Extracting Associations of Intersectional Identities with Discourse about Institution from Nigeria

Pavan Kantharaju and **Sonja Schmer-Galunder**
SIFT
319 N. 1st Avenue, Suite 400 Minneapolis, MN 55401-1689
{pkantharaju, sgalunder}@sift.net

Abstract

Word embedding models have been used in prior work to extract associations of intersectional identities within discourse concerning institutions of power, but restricted its focus on narratives of the nineteenth-century U.S. south. This paper leverages this prior work and introduces an initial study on the association of intersected identities with discourse concerning social institutions within social media from Nigeria. Specifically, we use word embedding models trained on tweets from Nigeria and extract associations of intersected social identities with institutions (e.g., domestic, culture, etc.) to provide insight into the alignment of identities with institutions. Our initial experiments indicate that identities at the intersection of gender and economic status groups have significant associations with discourse about the economic, political, and domestic institutions.

1 Introduction

Social scientists have leveraged quantitative methods to extract cultural knowledge from text, such as semantic networks (Hoffman et al., 2018), topic modeling (Mohr and Bogdanov, 2013), and language models (Friedman et al., 2021). Recent work by Nelson (2021) focused on using language models (specifically word embedding models) to extract intersectional identity associations inherent in narrative texts. Intersectionality (Crenshaw, 1989) is a theoretical framework for understanding how social identities such as gender and race, compound to create experiences that would otherwise be obscured by focusing on the identities separately.

Specifically, Nelson (2021) applied Word2Vec (Mikolov et al., 2013) to understand how intersected social identities associate with discourse about institutions of power (e.g. domestic, culture, etc.) from narratives of the nineteenth-century U.S. south. While this method was successfully able to extract intersectional associations from U.S. narratives, it remains an open question whether this method generalizes to other forms of text from outside the U.S., such as social media data from Nigeria. Social media data outside the U.S. presents an interesting challenge as social media may not be accessible or used by everyone outside the U.S. This means that these types of datasets can inherently contain an imbalance in population representation, making analyses with them need careful attention.

This paper presents an initial study on using word embedding models to understand how intersected identities associate with discourse concerning institutions found within social media text from Nigeria. Our main contributions are (1) the application of prior work by Nelson (2021) to tweets from Nigeria, and (2) an analysis of intersected gender and economic identities and their associations to the domestic, economic, political, and cultural spheres. We leverage Skip Gram with Negative Sampling (SGNS) (Mikolov et al., 2013) models and look at the relationship of intersected gender and economic identities within discourse concerning the political, cultural, domestic, and economic spheres within tweets. Our results indicate that a *female, poor* category of individuals is more associated with discourse from Lagos and Federal Capital Territories (FCT) concerning the domestic sphere while a *male, poor* category is associated with economic and political spheres.

This paper is structured as follows. Section 2 provides prior work on intersectionality and the extraction of cultural associations from language models. Section 3 describes the Twitter dataset used to train SGNS models used in our analysis. Section 4 describes the method used by Nelson (2021), which is leveraged for our analysis in Section 5. Section 6 provides a discussion about our analysis and Section 7 provides our conclusion.

2 Related Work

Our analysis is situated at the crossroads of intersectionality and extraction of cultural associations

Proceedings of the Fifth Workshop on Natural Language Processing and Computational Social Science (NLP+CSS), pages 164 - 169
November 7, 2022

from word embedding models. The concept of intersectionality can be traced to Crenshaw (1989), who argued and showed that the experiences of inequality of black women were obscured by the experiences of inequality of women and black people. Both quantitative and qualitative methods have been used to analyze intersectionality. Bright et al. (2016) argues that graphical causal models can be used to represent claims about the causal effects of occupying intersected social identities. A survey of quantitative research that uses the intersectionality framework is provided by Bauer et al. (2021). There has also been qualitative work by Sekoni et al. (2022), who analyzed the intersection of LGBT+ and other social identities in the context of the healthcare in Nigeria, discovering that sub-identities within LGBT+ suffer from bias more than their peer sub-identities, particularly when intersected with mental and sexual health conditions.

