HausaHate: An Expert Annotated Corpus for Hausa Hate Speech Detection

Francielle Vargas[1,2], **Samuel Guimarães**[2], **Shamsuddeen Hassan Muhammad**[3]
Diego Alves[4], **Ibrahim Said Ahmad**[5], **Idris Abdulmumin**[6], **Diallo Mohamed**[7]
Thiago A. S. Pardo[1], **Fabrício Benevenuto**[2]
[1]University of São Paulo, Brazil [2]Federal University of Minas Gerais, Brazil
[3]Imperial College London, UK [4]Saarland University, Germany [5]Northeastern University, US
[6]University of Pretoria, South Africa [7]University of Saint Thomas Aquinas, Burkina Faso
francielleavargas@usp.br

Abstract

We introduce the first expert annotated corpus of Facebook comments for Hausa hate speech detection. The corpus titled HausaHate[1] comprises 2,000 comments extracted from Western African Facebook pages and manually annotated by three Hausa native speakers, who are also NLP experts. Our corpus was annotated using two different layers. We first labeled each comment according to a binary classification: offensive versus non-offensive. Then, offensive comments were also labeled according to hate speech targets: race, gender and none. Lastly, a baseline model using fine-tuned LLM for Hausa hate speech detection is presented, highlighting the challenges of hate speech detection tasks for indigenous languages in Africa, as well as future advances.

1 Introduction

In African countries, the hate speech phenomenon is especially serious due to a historical problem regarding ethnic conflicts. Specifically, the Western region still lacks more research on hate speech focusing on its indigenous languages. Moreover, as most of the existing hate speech data resources are developed for the English language, the research and development of hate speech technologies for African indigenous languages are less developed.

Hate Speech (HS) is defined as any expression that attacks a person or a group based on identity factors, such as ethnicity, religion, origin, gender identity, sexual orientation, or disability (Zampieri et al., 2019; Fortuna and Nunes, 2018). Furthermore, hate speech is a particular form of offensive language that considers stereotypes to express an ideology of hate (Warner and Hirschberg, 2012), which may be used by terrorist groups to justify their acts by attacking targets, or even serve to propagate its ideology, acting as propaganda. In this regard, in Nigeria, which was divided into ethnic lines during independence, online hate speech and hate crimes have been a recurring issue.

Most existing conflicts in Nigeria are due to differences between Hausa and Fulani ethnic groups concentrated in the north, and between Yoruba and Igbo in the southwest, in which there are continuing ethnic tensions. In recent years, there was an increase in the hate rhetoric against the Fulani group (Nwozor et al., 2021), which lives as herdsmen, migrating across the region, and the ethnic-religious differences between the Igbo and the Fulani, the first being majority Christians and the second Muslims, which fuel hateful rhetoric in the country. Table 1 shows examples of offensive comments and hate speech targets in Hausa.

According to Ezeibe (2021) and Ridwanullah et al. (2024), the culture of hate speech is an often neglected major driver of election violence in Nigeria. Nevertheless, although the implementation of existing anti-hate speech laws presents an opportunity for protecting the rights of minorities and preventing election violence, its regulation is still not effective due to the difficulty of identifying, quantifying and classifying online hateful content.

Here, we introduce a benchmark corpus for Hausa hate speech detection. The corpus titled HausaHate comprises 2,000 comments extracted from the Western African Facebook pages and manually annotated by three Hausa native speakers, who are also NLP experts. Our corpus was annotated according to two layers: (i) a binary classification (offensive versus non-offensive), and (ii) hate speech targets (race, gender and none). We also describe our methodology to build data resources for indigenous languages in Africa that comprises data collection, data annotation, and annotation evaluation. Finally, a baseline model using fine-tuned LLM for Hausa hate speech detection is presented, highlighting the challenges of hate speech detection tasks for African indigenous languages.

