Journal of Computational Social Science manuscript No.
(will be inserted by the editor)

Understanding Sensitive Social Issues in Senegal through Language Use: A Machine Learning Approach

Kamwoo Lee[1] · Jeanine Braithwaite[2] · Michel Atchikpa[3]

Received: date / Accepted: date

Abstract Language is a reflection of issues and value systems of a society. This study tries to understand sensitive public issues in Senegal through language use. To this end, we utilize word embeddings, a numerical word representation, to analyze concepts, connotations, and nuances of several words. The state-of-art machine learning method can effectively extract the word embeddings from a collection of texts. Since people in different societies possess different mindsets and language uses, comparing semantic differences of words in different corpora is an efficient way to draw cross-cultural insights and implications. In this study, we extract word embeddings from Senegalese newspapers and French Wikipedia and then compare the results to identify different word sentiments in Senegalese cultures to understand the past, present, and future of the country.

Keywords Word Embeddings · French · Senegal · Social Analysis

1 Introduction

Interviews and surveys are often embraced as primary methods for social science studies. However, there is a limitation on collecting truthful data, especially opinions, through questionnaires when the subject matter is inherently sensitive, such as religion, race, and gender. Interviewee or survey participants might not give honest answers or, even if they are willing to, they might not realize their unconscious thoughts and behaviors. When it comes to sentiments of the general public, using language and texts can be an option since language collectively reflects values and concepts within society.

✉ Kamwoo Lee
kl9ch@virginia.edu
[1] Systems and Information Engineering, University of Virginia
[2] Frank Batten School of Leadership & Public Policy, University of Virginia
[3] Centre de Recherches Économiques Appliquées, Université Cheikh Anta Diop de Dakar

In this paper, we present our analysis of sensitive social issues that have been captured in Senegalese texts. We utilize the state-of-art machine learning techniques to extract word representations called word embeddings, which map vocabulary to a multi-dimensional semantic vector space. We create two sets of word embeddings from Senegalese newspapers (SnText) and Wikipedia pages in French (FrText), separately. We treat FrText as a reference point as it has more diverse context, larger size, and better text quality than any other French text corpus. We then compare those semantic representations word-by-word between the two sources to expose Senegalese specific viewpoints and perceptions.

Broadly stated, we create conceptual "semantic scales" with word vectors and "measure" the value of each word. Any pair of antonyms can act as a semantic scale. For example, ⟨*haut-bas* (high-low)⟩ scale measures a symbolic height of a word concept. ⟨*bien- mal* (good-evil)⟩ scale measures moral standards associated with a word. On the ⟨*bien-mal* (good-evil)⟩ scale, the word "*Allah*" is inclined to ⟨*bien* (good)⟩ in SnText whereas it is more close to ⟨*mal* (evil)⟩ in FrText. This difference is easily understandable considering that the vast majority of the Senegalese population is Muslim, whereas European French speaking countries have historically been at odds with Islam as shown through FrText. Similarly, a lot of words such as *pays* (country), *culture* (culture), *résident* (resident), and *norme* (norm) have contrasting conceptual values between SnText and FrText on the ⟨*blanc-noir* (white-black)⟩ scale, which obviously reflects the difference in race. We seek to take this approach an additional step further and expose some sensitive issues in Senegal that are hard to quantify otherwise.

2 Background and Related Work

This paper proposes a new text-based approach to study social issues of Senegal. This section summarizes selected findings that comprise the methodological foundation of our study. It also provides an overview and brief history of Senegal.

2.1 Natural Language Processing and Word Embeddings

Natural Language Processing (NLP) is a subfield of information engineering and artificial intelligence, broadly defined as the automatic manipulation of speech and text. NLP applications explore how computers can be used to understand and manipulate natural language for high-level tasks, including semantic analysis, document classification, information retrieval, and machine translation. Transforming text into something a computer algorithm can digest, or in other words, representing text numerically, typically involves feature extraction.