Language models have been shown to be effective at extracting cultural associations (Garg et al., 2018; Kozlowski et al., 2019; Nelson, 2021) and bias (Caliskan et al., 2017; May et al., 2019; Zhao et al., 2019; Tan and Celis, 2019; Guo and Caliskan, 2021) from text; our work focuses on extracting cultural associations from text. Garg et al. (2018) studied how word embedding models could be used to understand trends in gender and ethnic stereotypes in the U.S. over time. Kozlowski et al. (2019) studied how word embedding models could be used to construct cultural vectors, and applied this to understand social class in the U.S. The work closest to ours was done by Nelson (2021), who studied how intersectional identities associated with U.S. narratives about institutions of power. Our work differs from Garg et al. (2018), Kozlowski et al. (2019), and Nelson (2021) in that we apply our analysis to texts outside the U.S. (namely Nigeria).

3 Social Media Dataset

Table 1: Twitter Dataset Metrics/Measures

Metric/Measure	Value
Number of Tweets in Dataset	30,883,364
Vocabulary Size	2,000,381
Tweet Length (min)	2
Tweet Length (mean)	19.6
Tweet Length (median)	15.0

The present work leverages language models trained on an international social media dataset used in prior work (Friedman et al., 2019) for the DARPA Understanding Group Bias (UGB) project and approved for use by an independent IRB. Among other countries, the original UGB-gathered dataset includes approximately 30 million tweets from various states in Nigeria from 2018, gathered by a university teammate. This data is not used directly in this work, but the derived word embedding models are. To create the word embedding models for UGB, tweets were tokenized for whitespace and lower-cased. No stemming or lemmatization was performed, thereby preserving the original vocabulary for our analysis (preservation was necessary as the vocabulary affects the seed words used in the analysis).

Table 1 describes the original dataset from which our language models were derived. *Tweet length* measures the words in the tweet after tokenization (Twitter imposes its own character limit). Approximately 0.2% of the tweets in the dataset have a length of 100 or words and 35.2% are greater than or equal to the mean tweet length, so the majority of tweets are relatively short. We note that this dataset has an uneven distribution of tweets per state in Nigeria. More specifically, approximately 60% of tweets come from Lagos, with the Federal Capital Territory (FCT) being a far second (approximately 20%). As such, our analysis will be biased towards views from Lagos and FCT.

4 Extracting Intersectional Associations from Word Embedding Models

The main goal of our analysis is to extract intersectional associations within discourse about social institutions found in tweets from Nigeria. To this end, we leverage recent work by Nelson (2021) which used a Word2Vec model to understand how intersected social identities (black and white men and women) mapped within four social institutions (domestic, economic, polity, and culture) in a corpus of first-person narratives from the U.S. south. The method used by Nelson (2021) required constructing geometric vectors and spaces for the institutions and identities using trained Word2Vec models. This section describes their construction.

4.1 Intersectional Social Identity Vectors

Intersectional identity vectors provide meaning to each intersected social identity in vector space. Our analysis focuses on two social identity groups: *gender* and *economic status*. As such, we will use them in a running example showing how the intersected identity vectors are constructed.

Table 2: Social Identities and Corresponding Seed Words

Identity Category	Social Identity	Seed Words
Gender	Male	men, man, boy, boys, he, him, his, himself
	Female	women, woman, girl, girls, she, her, hers, herself
Economic Status	Rich	rich, richer, richest, affluence, affluent, expensive, luxury, opulent
	Poor	poor, poorer, poorest, poverty, impoverished, inexpensive, cheap, needy

First, the social identity groups *gender* and *economic status* are split into two identities: gender into *male* and *female* and economic status into *rich* (high) and *poor* (low). Each identity is associated with a set of seed words. Table 3 contains the social identities and seed words used in our analysis. Seed words add context about a particular concept to provide a geometric description of the concept. For example, if we wanted to describe the concept of *man*, we would construct a set of seed words corresponding to *men*, *males*, and *boys*. The addition of other seed words would further contextualize the concept and possibly change the description of the concept (i.e., adding seed words associated with *human* would change the description of *man*).

The gender seed words come from Nelson (2021) while the economic status words come from Kozlowski et al. (2019) and Antoniak and Mimno (2021). We focused on these seed words as they were successfully used in prior work on extracting associations from word embedding models; we plan to create our own seed words in future work.

Next, the cross product of the identities and their corresponding seed words is computed, giving us four intersected social identities (in our running example, we get *male rich*, *male poor*, *female rich*, and *female poor*) and a set of word pairs W_{id} (e.g., *(men, rich)*, *(woman, rich)*, etc.) for each intersected identity *id*. The set of word pairs effectively represent a joint space that provide meaning to an intersected identity. To construct an intersected identity vector $\vec{v_{id}}$, the word embeddings in each pair are summed to construct an embedding representing the pair (summing the embeddings for *men* and *rich* provides an embedding for *men rich*), and the pairs are subsequently averaged:

$$\vec{v_{id}} = \frac{1}{|W_{id}|} \sum_{(w_1,w_2)\in W_{id}} \vec{w_1} + \vec{w_2}$$

where $\vec{w_1}$ and $\vec{w_2}$ are word embeddings for words w_1 and w_2. This results in a set of four intersected identity vectors that capture the meaning of the identity in vector space.