[1] **HausaHate corpus**: https://github.com/franciellevargas/HausaHate

Proceedings of the The 8th Workshop on Online Abuse and Harms (WOAH), pages 52–58
June 20, 2024 ©2024 Association for Computational Linguistics

Comment	Offensive	HS Target
Ai abun Nace allah ne shike rayawa shike kashewa **Translation:** God is the one who gives life and takes it away.	No	No
To angaya muku mu wawaye kamar iyan kauye **Translation:** Who told you we are stupid like your parents.	Yes	None
95% Fulani makiya suna da hanu a Taadacin Arewa kasa Nigeria. **Translation:** 95% of Fulani herdsmen are involved in the crisis in Northern Nigeria.	Yes	Race
Ai Mata masu gemu nan akwai Dan Karin Gulma Masifa **Translation:** All women with beards, are hypocrite.	Yes	Gender

Table 1: Examples of Hausa comments annotated with offensive, non-offensive and hate speech targets.

2 Related Work

While most hate speech technologies are developed for English, African indigenous languages lack data resources to counter this problem. Towards addressing online hate speech in African countries, Ababu and Woldeyohannis (2022) proposed a corpus and baselines for Afaan Oromo hate speech detection. They obtained an accuracy of 0.84 using word2vec and BI-LSTM. Oriola and Kotzé (2019, 2020) proposed and evaluated different Machine Learning (ML) classifiers for hate speech detection in South African tweets. Reddy (2002) proposed a study on hate speech against LGBT people in Africa. They analyzed linguistic choices in a particular context of use to explain their links with gender, language, and power. Oriola and Kotzé (2022) explored word embeddings and mBERT-case to classify hate speech in South African social media texts. Taking into consideration the West African indigenous languages, there is a lack of papers that address hate speech detection (Ridwanullah et al., 2024; Abdulhameed, 2021; Auwal, 2018; Aliyu et al., 2022). Previous efforts analysed hateful content from Facebook pages data (Auwal, 2018), Twitter/X profiles (Abdulhameed, 2021) and Twitter/X interactions during an election campaign (Ridwanullah et al., 2024). In addition, an annotated hate speech corpus focused on Fulani herdsmen in Nigeria was released (Aliyu et al., 2022), which comprises three languages: English (97.2%), Hausa (1.8%) and Nigerian-Pigdin (1%). Another relevant resource called *PeaceTech HS Lexicon*[2], was proposed by the PeaceTech Lab[3] to address HS in Nigeria. It consists of a hateful lexicon for English, Fulani, Hausa, Igbo, Pidgin, and Yoruba.

[2] https://www.peacetechlab.org/nigeria-hate-speech-lexicon

[3] https://www.peacetechlab.org/history

3 Hausa Language

Hausa is a West Chadic branch of the Afro-Asiatic language family and a sub-Saharan African language with an estimated 30 million or more speakers (Chamo, 2011). Most Hausa speakers live in northern Nigeria and in southern areas of the neighboring Republic of Nigeria, where Hausa represents the majority language (Jaggar, 2001). Nigeria prior to British colonization existed as a sprawling territory of diverse ethnic groups with linguistic and cultural patterns expressed in traditional political, educational and religious systems (Dike, 1956), and there is an influence of the Hausa language in different ethnic groups in this region (Lambu, 2019). For instance, the Hausa ethnic group uses Hausa as the main language of communication. In addition, the Fulani ethnic group uses Hausa as their first language due to the historical relationship between the two groups (Hausa and Fulani) in terms of religion, inter-marriages, and social activities, which lead to the loss of their first language.

In northern Nigeria, the minority languages tend to lose their functional values due to the growing preference for Hausa. In contrast, in southern Nigeria, considering that the English language is the official communication medium, according to Chamo (2011), there has been a replacement of the mother tongues. Furthermore, the Hausa is a language of everyday communication for different domains in northern Nigeria. It is also a vehicle of specific domains in the whole country. Several business activities are dominated by the Hausa ethnic group, such as exchange of money, sales of domestic animals, trailer transportation, sales of second hand cars, etc. Hausa language is also regarded as the language of Muslim community in Nigeria. This identification is a sign of membership of the Hausa community (Chamo, 2011).