Words can be seen as discrete, categorical data since each word represents a distinct meaning. Feature extraction is a process of mapping from such

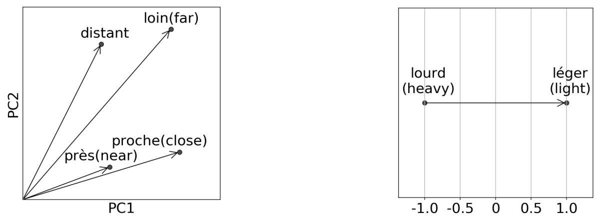


Fig. 2: Several words vectors in FrText displayed in 2-dimension using PCA. Close words have similar angles (left) and a pair of antonyms defines a semantic scale (right)

categorical data to real-valued numbers. Conceptually, it projects a word from a vocabulary size dimension to a lower dimensional space. One of the most recent word feature learning techniques is word embedding, which represents each word as a real-valued vector based on the textual context in which the word is found [19]. It is a powerful tool to extract both semantic and syntactic meanings from a large unlabeled corpus, with the idea being that more similar words are closer to one another in the embedding space.

Since word embedding is a numerical vector, it is possible to measure the similarity between different words. The most widely used measure is the cosine similarity defined by $\cos(w_x, w_y) = \frac{w_x \cdot w_y}{||w_x|| \cdot ||w_y||}$, which ranges from -1 (most distant) to 1 (closest). The reasoning behind the cosine similarity is that similar words have similar semantic vectors, and the angle between the vectors is small. Additionally, by subtracting two vectors of antonyms gives a new set of direction and magnitude, which defines a conceptual semantic scale between the two antonyms (Fig. 2). The cosine similarity between a semantic scale and a word gives a conceptual value of the word on the scale.

2.2 Biases and Stereotypes Captured by Word Embeddings

Bolukbasi et al. [2] suggested that the word embeddings trained on Google News articles exhibit female/male gender stereotypes to a disturbing extent. The word embeddings pinpoint sexism implicit in the text. For instance, given an analogy puzzle, "man is to computer programmer as woman is to x", the embedding vectors find that x=homemaker. Some of the most extreme occupations as projected on the female-male gender direction are homemaker, nurse, receptionist, librarian for woman and maestro, philosopher, captain, architect for man. Caliskan et al. [4] statistically showed that training word embeddings from text data absorbs more diverse semantic biases in the text. They replicated a spectrum of known biases, such as flowers (pleasant) vs. insects (unpleasant) and European-American names (pleasant) vs. African-American names (unpleasant), using word embeddings, demonstrating that this machine learning technique can be a tool for studying historical biases and prejudicial attitudes.

Considering the fact that the word embeddings blatantly reflect the biases and stereotypes of language users, it can be generalized to capture collective minds, and even the intuition, of the language users. Also, since people in different societies have different mindsets, we believe that the word embeddings can be a tool for better understanding the culture, value systems, and contemporary issues of a society through language. Thus, when trained on Senegalese text, the embeddings can reveal varying opinions or viewpoints on social issues expressed throughout the country.

2.3 Senegal

The Republic of Senegal is situated on the westernmost point of the African continent. The official language of Senegal is French, which was inherited from the colonial era dated from the mid-15th century to 1960. Currently, almost all official documents and newspaper articles are in French, but most Senegalese people also use ethnic spoken languages, such as Wolof, native to their particular area or people. Senegal is recognized by its historical past in colonization since it was one of the major sources of slave deportation to other continents [15]. As with many developing countries, Senegal has its own potentials and problems. We believe that NLP and word embedding techniques can contribute to identifying, and even prioritizing, the country's assets and liabilities by taking advantage of unstructured text data.

3 Training Word Embeddings

Training word embeddings involves a lot of different moving parts, from collecting texts from multiple websites to using various NLP techniques to extract relevant information. This section provides explanations for the sequential steps to show how each part contributes to our analysis of public sentiments, expressed in SnText and FrText. Since both are in French, some steps are different from English NLP.

3.1 Text Processing Steps

1) Data Collection
We collected past and current articles that are accessible through an exhaustive web search during the period of 4/1/2019 - 4/20/2019 from 15 newspaper websites*in Senegal. Although there are spoken languages like Wolof in Senegal, almost all newspapers and websites are in French. For FrText, we collected a database dump of all Wikipedia articles in French through CirrusSearch. The dump is a 4/15/2019 snapshot. We extracted plain text from the dump with Wikiextractor [1].

