immigrant corpora, which explicitly reference immigration, are more likely to focus on the identity of the outgroup. We explored language valence and concreteness in both minority group corpora and immigrant corpora. When extracting topics related to outgroups using LDA, we used only the immigrant corpora, since the articles there contained less information that is irrelevant to outgroup identity (e.g., Tokyo Olympics).

Of the 60 minority groups examined in this study, 50 were defined by country or region of origin; we selected the largest 50 groups (each more than 0.8% of the total population) reported in the American Community Survey (U.S. Department of Homeland Security, 2017). The remaining 13 minority groups consisted of eight social categories (e.g., African American, Muslim, Jew) used by Bogardus (1927) and Parrillo and Donoghue (2005) and a further two religious groups (Christian and Buddhist).

3.2. Language Valence and Concreteness

In order to examine features of language used to describe minority groups, we computed the language valence and concreteness for each group. Valence is an affective dimension underlying the meanings of words: Higher valenced words evoke pleasant emotions and lower valenced words evoke unpleasant emotions. Concreteness evaluates the degree to which the concept denoted by a word refers to a perceptible entity. Words like dog and computer are more vividly imagined than words like truth and feeling, and people easily report this difference. In this study, we retrieved valence and concreteness norms from Hollis et al. (2017), whose data set contains valence and concreteness ratings for 78,286 English words. The ratings are based on a well validated computational approach to extrapolating valence and concreteness information from human-rated scores of valence (Bradley & Lang, 1999; Warriner et al., 2013) and concreteness (Brysbaert et al., 2014).

We computed the language valance and concreteness for each group by averaging valence and concreteness of all words in its corresponding corpus (we did this for minority group corpus and immigrant corpus separately). Previous studies have shown that aggregating valence and concreteness over a large corpus reveals meaningful macro-level patterns that would otherwise be difficult to detect, such as the evolution of American English towards greater learnability (Hills & Adelman, 2015; Snefjella et al., 2019) and changes in national well-being over history (Hills et al., 2019).

3.3. Human-Rated Perceived Social Distance

To examine whether linguistic features of language describing outgroups reflects perceived social distance, we obtained human-rated perceived social distance from Parrillo and Donoghue (2005). They used the Bogardus social distance scale (Bogardus, 1927), in which participants were asked to evaluate their willingness to take members of the social group in question into their social circles at various degrees of intimacy. Social circles range from close relatives and personal friends to foreign visitors. One typical question was "Would you be willing to have a member of this group as your colleague at work?"

3.4. Topic Modelling

In the second part of this study, we used LDA to uncover the content of outgroup prejudice. LDA assumes that a set of latent patterns (or topics) explains and generates the structure of textual documents. It computes the distribution of topics over documents, with topics represented as distributions of words. We trained LDA on the immigrant corpora such that each news article was assigned a distribution of topics, and each topic consisted of a distribution of words.[3] For instance, "dangerous illegal workers" may be translated to "10 2 2," indicating that the last two words were generated by topic 2 and the first by topic 10. The same word may be assigned to different topics, allowing generic words (e.g., *make*, *take*) to appear in multiple topics.

We set the topic number to 15 to ensure that the model was sufficiently simple (to avoid overfitting) while providing adequate topic resolution (e.g., to avoid assigning different content to the same topic). No consensus has yet been reached on a nonarbitrary solution for

[3] We used R lda library (Chang, 2012) to train the LDA model for multiple numbers of topics (from 10 to 20) using 1,000 iterations. The hyperparameters alpha and beta were set to 0.01 to encourage the model to assign topics to documents such that each document was composed of a few topics and to learn topics that produce a few words with high probability.

determining the optimal topic number. In our analysis, the number of topics was chosen to maximize interpretability.

We examined the 10 most relevant words for each topic. We defined the relevance of term *w* to topic *k* (Sievert & Shirley, 2014) as:

$$\gamma(w, k|\lambda) = \lambda \log P(w|k) + (1 - \lambda) \log \frac{P(w|k)}{P(w)}, \qquad (1)$$

where $P(w|k)$ is the probability that term *w* is assigned to topic *k* and $P(w)$ is the marginal probability of term *w* in the corpus. The first component of the equation, $P(w|k)$, prioritizes terms with high frequency in a topic. However, it does not consider how unique term *w* is to topic *k*, which can be captured by $\frac{P(w|k)}{P(w)}$, a quantity that Taddy (2012) called *lift*. We set λ to 0.5 to take both components into consideration; λ determines the weight given to the probability of term *w* under topic *k* relative to its lift.

3.4.1. Topic Specificity

We used Equation 2 to compute the specificity of topic *k* to the immigrant corpus compared with the corpus as a whole:

$$\text{Specificity(k)} = \sum_{i=1}^{n} \left(\frac{\gamma(w_i|k)}{\sum_{i=1}^{n} \gamma(w_i|k)} * \frac{P(w_i|immigrant\ corpus)}{P(w_i|New\ York\ Times\ \ corpus)} \right), \qquad (2)$$

where $\frac{\gamma(w_i|k)}{\sum_{i=1}^{n} \gamma(w_i|k)}$ is the normalized relevance of word w_i to topic *k*, and $\frac{P(w_i|immigrant\ corpus)}{P(w_i|New\ York\ Times\ corpus)}$ is the ratio of the frequency of word *w* in the immigrant corpus to its frequency in the *New York Times* corpus. Specificity can range from 0 to near infinity. A specificity of 1 means that, on average, the words characterizing the topic have the same frequency in both the immigrant corpus and the *New York Times* corpus. Higher topic specificity suggests that they are more likely to occur in the immigrant corpus than elsewhere.