Furthermore, the permanent contact with different languages in communication of day-to-day life (e.g. it is contact between Hausa and English) lead in introducing of new words into the language. New vocabularies are generated by the group through discussion of political issues, presentation of new products or by commenting on films. The borrowings are usually inherited from English, although there are also words borrowed from Arabic and from other African indigenous languages. The reason for the use of these words is the lack of their equivalents in Hausa, when they are easily understood as terms of the source language. In general, this borrowings are considered a type of *Hausanized*, which it means new words are accepted in wide variety of communication spheres. This is reflected on the dictionaries (Chamo, 2011).

Finally, according to Ogunmodimu (2015), there is a constant concern related to language policy in order to recommend the adoption of indigenous languages (e.g. Hausa, Yoruba, Igbo, etc.) in African countries as national *lingua franca* towards obtaining emancipation from colonial legacy. In Nigeria, this would mean the promotion of Hausa over English, hence highlighting the importance of developing specific NLP data resources, methods and tools for the Hausa language.

4 HausaHate Corpus

4.1 Data Overview

We introduce a new expert annotated corpus for Hausa hate speech detection, and its statistics are shown in Table 2. Our corpus comprises 2,000 comments annotated according to two different layers: binary classification (678 offensive comments and 1,322 non-offensive comments), and hate speech targets: race (391 comments), none (222 comments), and gender (65 comments). In terms of percentage, 67.5% of comments are non-offensive and 32.5% are offensive. Regarding the hate speech targets, 57.66% are against race, 9.58% against gender, and 32.74% are non-target. In average, each comment comprises 1.31 sentences and 18.33 tokens. Specifically, hate speech targets against race and gender present 1.40 and 1.38 sentences, and 24.77 and 22.43 tokens, respectively. On a smaller scale, non-target hate speech and non-offensive comments present in average 1.17 and 1.31 sentences, and 14.22 and 16.92 tokens, respectively. In total, our corpus comprises 36,670 tokens, 2,637 sentences and 2,000 documents.

4.2 Data Collection

4.2.1 Automated Data Collection

We used the Meta CrowdTangle platform[4] to find relevant Facebook pages and posts. On this platform, it is possible to search for Facebook pages, public groups, or posts by keywords. Our main focus was on the Hausa language and Fulani group[5]. Hence, we asked to Hausa native speakers, who live in that region, potential keywords to identify hateful content in Hausa. Accordingly, we first searched keywords related to the Fulani group and also added a set of keywords directly related to terrorism (e.g. "terrorist", "terrorism", "the unidentified armed man", "fulani", "fula", "fulanin"). The search returned 1.968 posts from 11 pages and 8 groups written in Hausa, Yoruba, and Igbo. Thus, as expected, most comments comprised events and themes related to violence mainly triggered by the *racial* and *religious* beliefs. The collected comments were posted between 2021 and 2022, with 57.14% of the Facebook posts classified as photos, 28.57% as videos, and 14.29% as textual content. Lastly, we also used the Facebook Graph API [6] to collect public comments published as response. In total, we found 1,364 comments in Hausa from which 132 were responses to previous comments.

4.2.2 Manual Data Collection

During the data collection process, the platform restricted our API for keeping the automatic collection. As a result, we also manually collected 636 comments. The manual data collection relied on extraction of comments from Facebook pages identified by the automated data collection process. The majority of comments manually collected were extracted from the Facebook page called *Labarun Hausa*[7]. We randomly selected posts published in this page during 2021 and 2022 and then manually extracted their comments.

4.2.3 Data Anonymization

In order to anonymize our corpus, we first removed any user or account reference from the data automatically collected. Subsequently, during the manual data collection, we selected only the text content of comment, therefore, without any user or account reference.

[4] https://www.crowdtangle.com
[5] https://tinyurl.com/542x6svh
[6] https://developers.facebook.com/docs/graph-api/
[7] Hausa News: https://www.facebook.com/lbrhausa

Description	Offensive			Non-Offensive	All
	race	gender	non-target		
#Documents (comments)	391	65	222	1,322	2,000
#Sentences	548	90	261	1,738	2,637
#Tokens	9,686	1,458	3,157	22,369	36,670
#Avg Sentences/Document	1.40	1.38	1.17	1.31	1.31
#Avg Tokens/Document	24.77	22.43	14.22	16.92	18.33

Table 2: HausaHate corpus statistics.

