Which Nigerian-Pidgin does Generative AI speak?: Issues about Representativeness and Bias for Multilingual and Low Resource Languages

Anonymous ACL submission

Abstract

Naija is the Nigerian-Pidgin spoken by approx. 120M speakers in Nigeria and it is a mixed language (e.g., English, Portuguese and Indigenous languages). Although it has mainly been a spoken language until recently, there are currently two written genres (BBC and Wikipedia) in Naija as well. Through statistical analyses and Machine Translation experiments, we prove that these two genres do not represent each other (i.e., there are linguistic differences in word order and vocabulary) and Generative AI operates only based on Naija written in the BBC genre which leads to bias in terms of representativeness.

1 Introduction

Nigeria is a multilingual country in the western part of the African continent hosting over 500 different languages spoken by approx. 220 million people (Eberhard et al., 2019). English is the official language and acquired mostly through formal education (Agbo and Plag, 2020). Nigerian Pidgin (Naija) is a mix of different languages (e.g., English, Portuguese, Indigenous languages) and it is widely spoken (approx. 120M speakers) as a first and second language (Adelani, 2022) around the Southern part of Nigeria (e.g., Lagos and Niger-Delta) with origins going back to the English-Creole Atlantic Krio language family. It is also adopted as the unifying and unofficial language for communication across ethnically diverse groups.

Despite the large number of speakers, Naija has remained a spoken language until 2017 when the British Broadcasting Company (BBC) launched a news website in written Naija targeting audiences from West Africa. [1] Since 2022, Naija is also accepted as one of the languages on Wikipedia. [2]. Although both BBC and Wikipedia claim to represent Naija as one of their languages, there are linguistic and social differences between the two written genres. For example, Naija BBC genre resembles English in terms of word order and vocabulary (see example (1)) with a simplified grammar (i.e., lacking auxiliary "were"). However, Naija Wikipedia genre is different than English in terms of word order and vocabulary choice (e.g., ”moto” instead of a "car" and ”wund" instead of ”injured" or ”wounded").

Example (1)

BBC Genre: Two pesin in di car dey injured.
Wikipedia Genre: Na wund di two pesin get for di moto.
English Translation: Two persons in the car were injured.

From the social perspective, Naija BBC genre is favored mostly by educated Nigerians whereas Wikipedia genre is closer to spoken Naija in everyday life and it is more accessible for larger groups regardless of their background (e.g., educational status). So far, Naija BBC genre has also more data available on the Internet.

It is well-known that there is a need for more research for multilingual and low resource languages (Doğruöz and Sitaram, 2022; Doğruöz et al., 2021) in Generative AI systems. This need is even enhanced for pidgin and creole languages due to their high numbers of speakers but lack of data (Lent et al., 2022, 2023). However, it is equally important that the different genres of the same pidgin and creole languages are also represented in these systems to be inclusive and accessible for all speakers/users with diverse backgrounds (e.g., educated vs. less educated).

In our paper, we address these issues with the following contributions. We introduce WARRI as a new MT evaluation data set including the two genres (BBC and Wikipedia) of Naija as a multi-

[1] https://www.bbc.com/news/world-africa-40975399

[2] https://meta.wikimedia.org/wiki/Requests_for_new_languages/Wikipedia_Nigerian_Pidgin

lingual and low resource language. [3] Our paper is the first to systematically analyze the similarities and differences between the two written genres of Naija and find that they do not represent each other. Through a Machine Translation experiment, we find that Generative AI models (e.g., GPT-4-TURBO and LLAMA 2 13B (Touvron et al., 2023)) are biased towards the BBC genre (favored by more educated groups in Nigeria). Hence, these systems do not represent the Wikipedia genre and its speakers (without education barriers).

2 Related Work and Links with Our Study

Available research on representativeness originates from corpus linguistics where it is important to include samples from different textual sources to have a balanced and representative (smaller) corpus reflecting the variation in the (larger) corpora (Biber, 1993). Similarly, Crowdy (1993) states the significance of representative sampling corpora to minimize bias and maximize the credibility and consistency of the linguistic analyses. Therefore, representative sampling encompasses a broad spectrum of language usage across various contexts, genres, and demographic factors.