2) Tokenization and Lemmatization

* actusen.sn, www.dakaractu.com, www.dakarmatin.com, www.dakarmidi.net, www.ferloo.com, homeviewsenegal.com, www.impact.sn, www.koldanews.com, www.lequotidien.sn, www.leral.net, letemoin.sn, senego.com, www.sudonline.sn, teranganews.sn

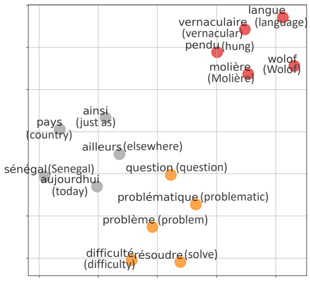
Tokenization is the process of breaking down a text into minimal meaningful units, called tokens. This process also splits contractions (e.g., "*l'hôtel*" → "*l*" and "*hôtel*"). Lemmatization performs vocabulary and morphological analysis of the word and converts each word to its original form (e.g., *fais, fait, faisons, faites, font* → *faire*). In many cases, especially in English, derived word forms are preserved in word embeddings for syntactic analysis such as tense and plurality. However, since our focus is semantic analysis and French has numerous conjugations for subject, gender, and tense than English, we decided to perform lemmatization to decrease the size of vocabulary and better utilize the limited corpus to learn the semantics of each word.

3) Detecting Phrased Words and Text Cleaning
In English NLP, a single word is usually treated as a token. However, since a lot of words are phrased in French (e.g., *salle de bain* for bathroom), it is necessary to extract phrased words and treat them as a single token for meaningful analysis. We extracted bigram phrases from FrText using the Gensim phrase detection algorithm [16] (min_count = 200, common_terms = all articles and prepositions) and then converted the phrases in FrText and SnText into tokens by concatenating with underscores (e.g., *salle de bain* → *salle_de_bain*). We further removed non-Roman characters such as punctuation and special characters to increase system performance and changed every word to lower case to have consistent capitalization. After the cleaning, there are around 100M tokens and 360K unique words in the SnText corpus, and there are 1B tokens and 7M words in the FrText corpus.

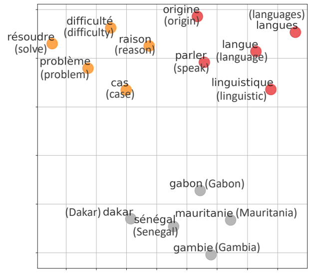
4) Training Word Embeddings
We chose the Global Vectors for Word Representation (GloVe) algorithm [14] to train our word embeddings. This algorithm first constructs a co-occurrence matrix that has the information on how frequently a word appears in high-dimensional contexts in a large corpus. Then, the algorithm reduces the dimension while retaining most of the variance in the high-dimension. The lower-dimensional matrix then becomes a set of word embeddings. We trained the word embeddings with 200 dimensions both on SnText and FrText.

3.2 Validation of the Word Embeddings

Since training word embeddings is an unsupervised machine-learning process, there are no right answers to test the result. However, it is still necessary to validate the word embeddings learned from our text collections before using them as the basis of our analysis. We first visually examined word groupings to see if similar and related words are located nearby in the semantic vector space, and spatially close words are similar and related. To visualize positions of words with 200 dimensions, we utilized T-distributed Stochastic Neighbor Embedding (t-SNE), which is a popular machine learning algorithm for visualizing high-dimensional data in two-dimensional space while keeping relative pairwise distance between points [13]. In Fig. 4, words that were far from each



(a) word groups in SnText



(b) word groups in FrText

Fig. 4: Closest 5 words to *sénégal* (Senegal), *problème* (problem), and *langue* (language) in SnText and FrText. The different words in the same groups, especially in the *sénégal* group, reflect cultural, geographical contexts of the word.

other in 200-D space are located far away in 2-D space, and initially close words are also converted to close ones in the lower dimension. Both word embeddings from SnText and FrText show consistent and coherent word groupings.

For more quantitative validation, we evaluated our word vectors on a word similarity task, based on French translation of the MEN dataset, which was developed specifically for computational semantics testing [3]. This dataset contains word pairs with human similarity ratings. We removed word pairs that have the same French translation (e.g., *voiture:voiture* for automobile:car) and some words that have significantly different two or more meanings in French (e.g., *essence* for gasoline, essence, and species). We measured the cosine similarities of the embeddings for each pair in the dataset and then calculated the Spearman's rank correlation coefficient (Rho) with the human judgment. The Spearman's Rho is 0.48 for SnText and 0.52 for FrText. We also used the Spearman's Rho to choose parameters of our word embedding model.