4.3 Data Annotation

4.3.1 Selection of Annotators

The first step of the annotation process comprises the selection of annotators. Given the complexity and subjectivity related to the annotation of hate speech and offensive language, only experts should be selected (Vargas et al., 2022, 2021). Accordingly, we selected three Hausa native speakers annotators, who are NLP experts with high education level (at least a Ph.D. degree) from Nigeria.

4.3.2 Annotation Schema

We adopted an annotation schema proposed in Vargas et al. (2022), which provides a distinguish definition for offensive language and hate speech described as follows.

For **offensive language classification**, the annotators classified as offensive, the comments with any term or expression used with *pejorative connotation*, otherwise, it was classified as non-offensive. Examples of offensive and non-offensive comments are shown in Table 1.

For **hate speech classification**, offensive comments were annotated according to hate speech targets: race, gender and none. We used the definition of racial categories (ethnicity, religion, and color) proposed by Silva et al. (2016). Moreover, we assumed that comments with gender discrimination comprises hostility against self-identified people as female gender, treated them as objects of sexual satisfaction, reproducers, labor force, or new breeders (Garrau, 2020). Examples of hate speech targets are shown in Table 1.

It should be pointed out that our annotators also had access to the context of the comments (i.e., link to the original post with information related to neighboring comments, post topic, and domain). Finally, we selected the final label for HausaHate corpus taking into consideration the majority of votes among the three annotators.

4.3.3 Annotation Evaluation

We used the Cohen's kappa inter-annotator agreement to evaluate our corpus and the results are shown in Table 3. Observe that our annotation process presents substantial results achieving an inter-annotator agreement of 79% for offensive language annotation (offensive and non-offensive), and 60% for hate speech targets annotation (race, gender and none).

Peer Agreement	AB	BC	CA	AVG
Offensive language	0.81	0.82	0.75	**0.79**
Hate speech targets	0.60	0.61	0.59	**0.60**

Table 3: Cohen's kappa.

5 Baseline Experiments

5.1 Model Architecture and Settings

We split the data into 80% train (1,599 instances), 10% test (201 instances), and 10% dev (200 instances). Then, we fine-tuned various LLMs adding a binary offensive classification task layer on top of the encoder, and training the whole model end-to-end, described as follows. It should be pointed out that although the annotation of hate speech targets may be used to better understand hatred comments in West Africa, we did not implement hate speech targets classifiers due to their smaller size.

AfriBERTa-base[8] (Ogueji et al., 2021) consists of 126 million parameters, 10 layers, 6 attention heads, 768 hidden units, and 3,072 feed-forward sizes. This multilingual model was pretrained on 11 African languages including Hausa.

Afro-XLMR-base[9] (Alabi et al., 2022) was created using MLM adaptation of XLM-R-large model on 17 African languages including Hausa.

[8] https://huggingface.co/castorini/afriberta_large

[9] https://huggingface.co/Davlan/afro-xlmr-base

mBERT-cased[10] (Devlin et al., 2019) consists of multilingual Bidirectional Encoder Representations from Transformers. We held batch size at 64, a maximum of 500 features, a learning rate at 2e-05, the number of epochs at 4, and utilized Keras.

XLM-R-base-Hausa[11] (Adelani et al., 2021) is a "Hausa RoBERTa" model obtained by fine-tuning xlm-roberta-base on the HausaHate corpus. It presents better performance compared to the XLM-RoBERTa on text classification and Named-Entity Recognition (NER) tasks.

6 Evaluation and Results

We evaluated the implemented LLMs described above using Precision, Recall, and F1-Score measures, as shown in Table 4.

Models	Precision	Recall	F1
AfriBERTa_base	80.3	80.1	80.2
Afro-XLMR-base	74.8	75.6	74.8
mBERT-cased	74.3	75.1	73.7
XLM-R-base-Hausa	**85.9**	**86.1**	**85.8**

Table 4: Performance of various fine-tuned LLMs.