Generative AI systems depend on the availability of large data sets on the Internet. However, this assumption does not consider the representativeness of the variation in the available data sets which is especially difficult for multilingual and low resource languages (Doğruöz et al., 2023). While developing language technologies for multilingual and low resource languages, it is crucially important to be aware of the linguistic variation (e.g., across genres) in these languages and aim for representing the variation in a balanced way to prevent potential bias.

To investigate bias (e.g., favoring one genre over the other one in Naija), the first step is to establish to what extent these data sets (i.e., BBC and Wikipedia genres of Naija in our case) represent each other through establishing the linguistic similarities and differences between them. If there are many differences, these data sets will not be representative of each other. Hence, they should be integrated into the systems separately to represent the variation in a balanced way.

[3] We will release the dataset in CC-BY-4.0 license.

Dataset	Genre	TRAIN	DEV	TEST
MAFAND	-	4,790	1,484	1, 564
WARRI (single-way)	BBC	5	-	500
WARRI (multi-way)	BBC & Wiki	5	-	500

Table 1: **WARRI and MAFAND dataset:** WARRI is only used for evaluation in zero or few-shot (e.g. 5) setting. WARRI (multi-way) have the same sentences in both BBC and Wiki genre unlike WARRI (single-way).

3 WARRI MT benchmark dataset for Naija Genres

To establish the WARRI data set, we used the Naija data in BBC genre from the MasakhaNER dataset (Adelani et al., 2021b) and downloaded Wikipedia articles in Naija from the HuggingFace dataset hub.[4] After the data collection, we created the parallel data set in English by recruiting two bilingual (i.e., in Naija and English) speakers. They translated about 505 sentences from the BBC news data and Wikipedia articles (both in Naija) into English. In this way, we maintained a high-quality dataset for the two genres (i.e., BBC and Wikipedia) in Naija, prevented the translators from mixing the features of the two different genres into one and we obtained two test sets for each genre. However, this also introduced a new obstacle (i.e., comparison of two test sets from slightly different domains (news vs. Wikipedia).

To handle the domain related obstacle, we created a **multi-way parallel dataset** for Naija Wikipedia genre. First, we asked a bilingual speaker to translate the Naija (Wikipedia genre) sentences into English. Then, we asked a professional translator (a different person), to translate the English sentences into Naija BBC genre.

Table 1 provides the details of our new WARRI dataset, containing a **single-way parallel sentences** translated by two native speakers from Naija BBC genre to English, and the **multi-way parallel sentences** where the English sentences are translated into Naija BBC and Wikipedia genres separately. Our test set composed of 500 sentences and the remaining five sentences were for few-shot/in-context learning for LLMs.

Other datasets in Naija There are also other parallel datasets available in Naija (e.g., Bible (Akerman et al., 2023), JW300 (Agić and Vulić, 2019),MAFAND (Adelani et al., 2022), UD-

[4] https://huggingface.co/datasets/wikimedia/wikipedia

Metric	single-way BBC genre	multi-way BBC genre	multi-way Wikipedia
Jaccard Similarity ([0,1] range)			
Unigram	0.712	0.802	0.517
Bigram	0.289	0.371	0.167
Trigram	0.151	0.207	0.084
Levenshtein distance	26.6	21.6	53.6

Table 2: **Lexical overlap and Levenshtein distance on WARRI benchmark**. Lexical overlap is measured by Jaccard similarity between English and Naija BBC and Wikipedia genres.

Pidgin,[5]). The first two data sets are in religious domain and the last one is based on a spoken conversation and contains only short sentences. To disentangle the genre effect from the domain effects (since domain transfer can lead to huge drop in performance (Adelani et al., 2021a; Lee et al., 2022)), we did not consider the conversational and religious parallel data sets for this study. The only training data we considered is MAFAND which is a parallel corpus based on news domain covering about 4,790 high-quality parallel training sentences in Naija.

4 Experimental setup

We conduct two types of experiments: (1) Statistical analysis of the texts obtained from two written genres in Naija (i.e., BBC and Wikipedia) to find out whether they represent each other. (2) Evaluation of WARRI MT benchmark dataset trained on MAFAND or prompting an LLM to find out which Naija genre is represented in Generative AI.