4 Social Issue Analysis with Word Embeddings

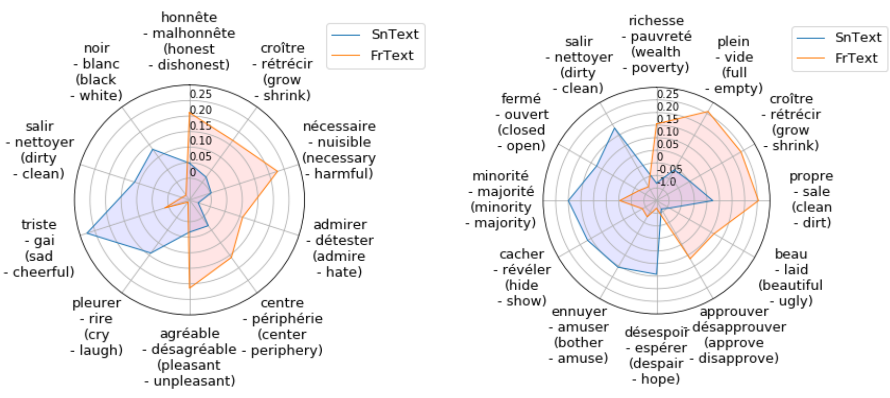
The essence of word embeddings for our study is being able to measure conceptual values of a word on a chosen semantic scale in the vector space. In order to examine the possibilities of word embeddings as a social issue analysis tool, we try three approaches: 1) comparing the conceptual values of words on the semantic scales, 2) examining orders of conceptual values for a list of words, and 3) inspecting orientations of the semantic scales. We present the three approaches with three different social issues in the following subsections. For notational simplicity, we denote a semantic scale as $\langle$A - B$\rangle$ where A has (+) value and B has (-) value on the scale.

4.1 Legacy of French Assimilation

During the colonial period, the French had a policy of assimilation [12]. Influenced by their colonial ideology, the French claimed to "civilize" indigenes and taught their subjects that, by adopting French language and culture, they could eventually become black Frenchmen. The 'Four Communes' (Goree, Saint Louis, Dakar, and Rufisque) in Senegal exemplify this policy [7]. The French colonists give citizenships to inhabitants of those towns when they were "civilized" enough. As a result, deliberately by the French or inadvertently by the public, the Senegalese society developed unfavorable views for some of its indigenous tradition [17]. We see an aspect of this legacy through the words "*tradition* (tradition)" and "*indigène* (indigenous)" that have significantly different conceptual values in Senegalese texts (Fig. 6).

Senegalese texts collectively relate the words "*tradition* (tradition)" and "*indigène* (indigenous)" to negative words such as "*triste* (sad)", "*salir* (get dirty)", "*cacher* (hide)", and "*déspoir* (despair)". Similar words like *traditionnel* (traditional), "*culture* (culure)", and "*coutume* (custom)" have a similar tendency. Although we cannot decisively conclude that these differences originated only from colonialism, it is hard to make these disturbing connections without implicating the colonial legacy. French colonialism still has an impact on the Senegalese society in many arenas, notably in the education system. Our study suggests, for example, that language education most directly affects the perspectives of the people, including the writers of Senegalese newspapers, as captured in the SnText dataset.

The French colonists vigorously and systematically oppressed indigenous African languages, cultural values and local religious beliefs [17]. Through legislation and its administrators, the French authorities tried to maintain a



(a) *tradition* (tradition) (b) *indigène* (indigenous)

Fig. 6: Measurement of words on semantic scales that give most contrasting values between SnText and FrText. Each spoke represents a semantic scale with positive value outside and negative value inside.

highly centralized system whereby they controlled the language of instruction, the curriculum, and the aims of the education system. Therefore, throughout colonial rule, the French language was imposed, and the indigenous languages were marginalized and stigmatized [6]. After its independence, Senegal chose French as the official language of the country, keeping its high status as the language of administration, media, education, international communication, etc. Despite numerous efforts undertaken, such as recognizing indigenous languages as the national languages, the national languages remain markedly absent from the official domains. French continues to dominate virtually the entire education system, from kindergarten to university. Currently, there are no government or private schools that systematically teach literature, mathematics, physics, finance or computer sciences, or any other school subjects in the national languages [6]. Considering that language is a vehicle of tradition, custom, and culture, it is understandable that Senegalese texts have negative connotations of those words in the French language as used in Senegal.