4.2 Social Institution Vectors

Table 3: Social Institutions and Corresponding Seed Words (Words in bold are those used by Nelson, 2021)

Social Institutions	Seed Words
Polity	**nation**, government
Economy	**money**, finance
Culture	**culture**, tradition
Domestic	**housework**, **children**

Social institution vectors provide meaning to each social institution in vector space. Each institution vector $\vec{v_{inst}}$ is constructed by defining a set of seed words W_{inst} for each institution *inst*, and averaging the word embeddings of the seed words. Table 3 contains the social institutions and seed words used in our analysis. We use the same set of institutions as those used by Nelson (2021), but we extend their seed words set such that each institution has an equal number of words. These new seed words were curated by the researchers of this paper by looking for related words to the institutions. We focus on these institutions as we wanted to keep as close as possible to the original analysis; we will look at other institutions in future work.

4.3 Social Institution Discourse Spaces

Table 4: Top 10 Words in each Social Institution Discourse Space

Polity	Domestic	Economy	Culture
govt	kids	finance-	traditions
country	baby-sit	vaid	cultural
gov't	house-helps	funds	cultures
goverment	homeworks	recapitalised	religion
governement	childcare	fgns	patriachal
counrty	pre-k	harmonising	unafrican
governments	great-grandchildren	remiting	norms
governnent	under-privileged	countingup	bidia
administration	godchildren	alison-madukwe	supremacism
reponsibility	#mychildmypride	slac	heritages

A discourse space for each social institution is constructed to compute an association score with discourse surrounding the institutions. This space provides a discourse-centric meaning to the social institution compared to the institution vectors from Section 4.2, which provide a concept-centric meaning. More specifically, this discourse space is constructed by finding K words closest to each institution vector (in our work, $K = 50$ and closest is

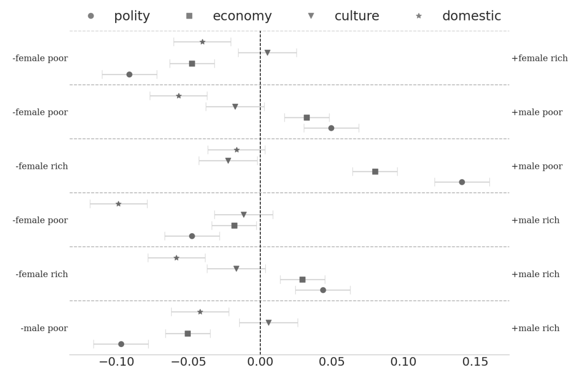

Figure 1: Gender vs Economic Status - 95% Confidence Interval ($n = 40$)

defined by cosine similarity). Table 4 provides the top 10 words in the discourse space for each social institution. Here, we see some challenges with using social media data: words may not always be grammatically correct (e.g., “governement" under *polity* column) and we may have non-word terms such as hashtags (e.g., “#mychildmypride" under *domestic* column). Given a discourse space for an institution, an association score can be computed for any intersected identity by taking the average cosine similarity between the identity vector and the words in the discourse space.

5 Analysis

Figure 1 provides the results of our analysis for the gender and economic status identity groups.[1] For each social institution, we compute a 95% confidence interval for the difference between the association scores (described in Section 4.3) for pairs of intersected identities (e.g., the first row of Figure 1 compares the difference between association scores for *female, poor* and *female, rich* for each social institution). We note that any confidence intervals that contain a difference of 0 (middle black dotted line in Figure 1) is not statistically significant.

Similar to Kozlowski et al. (2019) and Nelson (2021), we use the percentile bootstrap method to construct the confidence intervals, where the number of samples used is 40 (interval spans the 2^{nd} and 39^{th} association score differences). We used 40 pretrained SGNS models that were each trained on datasets generated by sampling the original Twitter dataset of the same size with replacement (any words whose frequency is less than five were removed). We then compute differences between association scores using the process in Section 4 for each SGNS model. The pretrained models have an embedding size of 200, context window of five, and were trained using five negative samples.