Notice that the best performance was obtained using the XLM-R-base-Hausa model with an F1-Score of 85.8, in contrast with the mBERT-cased, which presented the worst performance for the task. This result is based on the fact that multilingual models such as mBERT-cased tend to be more successful to predict texts in English given that they are pretrained on English data. Furthermore, African languages have distinct linguistic characteristics and cultural aspects that may be not totally covered by this multilingual model. Consequently, for subjective tasks such as hate speech and offensive language detection, which are also culturally dependent, monolingual models tend to be more realistic. Lastly, we also observed that AfriBERTa-base is the second-best model. Meanwhile, the Afro-XLMR-base model has a worse result than the XLM-R-base-Hausa, which is a smaller model compared to XLM-R-base-Hausa. Furthermore, the XLM-R-base-Hausa was pretrained on social media data, which is from the same domain as our corpus, thus, showing that LLMs tend to perform better when trained on data from the same domain.

[10]https://huggingface.co/bert-base-multilingual-cased

[11]https://huggingface.co/Davlan/xlm-roberta-base-finetuned-hausa

6.1 Error Analysis

Finally, we also rely on a ROC error analysis of LLMs, as shown by Figure 1. Observe that the XLM-R-base-Hausa, AfriBERTa and Afro-XLMR-base models are most successful to predict Hausa hate speech compared to mBERT-cased multilingual model.

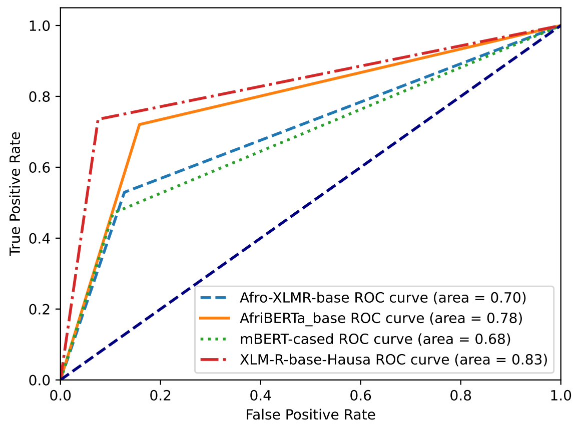

Figure 1: HausaHate Receiver Operating Characteristic (ROC) curves for the various implemented models.

7 Final Remarks and Future Work

This paper provides a benchmark corpus and baseline models for Hausa hate speech detection. The HausaHate corpus was manually annotated by three NLP experts and Hausa native speakers according to two different layers: binary classification (offensive and non-offensive), and hate speech targets (race, gender and none), which obtained substantial annotators agreement. Based on our findings, we concluded that the efforts to counter HS in West Africa should be focused on the detection of racist comments since comments classified as offensive in our corpus are composed mostly of racial hate. Furthermore, a suitable understanding of political conflicts by region is crucial to propose effective HS classifiers for African indigenous languages.

Acknowledgements

This project was partially funded by the CNPq, FAPEMIG, and FAPESP, as well as the Ministry of Science, Technology and Innovation, with resources of Law N. 8.248, of October 23, 1991, within the scope of PPI-SOFTEX, coordinated by Softex and published as Residence in TIC 13, DOU 01245.010222/2022-44.

Ethics Statement

The data collection process was performed using only the publicly available data via the Facebook Graph API [12], along with the CrowdTangle platform. By the very nature of the access used, any users with privacy restrictions are not included in our dataset. Data is downloaded from Facebook pages that are public entities. The content of the comments published on such pages is also available on the Graph API to Facebook developers that are authenticated to access the public data of all pages. If any user has privacy settings changing the privacy of its comments from the default, they become unavailable to us.

Furthermore, we followed the steps to anonymize the data describe in Section 4.2.3, as it is standard for papers with this kind of data. There are public corpus of anonymized Facebook comments available, e.g. Chowdhury et al. (2020). However, since the last change on the Meta platform terms of service was in 2020, we only decided to disclose the ids of the comments (only when requested) in order to allow the reproducibility, while also compelling researchers to pass through Meta's authorization procedures to access the full data.