3.4.2. Topic Valence and Concreteness

LDA assigned one topic to each word token. Therefore, a topic can be represented as a probability distribution of words. We computed topic valence and concreteness by a probability-weighted averaging of the valence and concreteness ratings of the individual words assigned to each topic by LDA.

3.4.3. Associations Between Immigrant-Related Topics and Minority Groups

To determine the strength of association between immigrant topics and each minority group (e.g., whether the topic "illegal workers" is associated more closely with Mexicans or Japanese people), we computed the document-normalized probability distribution of words in immigrant corpora over the 15 topics, with the association between an immigrant group and topic *t* being

$$l_t = \frac{\sum_{d \in D} P_{dt}}{\sum_{t \in T} \sum_{d \in D} P_{dt}} , \qquad (3)$$

where *d* is a document from an immigrant group corpus *D, t* is one of the 15 topics, and P_{dt} is the proportion of words in document *d* assigned to topic *t*.

4. Results

4.1. Linguistic Footprints of Prejudice

4.1.1. Relationship Between Linguistic Features and Social Distance

First, we examined whether human-rated perceived social distance of the 30 social and religious groups in Parrillo and Donoghue's 2005 study was reflected in the linguistic features underlying language referring to minority groups. Comparing the valence and concreteness of the language in both minority group corpora and immigrant corpora, we found that both valence and concreteness were strongly correlated with Parrillo and Donoghue's survey of social distance (Table 1). Although the construal level theory holds that concreteness is a more direct factor underlying perceived social distance, we found that valence was more strongly correlated with human-rated social distance. This is not entirely surprising because instead of capturing actual interpersonal contact with minority groups, Parillo and Donaghue's survey used hypothetical questions to capture willingness to contact, and thereby reflected a mixture of perceived social distance, affective feelings towards minority groups, and possibly moral considerations.

The key distinction between our minority group corpora and the associated immigrant corpora is whether the group was mentioned in the context of immigration. Using explicit ethnic or religious labels in a text about a minority group clearly signals that the text is about an outgroup; referring to immigration further amplifies that signal. Correspondingly, we found that the correlation between linguistic features and human-rated social distance was stronger in immigrant corpora than in minority group corpora (Table 1).

Table 1

Correlation Between Linguistic Features and Human-Rated Social Distance

	Minority group corpora (N=30)	Immigrant corpora (N=30)
Valence	-0.68***	-0.72***
Concreteness	-0.37*	-0.55**

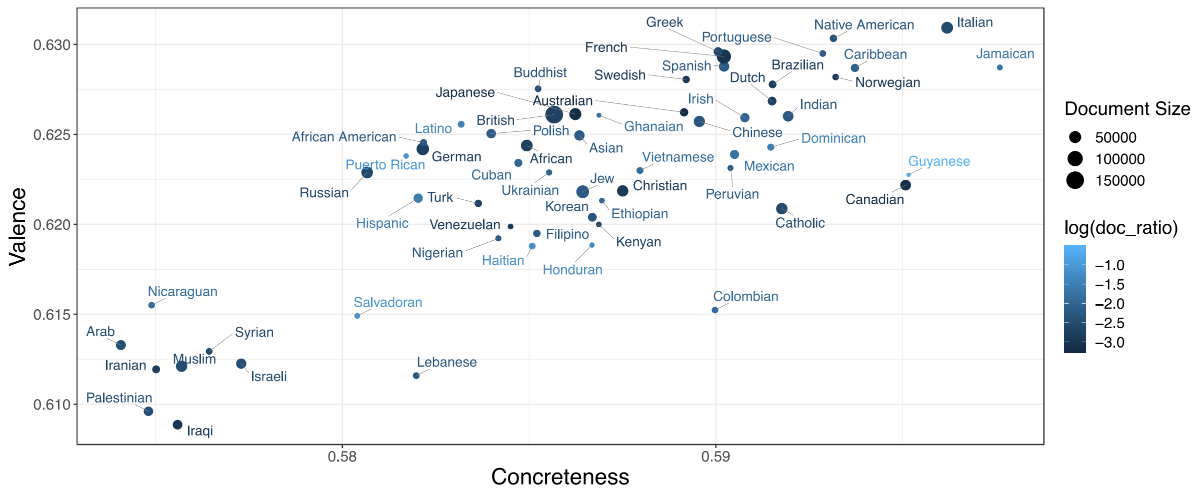
Note. * p<0.05, ** p<0.01, *** p<0.001.

4.1.2. Relationship Between Valence and Concreteness

We extracted language valence and concreteness for each minority group corpus. The group described in the most positive terms was Italian; the group described in the least positive terms was Iraqi. The Italian group also had a high concreteness rating, while that of the Iraqi group was low. Indeed, across groups, the language associated with more positively viewed groups was reliably more concrete, r (59) = 0.77, p < 0.001, 95% CI = 0.64–0.86 (Figure 2). This strong correlation also held when we used the immigrant corpora to compute language valence and concreteness for each group, r (59) = 0.65, p < 0.001, 95% CI = 0.48–0.78 (Appendix Figure S1).