Statistical analysis of the texts First, we compute lexical similarity between the English portion of WARRI dataset and Naija by measuring **Jaccard similarity** (in percentage) for each corpus unigram, bigram, and trigram tokens. Secondly, we compute the Levenshtein distance (Levenshtein, 1965) which is an edit distance between WARRI (multi-way) English test sentences and their translations in either BBC genre or Wikipedia genre.

Evaluation of WARRI MT benchmark dataset Following Adelani et al. (2022), we leveraged a pre-trained model to train an MT model by fine-tuning M2M-100 (418M) on a few thousand sentences (i.e. 4,790 sentences of MAFAND training set), and evaluated on WARRI test sets. Furthermore, we prompted GPT-4-TURBO [6] and LLAMA 2 13B (Touvron et al., 2023) to generate translations in either Naija or English in both zero-shot or few-shots settings (with one or five examples). We also provide a sample prompt in Appendix A.

5 Experimental Results

5.1 Statistical analysis results

In Table 2, by computing a lexical similarity between the n-gram tokens of each genre, we show that Naija Wikipedia genre consistently has a lower Jaccard similarity score with its parallel English corpus for all n-grams. For example, the unigram similarity score for Naija BBC genre was around $0.712 - 0.802$ while Naija Wikipedia genre only achieved 0.517. Furthermore, Levenshtein distance provides an additional evidence of a clear difference between Naija BBC and Wikipedia genres. We find that it takes more than twice edit-distance to transform the English sentences to Naija Wikipedia genre than to Naija BBC genre. Naija Wikipedia genre requires more edits in characters, which shows that it is farther from English compared to the BBC genre. In other words, these two genres are quite different than each other linguistically and they do not represent each other.

5.2 WARRI MT Benchmark Results

While statistical analysis are useful, evaluation on a practical NLP task (e.g., MT) shows the benefit of having a MT system that supports different genres which are accessible for different communities of Naija speakers.

MAFAND MT model and LLMs represent BBC genre more Table 3 shows the result of evaluation of the WARRI MT results. In the direction of **pcm→en**,[7] adapting MAFAND MT model to BBC genre gave an impressive result in both single-way (76.7 ChrF++) and multi-way parallel (83 ChrF++) scenarios, however the performance on Wikipedia genre is much worse (-24.3 drop in ChrF++). This shows that the fine-tuning corpus most likely represents the Naija BBC genre. Similar observation was found in GPT-4-TURBO and LLAMA 2 13B evaluation, although the performance of the latter was worse especially on the Naija Wikipedia genre. Similarly, for the **en→pcm**, in zero-shot setting, MAFAND MT model gave the best performance over GPT-4-TURBO and LLAMA 2 13B on the BBC genre, but competitive performance on the

[5] https://github.com/UniversalDependencies/UD_Naija-NSC

[6] GPT-4-TURBO pre-training data is up to December 2023.

[7] **pcm** is the ISO 639-3 code for Naija

Evaluation Task	Single-way (news) BBC genre BLEU	Single-way (news) BBC genre ChrF++	Multi-way parallel (Wiki) BBC genre BLEU	Multi-way parallel (Wiki) BBC genre ChrF++	Multi-way parallel (Wiki) Wikipedia genre BLEU	Multi-way parallel (Wiki) Wikipedia genre ChrF++
`pcm` → `en`						
0-shot: MAFAND → WARRI	**57.9**	**76.7**	**68.7**	83.4	35.0	59.1
0-shot: GPT-4-TURBO	52.4	75.7	65.8	**84.6**	**43.0**	**70.5**
0-shot: LLAMA 2 13B	43.1	70.6	44.1	73.1	29.5	58.5
`en` → `pcm`						
0-shot: MAFAND → WARRI	**55.0**	**75.5**	**61.0**	**79.5**	26.5	51.8
0-shot: GPT-4-TURBO	45.2	69.2	39.5	68.6	26.4	54.4
1-shot: GPT-4-TURBO	50.2	72.9	51.3	76.0	28.9	56.7
5-shot: GPT-4-TURBO	54.1	**75.5**	56.1	77.7	**29.2**	**57.1**
0-shot: LLAMA 2 13B	31.6	62.4	33.9	65.7	19.2	48.5
1-shot: LLAMA 2 13B	36.4	66.8	37.2	68.5	15.2	46.3
5-shot: LLAMA 2 13B	36.2	64.5	42.8	69.3	18.4	44.7

Table 3: **Evaluation on WARRI dataset: single-way and multi-way parallel (same sentences translated to both genres) test sets**: We compared the performance of MT to different genres using GPT-4-Turbo and adapted M2M-100 (418M) from MAFAND training set.