4.2 Environmental Problems

Another possibility in utilizing the word embeddings is to look at vocabulary for well-known social issues and estimate their perceived relative order. There are many current, pressing issues in Senegal. According to the World Bank's Systemic Country Diagnostic (SCD), the poverty rate is over 40%, indicators continue to lag behind for maternal health, nutrition and education, and the labor market features high inactivity and underemployment rates [20]. We measured the following 14 words on ⟨*difficile-facile* (difficult-easy)⟩ scale to try to compare them:

> *pauvreté* (poverty), *chômage* (unemployment), *crime* (crime), *corruption* (corruption), *maladie* (disease), *illettrisme* (illiteracy), *discrimination* (discrimination), *guerre* (war), *famine* (famine), *sécheresse* (drought), *sanitaire* (sanitation), *malnutrition* (malnutrition), *inégalité* (inequality), *inondation* (flood)

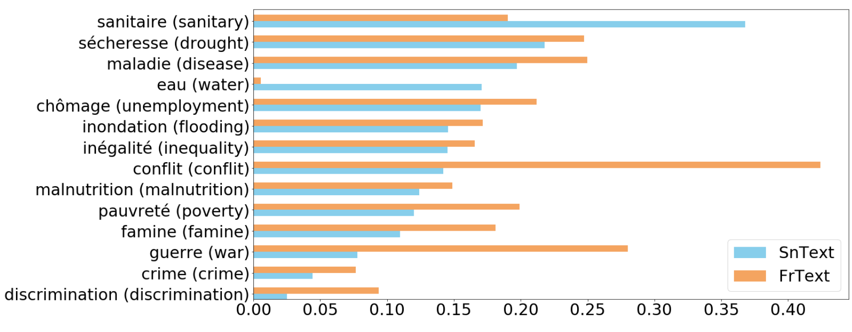


Fig. 7: Measurement of words for typical public problems on ⟨*difficile-facile* (difficult-easy)⟩ scale. The word *eau* (water) is included for further explanation.

The word *sanitaire* (sanitation) has a significantly high value when compared with other problems in SnText. The value is also twice as high as the same word in FrText (Fig. 7). When looking at the order of the words, we notice that the top difficult problems in SnText are closely related to water. Even the word *eau* (water) is much closer to ⟨*difficile* (difficult)⟩ side while it is almost neutral in FrText.

Senegal, like all other Sahelian countries, has experienced serious droughts in the past. Today, with the impact of climate change, droughts have more serious consequences for rural and urban populations [18]. Water scarcity first impacts the agriculture sector, and the economic impacts propagate to all sectors of the agriculture-based country, which then exacerbates other major social problems such as unemployment, poverty, and malnutrition [9]. The complication arises in that water is not a solution for the problems. In big cities including Dakar, Senegal's most populated city, sanitation infrastructures, such as sewers, are in bad condition. The sanitary sewers in cities are not deep enough so that they easily overflow when rainfall exceeds their low capacity resulting in larger after-effects such as strong smells and disease spread [5]. Although social problems cannot be measured or prioritized by a single criterion, it is useful to know which issues receive most public attention. In this sense, the word embedding analysis can give a collective summary of a social issue.

4.3 Past - Present - Future in Senegal

Since the semantic scales derived from the word embeddings exhibit the perspectives of people who use the language, we could further examine the perspectives by comparing different semantic scales. Specifically, it is possible to look at how orientations of different values scales are aligned. We can check the orientations of two scales by calculating the cosine distance. Normally, many semantic scales are independent. For instance, the ⟨*femme-homme* (woman-man)⟩ scale usually does not bear a strong relation to the ⟨*difficile-facile* (difficult-easy)⟩ scale. However, in SnText, we found interesting alignments between ⟨*passé-présent* (past-present)⟩ scale and others, such as ⟨*abnormal-normal* (abnormal-normal)⟩, ⟨*amateur-profesionnel* (amateur-professional)⟩, and ⟨*haine-amour* (hatred-love)⟩. Similar patterns can be found with ⟨*passé-futur* (past-future)⟩ scale (Fig. 9). When we ran the same analysis in FrText, the cosine similarities of all of these scales are less than 0.05, suggesting that they are independent in general.