Figure 2

Relationship Between Valence and Concreteness of Language in the Minority Group Corpora



Note. Dot size represents number of articles in the corpus. Color denotes immigrant status, operationalized as a log-transformed ratio between size of an immigrant corpus and size of its corresponding minority group corpus.

We can rule out an alternative explanation for these findings—that the strong linear correlation between valence and concreteness is a linguistic property of the English language. At the individual word level, relation between valence and concreteness is likely to be nonlinear. For instance, there is only a weak positive correlation between valence and concreteness across the 13,384 English words in the Warriner et al. (2013) data set, Pearson's r (13,383) = 0.10, p < 0.001, 95% CI = 0.08–0.11. In contrast, both linguistic and neuroscience studies find that abstract words are more emotionally loaded while concrete words are more likely to be emotionally neutral (Kousta et al., 2011; Vigliocco et al., 2014). Most importantly, when we computed valence and concreteness for each article instead of aggregating across all articles in a minority group corpora, the correlation between valence and concreteness of these articles was only 0.16 (r [1,260,046] = 0.16, $p < 0.001$, 95% CI = 0.16–0.17). Therefore, the substantial correlations we found across minority groups are unlikely to be an artefact of linguistic properties of the English language.

We also explored two further alternative explanations. The first was media exposure, operationalized in terms of the number of articles mentioning the respective target group. If social contact reduces intergroup prejudice, frequency of exposure to outgroup information may achieve a similar effect. The second was that a disproportionate emphasis on immigrant status may be associated with more negative attitudes. We operationalized immigrant status as the ratio between the number of articles mentioning a minority group in immigrant contexts (size of immigrant corpus) and the number of articles mentioning that minority group (size of minority corpus).

We controlled for both above factors in two regression models that predicted valence using concreteness. We did this separately for minority group corpora and immigrant corpora (Table 2). In the first regression model, we included year as a fixed effect (Table 2, “Year fixed-effect”) in order to control for potential biases generated by shocks common to all minority groups in a given year (e.g., the 9/11 terrorist attack in 2001). In other words, introducing year fixed effects allowed us to examine the relationship between valence and concreteness for all minority groups within each year. For both corpora, the strong positive relationship between valence and concreteness was robust to the introduction of year as a fixed effect, as well as to the inclusion of media exposure and immigrant status.

Introducing group fixed effects in the second regression model allowed us to explore the relationship between valence and concreteness for each minority group over the 20 years (Table 2, “Group-specific trends”). The results from both corpora suggest that the positive relationship between valence and concreteness is weaker at the intragroup level. This may be because the limited time span covered by the corpus was too short to encompass large changes in public perceptions towards minority groups. The large difference between marginal R^2 (variance explained by fixed effects) and conditional R^2 (variance explained by fixed effects and random effects) suggests that the majority of the variance was not explained by intragroup differences. Lastly, the coefficient of year ($\beta = 0.03, 95\%, CI = 0.03 - 0.04$) indicates that the sentiment towards minority groups became more positive over time. In sum, the relationship between valence and concreteness stands up to various statistical checks.

Table 2

Language Concreteness Predicts Valence

	Minority Group Corpora		Immigrant Corpora	
	Year fixed effect	Group-specific trends	Year fixed effect	Group-specific trends
Concreteness	0.65*** (0.61 - 0.69)	0.15*** (0.12 - 0.19)	0.51*** (0.46 - 0.56)	0.23*** (0.18 - 0.27)
Exposure (Corpus size)	0.08 ** (0.03 - 0.13)	-0.37*** (-0.44 - -0.30)	0.13 *** (0.08 - 0.17)	-0.06 (-0.15 - -0.02)
Immigrant Status	-0.05 * (-0.10 - 0.00)	-0.04* (-0.09 - 0.00)	-0.19 *** (-0.24 - -0.14)	-0.16*** (-0.23 - -0.09)
Year	-	0.03*** (0.03 - 0.04)	-	0.03*** (0.02 - 0.04)
Marginal R^2	0.43	0.17	0.26	0.12
Conditional R^2	0.44	0.89	0.28	0.59

Note. The dependent variable is valence per minority group per year. Variables are normalized so that they are all centered at 0 with standard deviation equaling 1. The 95% confidence intervals are included inside the parentheses.
p* < 0.05, ** *p* < 0.01, * *p* < 0.001.

4.2. Immigrant-Related Topics

Next, we investigated the content of outgroup prejudice by applying LDA to extract topics from language referencing minority groups in the context of immigration, which highlights their outgroup identity. Table 3 shows the 10 most relevant words in each topic (see Equation 1 for a definition of relevancy of words to a topic). Keywords for a particular topic were strongly associated with each other and were clearly distinguishable from keywords of other topics. We labelled the topics by summarizing their top 20 keywords. The results indicate a wide array of topics. Crime, terrorism, and geopolitical conflict were among the most negative topics and museums, music and movies, and restaurants were among the most positive. These topics reflect many of the issues commonly associated with immigration (Alexander et al., 1999; Borjas, 1990; Carens, 2013; Cuddy et al., 2007; Peabody, 1985; Poppe, 2001; Skeldon, 1997).