Method	Translation
English	The name "ABIA" comes from the first letter of the four highly populated places in the state.
Reference (Wikipedia)	Di fest leta for di four ples for di Stet wey pipul plenti pass mek-up di nem ABIA dey bi
Reference (BBC)	Di name "ABIA" be from di first letter of di four highy populated place dem for di state.
`Wikipedia genre`	
0-shot: GPT-4-TURBO	Di name "ABIA" na from di first letter of di four places wey get plenty people for di state.
1-shot: GPT-4-TURBO	Di name "ABIA" na from di first leta of di four plenti pipul dey areas for di state.
5-shot: GPT-4-TURBO	Di name "ABIA" na from di first leta of di four places wey get plenti pipul for di state.
`BBC genre`	
0-shot: GPT-4-TURBO	Di name "ABIA" na from di first letter of di four places wey people full well well for di state.
1-shot: GPT-4-TURBO	Di name "ABIA" na from di first letter of di four places wey get plenty people for di state.
5-shot: GPT-4-TURBO	Di name "ABIA" come from di first letter of di four places wey people plenty well well for di state.

Table 4: **Qualitative analysis on Predicted translations on WARRI dataset:** multi-way parallel output

Wikipedia genre (51.8 ChrF++) compared to GPT-4-TURBO (54.4 ChrF++).

Can we teach LLMs different genres with only a few examples? Our result (Table 4) of prompting both GPT-4-TURBO and LLAMA 2 13B shows that providing one or five examples is effective in providing more boost in performance for generating Naija sentences. For GPT-4-TURBO, the performance improved over zero-shot result by +7.4 ChrF++ with the LLM is prompted with one example translation of the BBC genre, and +8.4 ChrF++ when prompted with five examples, on the multi-way test set. However, the boost in performance is very small when Naija Wikipedia genre examples are provided. The boost is only +2.3 and +2.7 when one example and five examples are provided during the prompting of GPT-4-TURBO. This clearly shows that GPT-4-TURBO is more biased toward the Naija BBC genre than Naija Wikipedia genre, and it is difficult to teach LLM with few examples. We provide qualitative examples in Table 4, where we show that with one or five examples, the GPT-4 LLM slightly changes its writing style to be more similar to Naija Wikipedia genre but this change leads to loss of meaning as well. For example, "**four plenti pipul dey areas**" ("four plenty people are in the areas of the state" instead of "four highly populated places" lost the meaning but the Wikipedia genre is preserved.

6 Conclusion

Between 2017-2022, Naija was written only in the BBC genre which is closer to English and mainly used among educated populations in Nigeria. After 2022, a Wikipedia genre which is closer to the spoken Naija and used by a diversity of speakers was also available. In our short paper, we were able to prove that 1) these two genres do not represent each other (i.e., they are very different from each other linguistically) and 2) Naija Wikipedia genre is not represented in Generative AI models which are currently biased towards the BBC genre. The second finding is probably due to more availability of data on the Internet for Naija BBC genre which leads to bias towards favoring language preferences of certain groups (e.g. educated) instead of being more inclusive.

7 Limitation

There are few limitations of our work (1) Our evaluation dataset is small, although we argue that 500 may be good enough as a test set for MT, however, we only have maximum of 5 sentences we could use for few-shot learning or in-context learning. Moreover, with additional sentences (e.g. 2.5K-5K parallel sentences as recommended in (Adelani et al., 2022)), we may be able to adapt M2M-100 model to produce better generation of the Wikipedia genre. (2) Our analysis is limited to one task which is machine translation, we hope to extend this analysis to other tasks in the future. (3) our analysis only cover Naija spoken in Nigeria, we hope to extend our analysis to other English-based creole in West Africa like Ghananian Pidgin, Cameronian Pidgin, and Krio in the future.

