TANGALENLP: BUILDING PO TANGLE TO ENGLISH PARALLEL CORPORA AND MACHINE TRANSLATION OF THE TANGLE (TANGALE) LANGUAGE

Gideon George[1,2], Olubayo Adekanmbi[1], Anthony Soronnadi[1], Amina Sambo[2], Olufunke Vincent[2], and Nebath Tanglang[2]

[1]Data Science Nigeria (DSN)
[2]Africa Centre of Excellence on Technology Enhanced Learning (ACETEL)

ABSTRACT

In a digitally connected world, language barriers are silencing millions, leaving communities like Tangle (Tangale) with limited access to information and online opportunities, and their rich heritage fading. This research offers hope that natural language processing and machine translation can bridge this gap. Our efforts go beyond Po Tangle. We are paving the way for similar systems in other African languages and promoting a more diverse digital space. We have successfully created a Po Tangle-English machine translation system using state-of-the-art AI by fine-tuning the pre-trained M2M100 model using 1150 parallel sentences from the dataset and obtained results showing that the system works and produces translations. The system achieves an evaluation BLEU score of 6.7604 and a prediction BLEU score of 6.0101. This indicates the potential for fluent translations with more substantial data. By building a parallel corpus with native speakers to ensure cultural authenticity, we are discovering much more than just numbers. This empowers communities to take control, enabling socio-economic development and preserving linguistic heritage. Our research is having an impact in the form of more targeted interventions, better education, and more vibrant online communities. It is paving the way for a future where every voice is heard and celebrated, regardless of language. This is a movement towards inclusion and equality, we are breaking down language barriers, celebrating the symphony of human voices, and ensuring that no community is left behind in the digital age.

1 INTRODUCTION

Multilingualism is not only a linguistic reality but also a matter of fundamental rights in an increasingly digital world. According to Ethnologue (a), there are 7,164 languages worldwide, of which about 3,000 are African. Nigeria alone has about 525 languages. Many languages, such as Tangale, are either neglected or underserved by the natural language processing community despite this rich linguistic diversity. Ethnologue (b) Tangale is an endangered indigenous language of Nigeria and belongs to the Afro-Asiatic language family. It is the first language of all adults in the ethnic community, but not all young people speak it and it is not known to be taught in schools. With limited access to essential information and online resources, these linguistic communities risk cultural isolation and exclusion from the benefits of the digital age. Orife (2020) argued that many Nigerian languages have experienced a decline in status compared to English and Nigerian Pidgin. This has led to inequalities in access to information and social engagement. A promising way of preserving endangered languages and promoting economic and social inclusion is the advent of machine translation. In this paper, we present a machine translation approach for the language pair of Po Tangle-English. This approach could pave the way for the digitization of such languages in Nigeria and Africa at large.

Tangle (Tangale) is a language of the Western Chadian region of Nigeria. The vast majority of native speakers live in the Akko, Billiri, Kaltungo, and Shongom local government areas of Gombe State in

the northeast region of the country. Billiri/Akko belongs to the western Tangale communities, while Kaltungo/Shongom belongs to the eastern Tangale communities. Although the language speakers are less than a million, it has a unique linguistic identity and strong cultural heritage. In the digital space, however, the language has very little presence. This is a barrier to accessing vital information and learning materials. It hinders progress and the facilitation of communication in education, healthcare, and online businesses, and it also contributes to the perpetuation of economic and social inequalities.

Motivated by a commitment to linguistic inclusion and cultural equity, this research aims to empower speakers of the Po Tangle (Tangale language). A Po Tangle-English parallel corpus will be created. This will be a fundamental resource for the training of a robust machine translation system. By harnessing the power of generative AI models adapted to African languages, this system will bridge the communication gap between Po Tangle and English. It will ensure that Tangale communities continue to be able to access the wealth of information available in the digital world.

1.1 The Historical and Cultural Context of the Tangle Language

An examination of the orthography and linguistic features of the Tangle language, as studied by Tadi & Zakayo (2008b) on Po Tangle orthography, reveals a deep history of colonial influence and cultural identity. The term 'Tangle' itself was imposed by British colonialists and missionaries as a label to differentiate between eastern and western Tangale communities. It was a contentious term, particularly among the people of Kaltungo and Shongom, who rejected it and reasserted their identity by changing the title of their Christian hymn book. The prefix 'Po', as in 'Po Tangle', which is used to denote the language, is central to Tangle's identity. This linguistic marker is a reflection of a wider cultural resistance to external categorization.

The review highlights that the issue with the language name, as stated by the British and missionaries, is on the word 'Po' alone, not on 'Po Tangle'. This is because even before the advent of the British and missionaries, the speakers called themselves and the language Tangle. It is also important to note that the well-known name 'Tangale', denoting this group of language speakers, was an aberration by the Hausas. Additionally, 'Pok' means mouth in Tangle, which is used to refer to language, so 'Po Tangle' can be understood as the language of Tangle. The omission of 'k' is due to the rich elision in the spoken form of Tangle.

As part of the Bole-Tangale group of Chadic languages, Tangle has significant dialectal differences, particularly in vocabulary and phonology, which challenge mutual understanding, especially between Kaltungo/Shongom and Billiri/Akko speakers. However, increased socialization and interaction between the clans are gradually reducing these differences.

There are many dialects within Tangle. These include Tangaltong, Kalmai, Banganje, Tanglang, Tal and Todi. Despite these variations, the Tangaltong dialect was strategically chosen for the orthography because of its importance in Billiri, the nerve center of Tangle speakers. The orthography of the Tangle language has evolved from an initial 25-letter alphabet to the current 21-letter system. This system is common to all Tangle dialects in the Billiri and Akko Local Government Areas.