Table 3

Top 10 Keywords for Each Topic (From Most Negative to Most Positive)

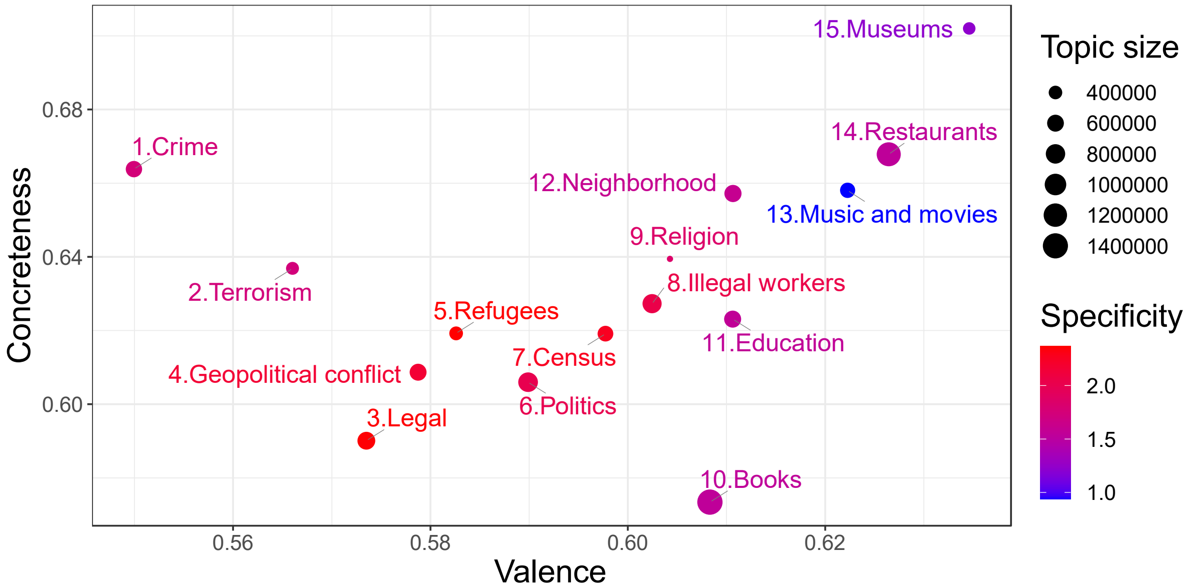
Index	Topic	Keywords
1	Crime	police, officer, arrest, charge, prosecutor, drug, kill, gang, crime, shoot.
2	Terrorism	Muslim, terrorist, bomb, attack, intelligence, Islamic, FBI, mosque, Sept, Iraq
3	Legal	immigration, law, court, alien, judge, legal, justice, case, federal, lawyer
4	Politics	Republican, Bush, Democrat, bill, president, vote, senate, senator, campaign, Clinton
5	Geopolitical conflict	Israel, minister, Soviet, France, Germany, Europe, party, prime, Palestinian, Jew
6	Refugees	refugee, Cuban, asylum, Haitian, unite, Miami, boat, Castro, state, official
7	Illegal workers	worker, border, Mexican, company, labor, job, wage, work, pay, illegal
8	Census	Hispanic, population, percent, Asian, Black, census, Chinese, Korean, Latino, immigrant
9	Neighborhood	city, build, house, neighborhood, county, resident, island, apartment, rent, community
10	Books	write, book, life, American, world, think, history, story, time, way
11	Religion	church, Catholic, Irish, bishop, priest, Jewish, religious, parish, pope
12	Education	school, student, child, teacher, education, parent, program, health, care, college
13	Restaurants	restaurant, cook, eat, chicken, room, shop, soccer, dish, food, cup
14	Music & movies	theater, film, music, movie, play, art, direct, musical, dance, song, artist
15	Museums	museum, Sunday, tour, street, information, tomorrow, admission, exhibition, park, sponsor

Note. We combined inflections (e.g., *German*, *Germany*) to avoid unnecessary duplications. An interactive visualization of topic–word association with varying degrees of lambda can be accessed at https://liyingpsych.github.io/LanguageOfPrejudice/.The visualization was generated by R package LDAvis (Sievert & Shirley, 2014). Lambda was set to 0.3 when displaying keywords for topic 13 (Restaurants) because this topic was mixed with generic linguistic patterns underlying all articles (e.g., *say, like, one, day, get, come*). Reducing lambda further penalizes the weight of high frequency words that tend to appear across all articles.

Next, we analyzed three linguistic features of the topics: valence, concreteness, and topic specificity (Figure 3). Topic valence and concreteness were computed by the average of all words assigned to the given topic. We found no significant correlation between topic valence and concreteness, $r(13) = 0.36$, $p = 0.17$. Some topics that are per se more concrete (e.g., crime and terrorism) are not highly positive; similarly, positive topics are not necessarily to be more concrete (e.g., books). Thus, the strong correlation between language valence and concreteness across minority groups shown in Figure 2 was not the result of language distributed across topics. More specifically, if a group was associated with a concrete negative topic such as crime, it also tended to be associated with other topics featured by abstract language. Negative discussion of minority groups and abstract language tended to go hand-in-hand.

Figure 3

Valence and Concreteness of the 15 Immigrant Topics Identified Using LDA



Note. Dot size corresponds to number of words assigned to each topic. Color represents topic specificity, with higher values indicating that the topic was more likely to occur in the immigrant corpus than elsewhere in the *New York Times* corpus.