References

David Adelani, Jesujoba Alabi, Angela Fan, Julia Kreutzer, Xiaoyu Shen, Machel Reid, Dana Ruiter, Dietrich Klakow, Peter Nabende, Ernie Chang, Tajuddeen Gwadabe, Freshia Sackey, Bonaventure F. P. Dossou, Chris Emezue, Colin Leong, Michael Beukman, Shamsuddeen Muhammad, Guyo Jarso, Oreen Yousuf, Andre Niyongabo Rubungo, Gilles Hacheme, Eric Peter Wairagala, Muhammad Umair Nasir, Benjamin Ajibade, Tunde Ajayi, Yvonne Gitau, Jade Abbott, Mohamed Ahmed, Millicent Ochieng, Anuoluwapo Aremu, Perez Ogayo, Jonathan Mukiibi, Fatoumata Ouoba Kabore, Godson Kalipe, Derguene Mbaye, Allahsera Auguste Tapo, Victoire Memdjokam Koagne, Edwin Munkoh-Buabeng, Valencia Wagner, Idris Abdulmumin, Ayodele Awokoya, Happy Buzaaba, Blessing Sibanda, Andiswa Bukula, and Sam Manthalu. 2022. A few thousand translations go a long way! leveraging pre-trained models for African news translation. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 3053–3070, Seattle, United States. Association for Computational Linguistics.

David Adelani, Dana Ruiter, Jesujoba Alabi, Damilola Adebonojo, Adesina Ayeni, Mofe Adeyemi, Ayodele Esther Awokoya, and Cristina España-Bonet. 2021a. The effect of domain and diacritics in Yoruba–English neural machine translation. In *Proceedings of Machine Translation Summit XVIII: Research Track*, pages 61–75, Virtual. Association for Machine Translation in the Americas.

David Ifeoluwa Adelani. 2022. Natural language processing for african languages.

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. 2021b. MasakhaNER: Named entity recognition for African languages. *Transactions of the Association for Computational Linguistics*, 9:1116–1131.

Ogechi Florence Agbo and Ingo Plag. 2020. The relationship of nigerian english and nigerian pidgin in nigeria: Evidence from copula constructions in ice-nigeria. *Journal of Language Contact*, 13(2):351–388.

Željko Agić and Ivan Vulić. 2019. JW300: A wide-coverage parallel corpus for low-resource languages. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 3204–3210, Florence, Italy. Association for Computational Linguistics.

Vesa Akerman, David Baines, Damien Daspit, Ulf Hermjakob, Tae Young Jang, Colin Leong, Michael Martin, Joel Mathew, Jonathan Robie, and Marcus Schwarting. 2023. The ebible corpus: Data and model benchmarks for bible translation for low-resource languages. *ArXiv*, abs/2304.09919.

Douglas Biber. 1993. Representativeness in corpus design. *Literary and Linguistic Computing*, 8:243–257.

Steve Crowdy. 1993. Spoken corpus design. *Literary and Linguistic Computing*, 8:259–265.

A. Seza Doğruöz and Sunayana Sitaram. 2022. Language technologies for low resource languages: Sociolinguistic and multilingual insights. In *Proceedings of the 1st Annual Meeting of the ELRA/ISCA Special Interest Group on Under-Resourced Languages*, pages 92–97, Marseille, France. European Language Resources Association.

A. Seza Doğruöz, Sunayana Sitaram, Barbara E. Bullock, and Almeida Jacqueline Toribio. 2021. A survey of code-switching: Linguistic and social perspectives for language technologies. 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 1654–1666, Online. Association for Computational Linguistics.

A. Seza Doğruöz, Sunayana Sitaram, and Zheng Xin Yong. 2023. Representativeness as a forgotten lesson for multilingual and code-switched data collection and preparation. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 5751–5767, Singapore. Association for Computational Linguistics.