Within a particular gender identity, the poor are significantly more associated with the discourse about the political, economic, and domestic sphere than the rich ($p < 0.05$). This can be seen in the first and last rows of Figure 1. Within a particular economic identity, females are significantly more associated with discourse about the domestic sphere than males. On the other hands, males are more significantly associated with discourse concerning the economic and political sphere than females. This can be seen in the second and fifth rows of Figure 1. According to our results, discourse concerning the domestic sphere has an intersectional association towards female, poor individuals. This can be seen by the fact that female, poor individuals are always significantly associated with domestic sphere discourse compared to the other intersected identities (see the first, second, and fourth rows of Figure 1). Similarly, discourse about the economic and political spheres has an intersectional association towards male, poor individuals (second, third, and sixth rows of Figure 1).

Recall from Section 3 that the majority of tweets are from Lagos and FCT. As such, a majority of the discourse in the dataset is biased towards those two states in Nigeria. This means that the associations detected for the intersected identities are not representative of individuals in all of Nigeria, but rather those that live in Lagos and FCT.

6 Discussion

Our analysis provides insight into what social institutions are of discursive interest to intersected social identities in Nigeria with a bias towards individuals from Lagos and FCT. In particular, our results show who is more vocal about a particular institution, and which individuals are less vocal about a given institution, but it does not explicitly mention whose voice is the most marginalized. This analysis is a good starting point for detecting bias in discourse about an institution, but work is needed to extract the most marginalized voices.

Similar to Nelson (2021), we find that machine learning can enhance qualitative research methods, allowing us to juxtapose quantitative outcomes with qualitative examples. For example, “I came

[1] Graph generated based on code from Nelson (2021): https://github.com/lknelson/measuring_intersectionality

across a poor women who had recently delivered five children. She needed money for food and medical bills. Such a sad example of poverty in Nigeria" is reflective of the lived experience of intersectionality while our results provide evidence for how intersected identities are linked to particular institutions at a larger scale.

The results of our analysis also aligns with several recent qualitative works that look at discrimination and bias in Nigeria. Dosekun (2022) showed that females are heavily associated with the domestic sphere (i.e., having children and domestic skills). Additionally, Enfield (2019) mentioned that females are represented in the labor markets, but they are penalized through low wages and activity. Enfield (2019) also described that females (especially poor females) join the labor market late due to the cultural pressure of early marriage and having children. This implies that males have more freedom in the labor market than females, aligning with our results that males are more associated with the economic spheres than females.

7 Conclusion

This paper presents an initial study which uses SGNS models trained on Twitter data from Nigeria to determine how intersectional identities are associated with discourse on social institutions. Our results show that female, poor individuals are more associated with discourse from Lagos and FCT concerning the domestic sphere while male, poor individuals are associated with discourse about the economic and political spheres.

There are several avenues for future work. First, the efficiency of the analysis could be improved, particularly to handle large corpora. Second, the sensitivity of associations to model hyperparameters could be assessed to ensure the associations hold under different hyperparameter choices. Finally, the analysis could be made sensitive to dataset statistics such as geographic distribution.

Limitations

The analysis done in this paper has several limitations that would benefit future investigation. The first limitation is that the analysis assumes that all individuals in the population are represented equally in a dataset. As we stated, a majority of the tweets in the Twitter dataset come from Lagos and FCT, both of which may have the benefit of technological access and literacy. Unfortunately, this skews our analysis such that the associations extracted from the word embedding models is really representative of Lagos and FCT instead of Nigeria. The second limitations concerns the efficiency and computational resources required to run this analysis. Our analysis required using a number of SGNS models trained on nearly 30 million tweets. While training is done only once, it requires training on server-sized systems over several days.

Ethical Impacts Statement

This study was conducted as basic research using publicly available Twitter data that has been collected and approved for use by an independent IRB and a HRPO agency. The intent of this study was to replicate the approach by Nelson (2021) using social media data, showing that it is possible to quantify how intersectional identities are embedded in structural social inequalities. Such bias quantifications - while highlighting social inequalities - can serve to counter or strengthen social inequalities if applied in questionable contexts (e.g., marketing/targeting, rating systems, algorithmic decision making). However, our intention with this study is to highlight and quantify social inequalities as a way to provide evidence of its existence in society.

The research team consists of women and men with diverse ethnic backgrounds, trained in Western educational institutions. A limitation of our interpretation of these results is that we did not have individuals native to Nigeria be part of the research team. We used an intersectional theoretical framework to reduce bias, and believe that using inductive methods (e.g., grounded theory, machine-learning) to this research reduces biases that may be introduced by a researcher. Still, we acknowledge that social media data is in no way representative of a diverse population as the one in Nigeria with large parts of the population not having access to technology. Finally, the impact of an intersectionality analysis helps center marginalized voices.