[12]We followed the use case described on section 3.C fo the terms of service of Meta's platforms `https://developers.facebook.com/terms/dfc_platform_terms/#datause`.

References

Teshome Mulugeta Ababu and Michael Melese Woldeyohannis. 2022. Afaan Oromo hate speech detection and classification on social media. In *Proceedings of the 13th Language Resources and Evaluation Conference*, pages 6612–6619, Marseille, France.

Ridwanullah Abdulhameed. 2021. Analysis of machine sensing of hate speech on twitter in nigeria. *Journal of Artificial Intelligence, Machine Learning and Neural Network*, 1(01):28–44.

David Ifeoluwa Adelani, Jade Abbott, Graham Neubig, Daniel D'souza, Julia Kreutzer, Constantine Lignos, Chester Palen-Michel, Happy Buzaaba, Shruti Rijhwani, Sebastian Ruder, Stephen Mayhew, Israel Abebe Azime, Shamsuddeen H. Muhammad, Chris Chinenye Emezue, Joyce Nakatumba-Nabende, Perez Ogayo, Aremu Anuoluwapo, Catherine Gitau, Derguene Mbaye, Jesujoba Alabi, Seid Muhie Yimam, Tajuddeen Rabiu Gwadabe, Ignatius Ezeani, Rubungo Andre Niyongabo, Jonathan Mukiibi, Verrah Otiende, Iroro Orife, Davis David, Samba Ngom, Tosin Adewumi, Paul Rayson, Mofetoluwa Adeyemi, Gerald Muriuki, Emmanuel Anebi, Chiamaka Chukwuneke, Nkiruka Odu, Eric Peter Wairagala, Samuel Oyerinde, Clemencia Siro, Tobius Saul Bateesa, Temilola Oloyede, Yvonne Wambui, Victor Akinode, Deborah Nabagereka, Maurice Katusiime, Ayodele Awokoya, Mouhamadane MBOUP, Dibora Gebreyohannes, Henok Tilaye, Kelechi Nwaike, Degaga Wolde, Abdoulaye Faye, Blessing Sibanda, Orevaoghene Ahia, Bonaventure F. P. Dossou, Kelechi Ogueji, Thierno Ibrahima DIOP, Abdoulaye Diallo, Adewale Akinfaderin, Tendai Marengereke, and Salomey Osei. 2021. MasakhaNER: Named entity recognition for African languages. *Transactions of the Association for Computational Linguistics*, 9:1116–1131.

Jesujoba O. Alabi, David Ifeoluwa Adelani, Marius Mosbach, and Dietrich Klakow. 2022. Adapting pre-trained language models to African languages via multilingual adaptive fine-tuning. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 4336–4349, Gyeongju, Republic of Korea.

Saminu Mohammad Aliyu, Gregory Maksha Wajiga, Muhammad Murtala, Shamsuddeen Hassan Muhammad, Idris Abdulmumin, and Ibrahim Said Ahmad. 2022. Herdphobia: A dataset for hate speech against fulani in nigeria. *arXiv:2211.15262 [cs.CL]*, pages 1–3.

Ahmad Muhammad Auwal. 2018. Social media and hate speech: Analysis of comments on biafra agitations, arewa youths' ultimatum and their implications on peaceful coexistence in nigeria. *Media and Communication Currents*, 2(1):54–74.

Isa Yusuf Chamo. 2011. Language and identity: Hausa language of youth generation in northern nigeria. *Studies in African Languages and Cultures*, (45):23–38.

Shammur Absar Chowdhury, Hamdy Mubarak, Ahmed Abdelali, Soon-gyo Jung, Bernard J Jansen, and Joni Salminen. 2020. A multi-platform arabic news comment dataset for offensive language detection. In *Proceedings of the 12th language resources and evaluation conference*, pages 6203–6212, Marseille, France.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 17th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 4171–4186, Minnesota, United States.

Kenneth Dike. 1956. *Trade and Politics in the Niger Delta 1830–1835: An introduction to the economic and political history of Nigeria*. Oxford University Press, London.