David M Eberhard, Gary F Simons, and Charles D Fennig. 2019. Ethnologue: Languages of the world(22nd edn.). dallas, tx: Sil international. *Online version: http://www. ethnologue. com [08.08. 2020]*.

En-Shiun Lee, Sarubi Thillainathan, Shravan Nayak, Surangika Ranathunga, David Adelani, Ruisi Su, and Arya McCarthy. 2022. Pre-trained multilingual sequence-to-sequence models: A hope for low-resource language translation? In *Findings of the Association for Computational Linguistics: ACL 2022*, pages 58–67, Dublin, Ireland. Association for Computational Linguistics.

Heather Lent, Kelechi Ogueji, Miryam de Lhoneux, Orevaoghene Ahia, and Anders Søgaard. 2022. What a creole wants, what a creole needs. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 6439–6449, Marseille, France. European Language Resources Association.

Heather Lent, Kushal Tatariya, Raj Dabre, Yiyi Chen, Marcell Richard Fekete, Esther Ploeger, Li Zhou, Hans Erik Heje, Diptesh Kanojia, Paul Belony, Marcel Bollmann, Loic Grobol, Miryam de Lhoneux, Daniel Hershcovich, Michel DeGraff, Anders Sogaard, and Johannes Bjerva. 2023. Creoleval: Multilingual multitask benchmarks for creoles. *ArXiv*, abs/2310.19567.

Vladimir I. Levenshtein. 1965. Binary codes capable of correcting deletions, insertions, and reversals. *Soviet physics. Doklady*, 10:707–710.

Hugo Touvron, Louis Martin, Kevin R. Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, Soumya Batra, Prajjwal Bhargava, Shruti Bhosale, Daniel M. Bikel, Lukas Blecher, Cristian Cantón Ferrer, Moya Chen, Guillem Cucurull, David Esiobu, Jude Fernandes, Jeremy Fu, Wenyin Fu, Brian Fuller, Cynthia Gao, Vedanuj Goswami, Naman Goyal, Anthony S. Hartshorn, Saghar Hosseini, Rui Hou, Hakan Inan, Marcin Kardas, Viktor Kerkez, Madian Khabsa, Isabel M. Kloumann, A. V. Korenev, Punit Singh Koura, Marie-Anne Lachaux, Thibaut Lavril, Jenya Lee, Diana Liskovich, Yinghai Lu, Yuning Mao, Xavier Martinet, Todor Mihaylov, Pushkar Mishra, Igor Molybog, Yixin Nie, Andrew Poulton, Jeremy Reizenstein, Rashi Rungta, Kalyan Saladi, Alan Schelten, Ruan Silva, Eric Michael Smith, R. Subramanian, Xia Tan, Binh Tang, Ross Taylor, Adina Williams, Jian Xiang Kuan, Puxin Xu, Zhengxu Yan, Iliyan Zarov, Yuchen Zhang, Angela Fan, Melanie Kambadur, Sharan Narang, Aurelien Rodriguez, Robert Stojnic, Sergey Edunov, and Thomas Scialom. 2023. Llama 2: Open foundation and fine-tuned chat models. *ArXiv*, abs/2307.09288.

A Prompt Template

B MAFAND training

We fine-tune MAFAND dataset on M2M-100 (418M) using the same hyparameters stated in Adelani et al. (2022) i.e. number of training epochs of 10, batch size of 32, source and target maximum sequence length of 200, and beam size of 10.

C Licence of WARRI

We plan to release it publicly under the CC-4.0-NC due to BBC portion of the dataset that cannot be for commercial use. However, WARRI (multi-way) has a licence of CC-4.0 international.

D LLAMA 2 13B qualitative results

We added the LLAMA 2 13B results to the first example in Table 4 in the paper. LLAMA 2 13B struggles to generate translation in similar style as the Wikipedia genre, it only introduce the articles "dey" and "di". Prompting with zero or one example is mostly in English, while prompting with five examples gave us a sentence that is similar to the BBC genre despite prompting it to produce Wikipedia genre output. This highlights the bias of LLAMA 2 13B to the BBC genre. Morever, prompting with the BBC genre seem to work better, but the meaning of the sentence is often affected, and sometimes spelling issues like "firs" and "highly populate"—which is unnatural to the way a native speaker would speak.