This result could originate from Senegalese history, current optimism, or prospects for the future. Senegal's economic development trajectory supports this expectation. Senegal has one of the fastest-growing economies in Sub-Saharan Africa. Driven by diversified exports and favorable external conditions, such as the decline of global oil, commodities and food prices, the Senegalese economy has been expanding at more than 6 percent annually since 2014 [20]. The forecast for the future is hopeful, particularly with oil and gas

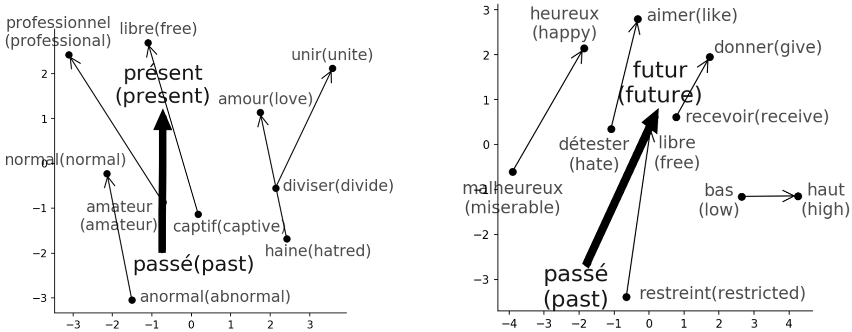


Fig. 9: Semantic scales that are aligned with ⟨*passé-présent* (past-present)⟩ scale (left) and ⟨*passé-futur* (past-future)⟩ scale (right) in SnText, displayed in reduced 2 dimension using PCA.

production expected in 2022. Growth accelerated to over 7% in 2017, and is expected to remain over 6% in 2018 and in the following years [21].

There have been other factors that have contributed to Senegal's optimism. Senegal has been one of the most stable and politically open countries in Africa, with three major peaceful political transitions since independence in 1960 [20]. It has been considered an "exception" in West Africa where many countries have been plagued by military coups, civil wars and ethnic/religious conflicts [8]. An additional illustration is that while Sufi Islam is the dominant religious practice in Senegal, religious minorities including Christianity have long been accepted and given the freedom to practice their religion. Furthermore, Senegal has a long history of fostering global intellectual, diplomatic, and financial connections [8].

It is hard to measure such vague concepts as optimism and hope in a society since a few socio-economic indicators cannot capture the future-oriented attitudes of the population. Quantifying them, however, can help guide public policy and embrace a more holistic approach. Specifically, optimism measurement with word embeddings can supplement existing surveys such as the World Happiness Index by the United Nations.

5 Discussion

This study emphasizes the value of word embeddings in studying social issues in Senegal. Several limitations remain, and many extensions are possible. The case studies that we presented in section 4 are by no means an exhaustive analysis on the issues. Each individual topic deserves a more dedicated, in-depth study. On the methodological side, we tried to find a recurrent tendency of difference between SnText and FrText under a chosen topic rather than showing statistical significance of difference for each word, which can lead to multiple research topics as shown by Caliskan et al. [4].

With a proper experimental design, such as language use of different groups in different periods in time, it could be possible to find causal relationships of why a word takes a specific connotation. Tracing semantic changes of words historically can be an important aspect of social issue research as presented by Hamilton et al. [11] and Yao et al. [22]. We only compared Senegalese texts and French texts. A possible next step is to expand this approach and compare multiple countries. One promising future project is collecting texts from all former French colonies and examining their similarities and differences. One can also inspect conceptual values of words in multiple, different languages by utilizing pre-trained word embeddings, such as Word2Vec for 157 languages introduced by Grave et al. [10].

6 Conclusion

This paper demonstrated that training word embeddings from Senegalese texts and analyzing representative words can shed light on social issues in the country. While French is the official language in Senegal, we found different connotations between FrText and SnText on the same words, and we believe these different interpretations reflect the mindsets of Senegalese society. Senegalese texts collectively (1) carry unfavorable views of the country's past, including tradition, (2) voice water-related problems as the most challenging of contemporary issues, and yet (3) also expresses public optimism about the future. We believe that quantifying multi-faceted semantics of words enables researchers and policy makers to utilize textual data, especially when there are not enough structured data, such as survey and census data, to study opinions on public issues.