Christian Ezeibe. 2021. Hate speech and election violence in nigeria. *Journal of Asian and African Studies*, 56(4):919–935.

Paula Fortuna and Sérgio Nunes. 2018. A survey on automatic detection of hate speech in text. *ACM Computing Surveys*, 51(4):1–30.

Marie Garrau. 2020. Une approche psychologique du patriarcat? *Multitudes*, 2(79):186–192.

P.J. Jaggar. 2001. *Hausa*. London Oriental and African Language Library. John Benjamins Publishing Company.

Ibrahim Badamasi Lambu. 2019. Hausanization of nigerian cultures. *Handbook of the Changing World Language Map*, page 145–1153.

Agaptus Nwozor, John Shola Olanrewaju, Segun Oshewolo, Anthony M Oladoyin, Solomon Adedire, and Onjefu Okidu. 2021. Herder-farmer conflicts: The politicization of violence and evolving security measures in nigeria. *African Security*, 14(1):55–79.

Kelechi Ogueji, Yuxin Zhu, and Jimmy Lin. 2021. Small data? no problem! exploring the viability of pretrained multilingual language models for low-resourced languages. In *Proceedings of the 1st Workshop on Multilingual Representation Learning*, pages 116–126, Punta Cana, Dominican Republic.

Morakinyo Ogunmodimu. 2015. Language policy in nigeria: Problems, prospects and perspectives. *International Journal of Humanities and Social Science*, 5(9):154–160.

Oluwafemi Oriola and Eduan Kotzé. 2022. Exploring neural embeddings and transformers for isolation of offensive and hate speech in south african social media space. In *Proceedings of the 22nd International Conference on Computational Science and Its Applicationss*, pages 649–661, Malaga, Spain.

Oluwafemi Oriola and Eduan Kotzé. 2019. Automatic detection of abusive South African tweets using a semi-supervised learning approach. In *Proceedings of the South African Forum for Artificial Intelligence Research FAIR*, pages 90–102, Cape Town, South Africa.

Oluwafemi Oriola and Eduan Kotzé. 2020. Evaluating machine learning techniques for detecting offensive and hate speech in south african tweets. *IEEE Access*, 8:21496–21509.

Vasu Reddy. 2002. Perverts and sodomites: homophobia as hate speech in africa. *Southern African Linguistics and Applied Language Studies*, 20(3):163–175.

Abdulhameed Olaitan Ridwanullah, Sulaiman Ya'u Sule, Bashiru Usman, and Lauratu Umar Abdulsalam. 2024. Politicization of hate and weaponization of twitter/x in a polarized digital space in nigeria. *Journal of Asian and African Studies*, 0(0):1–21.

Leandro Silva, Mainack Modal, Denzil Correa, Fabrício Benevenuto, and Ingmar Weber. 2016. Analyzing the targets of hate in online social media. In *Proceedings of the 10th International AAAI Conference on on Web and Social Media*, pages 687–690, Cologne, Germany.

Francielle Vargas, Isabelle Carvalho, Fabiana Rodrigues de Góes, Thiago Pardo, and Fabrício Benevenuto. 2022. HateBR: A large expert annotated corpus of Brazilian Instagram comments for offensive language and hate speech detection. In *Proceedings of the 13th Language Resources and Evaluation Conference*, pages 7174–7183, Marseille, France.

Francielle Vargas, Fabiana Rodrigues de Góes, Isabelle Carvalho, Fabrício Benevenuto, and Thiago Pardo. 2021. Contextual-lexicon approach for abusive language detection. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing*, pages 1438–1447, Held Online.

William Warner and Julia Hirschberg. 2012. Detecting hate speech on the world wide web. In *Proceedings of the Second Workshop on Language in Social Media*, pages 19–26, Montréal, Canada.

Marcos Zampieri, Shervin Malmasi, Preslav Nakov, Sara Rosenthal, Noura Farra, and Ritesh Kumar. 2019. Predicting the type and target of offensive posts in social media. In *Proceedings of the 17th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1415–1420, Minnesota, United States.