Topic specificity represents the strength of association between topics and immigration (see definition in Equation 2). It is clear from Figure 3 that some topics are highly specific to immigration, such as refugees and illegal workers, while others like museums and music and movies are less specific. We found that topic specificity was negatively correlated with valence, $r(13) = -0.60$, $p < 0.05$, 95% CI = −0.85 – −0.13, and concreteness, Pearsons' $r(13) = -0.59$, $p < 0.05$, 95% CI = −085 – −0.11. In other words, language was more abstract and negative when it was more specific to immigrant-related topics.

Figure 4 (color version)

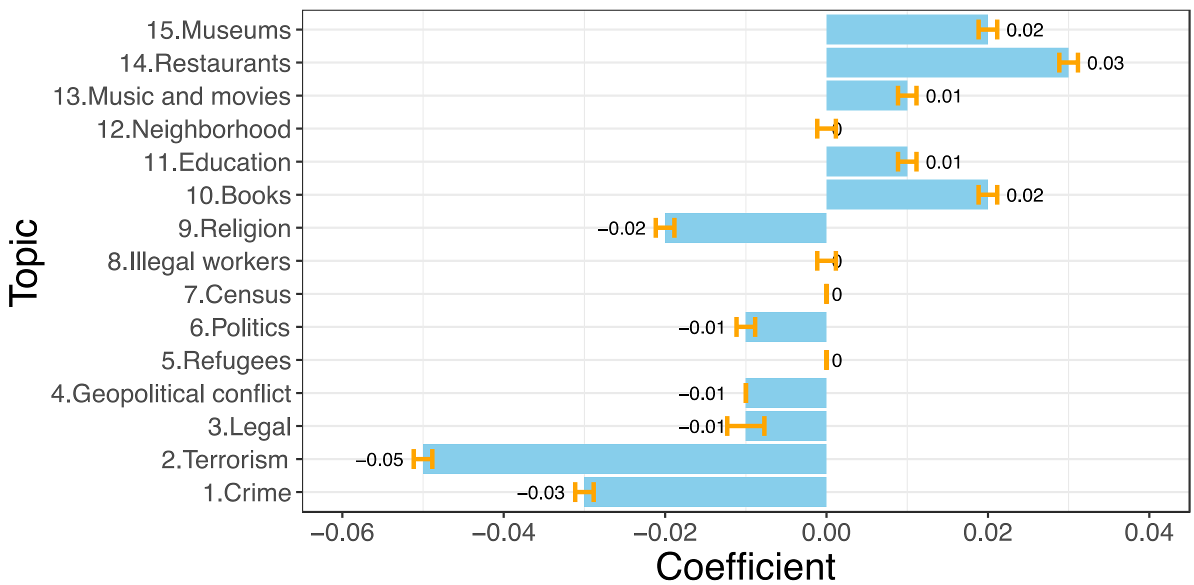
Distributions of Topics Over Groups Ranked by Valence

Note. The topics identified in Table 3 are plotted on the x-axes. The y-axes show the normalized weighting of each topic on each minority group. Topics are arranged by valence, with the lowest (red) on the left and the highest (green) on the right. Minority groups are also ranked by overall valence, with the most negative in the top left corner and the most positive in the bottom right.

To understand the association between each minority group and immigrant-related topics, we computed the document-normalized probability distribution of words in immigrant corpora over the 15 topics (see Equation 3). Figure 4 presents associations between the 15 immigrant-related topics and each minority group. Unsurprisingly, groups described in negative language (and also perceived as more socially distant) were associated primarily with negative topics. The negative topics varied across minority groups. For example, the Iraqi, Palestinian, Lebanese, and Syrian groups were represented mostly in terrorism and geopolitical conflict; the Cuban, Nicaraguan, Vietnamese, and Venezuelan groups were closely associated with refugees; and the Mexican group was mentioned primarily in illegal workers. In contrast, groups described in positive language (also perceived as socially proximal) were closely associated with positive, less immigrant-specific topics (e.g., restaurants, museums, music and movies) and were rarely represented in negative topics. Two minority groups, Native Americans and African Americans, cannot be classified as immigrants in the United States. It is therefore unsurprising that their associated immigrant-related topics ranked low on topic specificity (e.g., books, museums).

Figure 5

Using Association With Immigrant-Related Topics to Predict the Valence of Minority Groups



Note. Regression coefficients are from an averaged linear regression model. Error bars represent the 95% confidence interval.

To assess which immigrant topics had the largest impact on sentiments toward minority groups, we regressed the valence of each minority group (inferred from the minority group corpora) on its association with 15 immigrant-related topics. As the model contained

15 independent variables and just 60 data points, we used elastic net regularization, a combination of lasso regression and ridge regression. These techniques perform simple linear least squares regression but penalize the coefficients of the inputs based on their size. The penalty forces some regression coefficients to zero. We cross-validated our findings by dividing our data set into 10 equal groups, training our model on a random sample of seven groups, and predicting immigrant sentiment in the remaining three. This cross-validation exercise was repeated 1,000 times to calculate the average adjusted R^2 for the out-of-sample predictions and average regression coefficients. A total of 78% of the variance in sentiment toward minority groups can be explained by topic profiles of individual groups. Overall, the negative topics had a stronger impact on sentiment than the positive topics did (Figure 5). Three negative topics—crime, terrorism, and legal—significantly predicted negative sentiment toward immigrants. Restaurants was the topic most strongly predictive of positive sentiment. Politics, geopolitical conflict, refugees, illegal workers, and religion did not significantly predict sentiment toward immigrants.