Acknowledgements This research was supported by the Global Infectious Disease Institute at the University of Virginia (UVA). We also thank Senegal Research Group members at UVA for their feedback to this research: Grace Wood, Gabrielle Posner, and Jordan Beeker. We also would like to show our deepest gratitude to Fama Gueye, an exchange scholar from Université Cheikh Anta Diop de Dakar for her insights in social issues of Senegal.

On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

1. Attardi, G.: Wikiextractor: A tool for extracting plain text from wikipedia dumps (2009)
2. Bolukbasi, T., Chang, K.W., Zou, J.Y., Saligrama, V., Kalai, A.T.: Man is to computer programmer as woman is to homemaker? debiasing word embeddings. In: Advances in neural information processing systems, pp. 4349–4357 (2016)
3. Bruni, E., Tran, N.K., Baroni, M.: Multimodal distributional semantics. Journal of Artificial Intelligence Research **49**, 1–47 (2014)
4. Caliskan, A., Bryson, J.J., Narayanan, A.: Semantics derived automatically from language corpora contain human-like biases. Science **356**(6334), 183–186 (2017)
5. DE LA VILLE, P.D., DAKAR, D.: Republique du senegal (2012)

6. Diallo, I.: 'to understand lessons, think through your own languages.'an analysis of narratives in support of the introduction of indigenous languages in the education system in senegal. Language Matters **42**(2), 207–230 (2011)
7. Diouf, M.: The french colonial policy of assimilation and the civility of the originaires of the four communes (senegal): A nineteenth century globalization project. Development and Change **29**(4), 671–696 (1998)
8. Diouf, M.: Tolerance, democracy, and Sufis in Senegal. Columbia University Press (2013)
9. Garcia, L.: Impact du changement climatique sur les rendements du mil et de l'arachide au sénégal: Approche par expérimentation virtuelle. Ph.D. thesis, Montpellier SupAgro (2015)
10. Grave, E., Bojanowski, P., Gupta, P., Joulin, A., Mikolov, T.: Learning word vectors for 157 languages. arXiv preprint arXiv:1802.06893 (2018)
11. Hamilton, W.L., Leskovec, J., Jurafsky, D.: Diachronic word embeddings reveal statistical laws of semantic change. arXiv preprint arXiv:1605.09096 (2016)
12. Lewis, M.D.: One hundred million frenchmen: The "assimilation" theory in french colonial policy. Comparative Studies in Society and History **4**(2), 129–153 (1962)
13. Maaten, L.v.d., Hinton, G.: Visualizing data using t-sne. Journal of machine learning research **9**(Nov), 2579–2605 (2008)
14. Pennington, J., Socher, R., Manning, C.: Glove: Global vectors for word representation. In: Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP), pp. 1532–1543 (2014)
15. Rawley, J.A., Behrendt, S.D.: The transatlantic slave trade: a history. U of Nebraska Press (2005)
16. Řehůřek, R., Sojka, P.: Software Framework for Topic Modelling with Large Corpora. In: Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks, pp. 45–50. ELRA, Valletta, Malta (2010). http://is.muni.cz/publication/884893/en
17. Salhi, K.: Rethinking francophone culture: Africa and the caribbean between history and theory. Research in african literatures **35**(1), 9–29 (2004)
18. Sène, A., Sarr, M.A., Kane, A., Diallo, M.: L'assèchement des lacs littoraux de la grande côte du sénégal: mythe ou réalité? cas des lacs thiourour warouwaye et wouye de la banlieue de dakar. Journal of Animal &Plant Sciences **35**(2), 5623–5638 (2018)
19. Turian, J., Ratinov, L., Bengio, Y.: Word representations: a simple and general method for semi-supervised learning. In: Proceedings of the 48th annual meeting of the association for computational linguistics, pp. 384–394. Association for Computational Linguistics (2010)
20. World Bank Group: Systematic country diagnostic of senegal. World Bank, Dakar. http://hdl.handle.net/10986/30852 (2018)
21. World Bank Group: Senegal. https://www.worldbank.org/en/country/senegal/overview (2019). Accessed on 6/10/2019
22. Yao, Z., Sun, Y., Ding, W., Rao, N., Xiong, H.: Dynamic word embeddings for evolving semantic discovery. In: Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining, pp. 673–681. ACM (2018)