	Prompt
Task Description	You are a helpful assistant who is an expert in translating English sentences to Nigerian Pidgin using two genres: BBC genre and Wikipedia genre, I would provide you with one example of the different genres, your task is to follow the style of the writing of the specified genre when translating the sentences.
Example	**Example 1:** **English**: Innocent Ujah Idibia was born on 18 September 1975, that is well known as 2baba, a Nigerian singer, songwriter, producer, philantropist. **BBC Genre**: Innocent Ujah Idibia wey dem born for 18 September 1975, wey dem know as 2baba, be a Nigerian singer, songwriter, producer, philantropist. **Wikipedia Genre**: Innocent Ujah Idibia (dem bon am for 18 September 1975), wey pipul no wel wel as 2baba, na Naija singa, songraita an podusa an im sabi dash pipul moni an gift wel wel.
Example	**Example 2:** **English**: He was born in Jos, Nigeria **BBC Genre**: Dem born am for Jos, Nigeria **Wikipedia Genre**: Dem bon am for Jos for inside Naija.
Example	**Example 3:** **English**: He is from the Idoma ethnic group **BBC Genre**: Im be from di Idoma ethnic group **Wikipedia Genre**: Im na Idoma pesin.
Example	**Example 4:** **English**: Idoma is in the southern part of Nigeria **BBC Genre**: Na southern part of Nigeria Idoma dey **Wikipedia Genre**: Idoma dey for di south side for Naija.
Example	**Example 5:** **English**: Before July 2014, he used 2face Idibia as his stage name **BBC Genre**: Before July 2014, i dey use 2face Idibia as im stage name **Wikipedia Genre**: Bifor July 2014 na 2face Idibia bi di nem wey im dey yuz for stej.
Prompt	'Translate this sentence to Nigerian Pidgin using the **Wikipedia** genre'
Input	Alexander Abolore Adegbola Akande was born on 17 January 1980, well known as 9ice, a Nigerian singer, dancer, and songwriter.
Output:	Alexander Abolore Adegbola Akande (dem bon am for 17 January 1980), wey pipul sabi well well as 9ice, na Naija singa, dansa, an songraita.

Table 5: **Prompt template used for MT**. An example prediction by GPT-4-TURBO

Method	Translation
Example 1	
English	The name "ABIA" comes from the first letter of the four highly populated places in the state.
Reference (Wikipedia)	Di fest leta for di four ples for di Stet wey pipul plenti pass mek-up di nem ABIA dey bi
Reference (BBC)	Di name "ABIA" be from di first letter of di four highy populated place dem for di state.
`Wikipedia genre`	
0-shot: GPT-4-TURBO	Di name "ABIA" na from di first letter of di four places wey get plenty people for di state.
1-shot: GPT-4-TURBO	Di name "ABIA" na from di first leta of di four plenti pipul dey areas for di state.
5-shot: GPT-4-TURBO	Di name "ABIA" na from di first leta of di four places wey get plenti pipul for di state.
0-shot: LLAMA 2 13B	Name 'ABIA' dey come from first letter of four highly populated place for state.
1-shot: LLAMA 2 13B	The name 'ABIA' dey come from the first letter of the four highy populated places for the state.
5-shot: LLAMA 2 13B	The name "ABIA" come from di first letter of di four highly populate place for di state.
`BBC genre`	
0-shot: GPT-4-TURBO	Di name "ABIA" na from di first letter of di four places wey people full well well for di state.
1-shot: GPT-4-TURBO	Di name "ABIA" na from di first letter of di four places wey get plenty people for di state.
5-shot: GPT-4-TURBO	Di name "ABIA" come from di first letter of di four places wey people plenty well well for di state.
0-shot: LLAMA 2 13B	ABIA get am name pass ABC, na im first letter be four high populated place for the state.
1-shot: LLAMA 2 13B	ABIA de name come from the first letter of four highy populated places for the state.
5-shot: LLAMA 2 13B	ABIA dey come from di firs letter of di four highy populate place for di state.

Table 6: **Qualitative analysis on Predicted translations on WARRI dataset:** multi-way parallel output