5. Discussion

Our study makes a number of additional contributions to research on prejudice and stereotypes. First, we found that perceived social distance towards outgroups is reflected in language: Socially distant groups are more likely to be described in abstract and negative language. Second, there is a clear linguistic bias underlying media representations of minority groups; some groups are represented in much more negatively valenced contexts than others are. Third, we found a strong positive correlation between valence and concreteness that is unique to language concerning minority groups, suggesting a potential cognitive bias when communicating narratives of outgroup members. Lastly, we uncovered the content of outgroup prejudice and showed how those topics explain why some groups were represented move positively than others.

Our approach reveals rich diversity within outgroups. Although they are all minority groups, they differ substantially in terms of sentiment, perceived social distance, and the content of prejudice. Classic theories on outgroup negativity has often focused on an ingroup-versus-outgroup dichotomy, thus overlooking differences among outgroups—a cognitive bias that these prejudice theories have themselves identified as one of the symptoms of outgroup bias. In contrast, more recent work from Fiske et al. (2002) highlights how stereotypes can be different for each outgroup, proposing that outgroups are perceived along two basic dimensions: warmth and competence (the stereotype content model). We complement Fiske et al.'s (2002) work by offering a quantitative measure of social distance and the topic model approach to identify further distinctions in the qualitative content of prejudice.

The fact that our findings on social distance are largely consistent with the survey results of Parrillo and Donoghue (2005) suggests that our corpus approach captures meaningful patterns despite its possible limitations. As the second largest news distributor in the United States, with its headquarters in a metropolitan city, the *New York Times* is well positioned to offer wide coverage of issues concerning ethnic and religious minorities and to influence its readers' attitudes toward outgroups. Nevertheless, it is unlikely to represent the full diversity of public opinion. We further acknowledge that the topics identified may vary across media targeting different audiences. However, given the established theory on which we frame our approach, the relationship we found between social distance and sentiment represents a hypothesis about language that may exist in other contexts, such as everyday conversations or communication on social media. Unlike articles in the *New York Times*, face-to-face conversations and comments on social media are not bound by style and editorial rules to use formal and politically correct language. It is therefore likely that socially distant outgroups are associated with even more negative language in these channels.

Overall, we believe that the strengths of the corpora approach outweigh its limitations. These strengths include (a) ecological validity, achieved by studying perceptions of immigrants outside the laboratory, thus avoiding problems such as socially desirable response bias, (b) tracking perceived social distance and sentiment in a data set referencing a large variety of minority groups, and (c) providing a valid social scientific approach with specific language patterns to potentially flag certain outgroups at greater risk of prejudice.

References

Alexander, M. G., Brewer, M. B., & Hermann, R. K. (1999). Images and affect: A functional analysis of out-group stereotypes. *Journal of Personality and Social Psychology*, *77*(1), 78–93. https://doi.org/10.1037/0022-3514.77.1.78

Allport, G. W. (1954). *The nature of prejudice*. Addison-Wesley.

Bhatia, S. (2017). The semantic representation of prejudice and stereotypes. *Cognition*, *164*, 46–60. https://doi.org/10.1016/j.cognition.2017.03.016

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, *3*, 993–1022.

Bogardus, E. S. (1927). Race friendliness and social distance. *Journal of Applied Sociology, 11*, 272–287.

Bogardus, E. S. (1958). Racial distance changes in the United States during the past 30 years. *Sociology & Social Research, 43,* 127–134.

Bolukbasi, T., Chang, K.-W., Zou, J. Y., Saligrama, V., & Kalai, A. T. (2016, December 5–10). Man is to computer programmer as woman is to homemaker? Debiasing word embeddings. In D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, & R. Garnett, *Advances in Neural Information Processing Systems 29* (pp. 4349–4357). http://papers.nips.cc/paper/6228-man-is-to-computer-programmer-as-woman-is-to-homemaker-debiasing-word-embeddings.pdf

Borjas, G. J. (1990). *Friends or strangers: The impact of immigrants on the U.S. economy*. Basic Books.

Bradley, M. M., & Lang, P. J. (1999). *Affective Norms for English Words (ANEW): Instruction manual and affective ratings* (Tech. Rep. No. C-1). Gainesville, FL: University of Florida, The Center for Research in Psychophysiology.

Brewer, M. B. (1979). In-group bias in the minimal intergroup situation: A cognitive-motivational analysis. *Psychological Bulletin, 86*(2), 307–324. https://doi.org/10.1037/0033-2909.86.2.307

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods, 46*(3), 904–911. https://doi.org/10.3758/s13428-013-0403-5

Caliskan, A., Bryson, J. J., & Narayanan, A. (2017). Semantics derived automatically from language corpora contain human-like biases. *Science, 356*(6334), 183–186. https://doi.org/10.1126/science.aal4230

Carens, J. H. (2013). *The ethics of immigration*. Oxford University Press.