Acknowledgements

The research was supported by funding from the Defense Advanced Research Projects Agency (DARPA UGB HR00111890015, DARPA HABITUS W911NF-21-C-0007-04). The views, opinions and/or findings expressed are those of the authors and should not be interpreted as representing the official views or policies of the Department of Defense or the U.S. Government.

References

Maria Antoniak and David Mimno. 2021. Bad seeds: Evaluating lexical methods for bias measurement. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 1889–1904, Online. Association for Computational Linguistics.

Greta R. Bauer, Siobhan M. Churchill, Mayuri Mahendran, Chantel Walwyn, Daniel Lizotte, and Alma Angelica Villa-Rueda. 2021. Intersectionality in quantitative research: A systematic review of its emergence and applications of theory and methods. *SSM - Population Health*, 14:100798.

Liam Kofi Bright, Daniel Malinsky, and Morgan Thompson. 2016. Causally interpreting intersectionality theory. *Philosophy of Science*, 83(1):60–81.

Aylin Caliskan, Joanna J. Bryson, and Arvind Narayanan. 2017. Semantics derived automatically from language corpora contain human-like biases. *Science*, 356(6334):183–186.

Kimberle Crenshaw. 1989. Demarginalizing the intersection of race and sex: A black feminist critique of antidiscrimination doctrine, feminist theory and antiracist politics. *University of Chicago Legal Forum*, 1989.

Simidele Dosekun. 2022. The problems and intersectional politics of "#beingfemaleinnigeria". *Feminist Media Studies*, 0(0):1–17.

Sue Enfield. 2019. Gender roles and inequalities in the nigerian labour market. *Institute of Development Studies*.

Roni Friedman, Lena Dankin, Yufang Hou, Ranit Aharonov, Yoav Katz, and Noam Slonim. 2021. Overview of the 2021 key point analysis shared task. In *Proceedings of the 8th Workshop on Argument Mining*, pages 154–164, Punta Cana, Dominican Republic. Association for Computational Linguistics.

Scott Friedman, Sonja Schmer-Galunder, Anthony Chen, and Jeffrey Rye. 2019. Relating word embedding gender biases to gender gaps: A cross-cultural analysis. In *Proceedings of the First Workshop on Gender Bias in Natural Language Processing*, pages 18–24, Florence, Italy. Association for Computational Linguistics.

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.

Wei Guo and Aylin Caliskan. 2021. Detecting emergent intersectional biases: Contextualized word embeddings contain a distribution of human-like biases. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*, AIES '21, page 122–133, New York, NY, USA. Association for Computing Machinery.

Mark Anthony Hoffman, Jean-Philippe Cointet, Philipp Brandt, Newton Key, and Peter Bearman. 2018. The (protestant) bible, the (printed) sermon, and the word(s): The semantic structure of the conformist and dissenting bible, 1660–1780. *Poetics*, 68:89–103.

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.

Chandler May, Alex Wang, Shikha Bordia, Samuel R. Bowman, and Rachel Rudinger. 2019. On measuring social biases in sentence encoders. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 622–628, Minneapolis, Minnesota. Association for Computational Linguistics.

Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in Neural Information Processing Systems*, pages 3111–3119.

John W. Mohr and Petko Bogdanov. 2013. Introduction—topic models: What they are and why they matter. *Poetics*, 41(6):545–569. Topic Models and the Cultural Sciences.

Laura K Nelson. 2021. Leveraging the alignment between machine learning and intersectionality: Using word embeddings to measure intersectional experiences of the nineteenth century u.s. south. *Poetics*, 88:101539.

Adekemi Oluwayemisi Sekoni, Kate Jolly, and Nicola Kay Gale. 2022. Hidden healthcare populations: using intersectionality to theorise the experiences of lgbt+ people in nigeria, africa. *Global Public Health*, 17(1):134–149. PMID: 33287671.

Yi Chern Tan and L. Elisa Celis. 2019. Assessing social and intersectional biases in contextualized word representations. In *Advances in Neural Information Processing Systems*.

Jieyu Zhao, Tianlu Wang, Mark Yatskar, Ryan Cotterell, Vicente Ordonez, and Kai-Wei Chang. 2019. Gender bias in contextualized word embeddings. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 629–634, Minneapolis, Minnesota. Association for Computational Linguistics.