Chang, J. (2012). *lda: Collapsed Gibbs sampling methods for topic models* (Version 1.4.2). R Package. https://rdrr.io/cran/lda/

Collins, K. A., & Clément, R. (2012). Language and prejudice: Direct and moderated effects. *Journal of Language and Social Psychology*, *31*(4), 376–396. https://doi.org/10.1177/0261927X12446611

Cuddy, A. J. C., Fiske, S. T., & Glick, P. (2007). The BIAS map: Behaviors from intergroup affect and stereotypes. *Journal of Personality and Social Psychology, 92*(4), 631–648. https://doi.org/10.1037/0022-3514.92.4.631

Cuddy, A. J. C., Fiske, S. T., Kwan, V. S. Y., Glick, P., Demoulin, S., Leyens, J.-P., Bond, M. H., Croizet, J.-C., Ellemers, N., Sleebos, E., Htun, T. T., Kim, H.-J., Maio, G., Perry, J., Petkova, K., Todorov, V., Rodríguez-Bailón, R., Morales, E., Moya, M., …Ziegler, R.. (2009). Stereotype content model across cultures: Towards universal similarities and some differences. *British Journal of Social Psychology, 48*(1), 1–33. https://doi.org/10.1348/014466608X314935

Fennelly, K., & Federico, C. (2008). Rural residence as a determinant of attitudes toward US immigration policy. *International Migration*, *46*(1), 151–190. https://doi.org/10.1111/j.1468-2435.2008.00440.x

Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology, 82*(6), 878–902. https://doi.org/10.1037/0022-3514.82.6.878

Garg, N., Schiebinger, L., Jurafsky, D., & Zou, J. (2018). Word embeddings quantify 100 years of gender and ethnic stereotypes. *Proceedings of the National Academy of Sciences, 115*(16), E3635–E3644. https://doi.org/10.1073/pnas.1720347115

Greenwald, A. G., McGhee, D. E., and Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology, 74*(6), 1464–1480. https://doi.org/10.1037//0022-3514.74.6.1464

Hills, T. T., & Adelman, J. S. (2015). Recent evolution of learnability in American English from 1800 to 2000. *Cognition, 143*, 87–92. https://doi.org/10.1016/j.cognition.2015.06.009

Hills, T. T., Proto, E., Sgroi, D., & Seresinhe, C. I. (2019). Historical analysis of national subjective wellbeing using millions of digitized books. *Nature Human Behaviour*, *3*(12), 1271–1275. https://doi.org/10.1038/s41562-019-0750-z

Hollis, G., Westbury, C., & Lefsrud, L. (2017). Extrapolating human judgments from skip-gram vector representations of word meaning. *Quarterly Journal of Experimental Psychology, 70*(8), 1603–1619. https://doi.org/10.1080/17470218.2016.1195417

Jones, E. E., & Nisbett, R. E. (1987). The actor and the observer: Divergent perceptions of the causes of behavior. In E. E. Jones, D. E. Kanouse, H. H. Kelley, R. E. Nisbett, S. Valins, & B. Weiner (Eds.), *Attribution: Perceiving the causes of behavior* (pp. 79–94). Lawrence Erlbaum Associates.

Katz, D., & Braly, K. (1933). Racial stereotypes of one hundred college students. *The Journal of Abnormal and Social Psychology*, *28*(3), 280–290. https://doi.org/10.1037/h0074049

Kousta, S.-T., Vigliocco, G., Vinson, D. P., Andrews, M., & Del Campo, E. (2011). The representation of abstract words: Why emotion matters. *Journal of Experimental Psychology: General, 140*(1), 14–34. https://doi.org/10.1037/a0021446

Maass, A., Salvi, D., Arcuri, L., & Semin, G. R. (1989). Language use in intergroup contexts: The linguistic intergroup bias. *Journal of Personality and Social Psychology, 57*(6), 981–993. https://doi.org/10.1037/0022-3514.57.6.981

Markowitz, D. M., & Slovic, P. (2020). Social, psychological, and demographic characteristics of dehumanization toward immigrants. *Proceedings of the National Academy of Sciences, 117*(17), 9260–9269. https://doi.org/10.1073/pnas.1921790117

Oswald, F. L., Mitchell, G., Blanton, H., Jaccard, J., & Tetlock, P. E. (2013). Predicting ethnic and racial discrimination: A meta-analysis of IAT criterion studies. *Journal of Personality and Social Psychology*, *105*(2), 171–192. http://dx.doi.org/10.1037/a0032734

Paluck, E. L., Green, S. A., & Green, D. P. (2019). The contact hypothesis re-evaluated. *Behavioural Public Policy, 3*(2), 129–158. https://doi.org/10.1017/bpp.2018.25

Parrillo, V. N., & Donoghue, C. (2005). Updating the Bogardus social distance studies: A new national survey. *The Social Science Journal, 42*(2), 257–271. https://doi.org/10.1016/j.soscij.2005.03.011

Peabody, D. (1985). *European monographs in social psychology. National characteristics*. Cambridge University Press; Editions de la Maison des Sciences de l'Homme.

Pettigrew, T. F. (1979). The ultimate attribution error: Extending Allport's cognitive analysis of prejudice. *Personality and Social Psychology Bulletin, 5*(4), 461–476. https://doi.org/10.1177/014616727900500407

Pettigrew, T. F., & Tropp, L. R. (2006). A meta-analytic test of intergroup contact theory. *Journal of Personality and Social Psychology*, *90*(5), 751–783. https://doi.org/10.1037/0022-3514.90.5.751

Poppe, E. (2001). Effects of changes in GNP and perceived group characteristics on national and ethnic stereotypes in Central and Eastern Europe. *Journal of Applied Social Psychology, 31*(8), 1689–1708. https://doi.org/10.1111/j.1559-1816.2001.tb02746.x

Pronin, E., & Ross, L. (2006). Temporal differences in trait self-ascription: When the self is seen as an other. *Journal of Personality and Social Psychology, 90*(2), 197–209. https://doi.org/10.1037/0022-3514.90.2.197

Sandhaus, E. (2008). The New York Times Annotated Corpus LDC2008T19. Linguistic Data Consortium. https://catalog.ldc.upenn.edu/LDC2008T19

Sherif, M., Harvey, O. J., White, B. J., Hood, W. R., & Sherif, C. W. (1961). *Intergroup conflict and cooperation: The Robbers Cave experiment.* University of Oklahoma Book Exchange.

Sievert, C., & Shirley, K. E. (2014). LDAvis: A method for visualizing and interpreting topics. In J. Chuang, S. Green, M. Hearst, J. Heer, & P. Koehn (Eds.). *Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces* (pp. 63–70). Association for Computational Linguistics. https://www.aclweb.org/anthology/W14-3110.pdf

Skeldon, R. (1997). *Migration and development: A global perspective* (1st ed.). Routledge. https://doi.org/10.4324/9781315843346

Snefjella, B., & Kuperman, V. (2015). Concreteness and psychological distance in natural language use. *Psychological Science, 26*(9), 1449–1460. https://doi.org/10.1177/0956797615591771

Snefjella, B., Généreux, M., & Kuperman, V. (2019). Historical evolution of concrete and abstract language revisited. *Behavior Research Methods*, *51*(4), 1693–1705. https://doi.org/10.3758/s13428-018-1071-2

Taddy, M. (2012). On estimation and selection for topic models. In N. D. Lawrence & M. Girolami (Eds.), *Proceedings of the Fifteenth International Conference on Artificial Intelligence and Statistics* (pp. 1184–1193). http://proceedings.mlr.press/v22/taddy12/taddy12.pdf

Tajfel, H. (1982). Social psychology of intergroup relations. *Annual Review of Psychology, 33*(1), 1–39.

Tajfel, H., Billig, M. G., Bundy, R. P., & Flament, C. (1971). Social categorization and intergroup behaviour. *European Journal of Social Psychology*, *1*(2), 149–178. https://doi.org/10.1002/ejsp.2420010202

Trope, Y., & Liberman, N. (2010). Construal-level theory of psychological distance. *Psychological Review, 117*(2), 440–463. https://doi.org/10.1037/a0018963

U.S. Department of Homeland Security. (2017). *Yearbook of immigration statistics 2017.* https://www.dhs.gov/immigration-statistics

Vigliocco, G., Kousta, S.-T., Della Rosa, P. A., Vinson, D. P., Tettamanti, M., Devlin, J. T., & Cappa, S. F. (2014). The neural representation of abstract words: The role of emotion. *Cerebral Cortex, 24*(7), 1767–1777. https://doi.org/10.1093/cercor/bht025

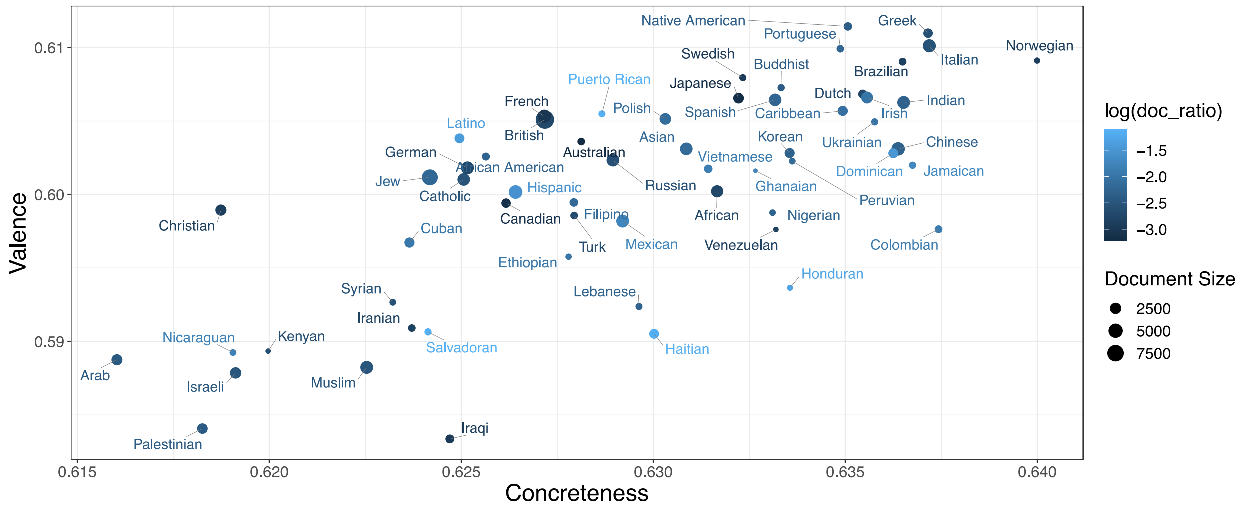
Warriner, A. B., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods, 45*(4), 1191–1207. https://doi.org/10.3758/s13428-012-0314-x

Wigboldus, D. H. J., Semin, G. R., & Spears, R. (2000). How do we communicate stereotypes? Linguistic bases and inferential consequences. *Journal of Personality and Social Psychology, 78*(1), 5–18. https://doi.org/10.1037/0022-3514.78.1.5

Appendix

Figure S1

Relationship Between Valence and Concreteness of Language in the Immigrant Corpora



Note. Dot size represents corpus size.