There are nine vowel sounds and twenty-eight consonant sounds in this system.

Elisions and diphthongs add to the complexity of the language's phonological structure. They reveal its richness and complexity. Labialized sounds further contribute to its unique phonetic features and enhance its distinctiveness within the Chadic language family.

Besides the language orthography, the Po Tangle-English Dictionary and Vocabulary, a publication of the Po-Tangle Committee is also instrumental in the translations of the Parallel sentences in the making of the dataset for this machine translation's task, which is authored by Tadi & Zakayo (2008a). This proves that there is a collection of corpora in Tangle.

The Tangle language is more than a means of communication. It is a testimony to cultural resilience and identity. While its orthographic and linguistic features are diverse and complex, they also reflect a deep-rooted attachment to heritage and community.

The goal is to go beyond the mere advancement of the technology. We envision a future in which Tangle users can seamlessly navigate the digital landscape, engage with news, participate in online discourse, and have access to essential, innovative, and cutting-edge products and information in

their native language. This machine translation system has the potential to empower the Tangale community, promote economic opportunity, and contribute to a more diverse and equitable digital space by fostering linguistic inclusion and cultural preservation of African languages.

2 BACKGROUND AND RELATED WORK

The African linguistic scene is rich with a wealth of diverse languages, each carrying a unique cultural message. However, many are hampering communication due to the lack of robust machine translation (MT) systems. In this review, we make considerable efforts to build on the abundant research that has been fashioned by past efforts.

We can envision a future where Tangale speakers can navigate the digital world seamlessly, harnessing emerging technological products like the accurate and culturally sensitive MT. This dream is in line with the path-breaking work of Adelani et al. (2022), who have demonstrated the potential of pre-trained models for news translations in Africa. We are inspired by their emphasis on fine-tuning with a limited amount of high-quality data, a technique that is crucial for resource-poor languages such as Po Tangle.

Just as different players are needed for a team, MT thrives on the participation of the local community. Stressing the vital role of collaborative efforts and larger parallel corpora, Nekoto et al. (2022) uphold this view. Their work provides the basis for our commitment to involving individuals from the Tangale community in building the parallel corpora and designing and maintaining the Machine Translation system.

In addition to accuracy, the issue of cultural sensitivity is also of utmost importance. The recent work of Vegi et al. (2022) on ANVITA-African, a multilingual neural MT system specifically designed for African languages, encourages us to investigate similar ways of making the translation of Po Tangle more effective in its diverse linguistic context.

The hard reality of languages such as Nko, highlighted by Doumbouya et al. (2023), further underlines the urgency of our mission. Their study of translation assisted by software reminds us of the challenges posed by limited corpora. It urges us to look for creative solutions for the TangaleNLP project.

Nwafor & Andy (2022) call for greater involvement in MT research in African languages which resonates with us entirely, hence we see this machine translation system not only as a technological triumph but also as an enabler for bridging the communication and socio-economic development gaps in Africa.

The key to progress is standardization indeed. The work of Reid et al. (2021) to establish AFROMT as a benchmark for African languages, and Ezeani et al. (2020) to create evaluation tools for Igbo, encourages us to work towards the development of a robust framework for TangaleNLP MT.

The power of community initiatives is immense. The efficacy of local efforts to create resources for low-resource languages is demonstrated by Akera et al. (2022). Their strong emphasis on practical applications and open datasets is consistent with our vision of making this machine translation system available and empowering the Tangle community.

There is hope, even for extremely resource-poor languages. Tapo et al. (2020) demonstrated this in their case study of *Bambara*, providing some valuable insights into how to cope with limited data and achieve acceptable translation quality. Their work fuels our optimism for TangaleNLP MT.

Orife et al. (2020) showed that MASAKHANE is an excellent example of collective action, bringing researchers together to tackle the challenges of resource scarcity and acceptance of African languages. We join them in believing that collaboration is the key to unlocking the transformable potential of MT for Tangle and beyond.

The innovative and collaborative research of Nekoto et al. (2020) paves the way for the inclusive and sustainable development of machine translation. With their commitment to human evaluation and the involvement of local agents, we are encouraged to prioritize cultural sensitivity and ethical practices at every stage of our journey.

Finally, Ogueji & Ahia (2019) progress in unsupervised MT for West African pidgin English provides us with valuable insights. Their work demonstrates the potential of MT to bridge linguistic divides and promote cross-cultural understanding.

This research reflects a common dream: a world where language barriers fall and are replaced by bridges of understanding. Following these threads, we are embarking on the construction of a robust and culturally sensitive TangaleNLP MT system, paving the way for communication, knowledge sharing, and cultural preservation. The journey has only just begun. But with lessons learned from past efforts and the unwavering support of Data Science Nigeria (DSN) and the Africa Centre of Excellence on Technology Enhanced Learning (ACETEL), we believe this dream cannot only become a reality but already awaiting imminent manifestation.

3 METHODOLOGY

3.1 PRE-TRAINING AND DATASET

We have adopted a transfer learning approach by fine-tuning a pre-trained M2M100 model (*Facebook/m2m100_418M*), harnessing its rich linguistic knowledge to improve Po Tangle's translation performance.

We translated 711 sentences from the Flores200 Dev Benchmark dataset. This dataset has 979 original sentences. Consequently, we added 62 sentences from the FloresDevTest dataset and 377 sentences from the Daily Trust Dev datasets to make 1150 translated sentences from English into Po Tangle, which is divided into 80% of the training set and 10% each for the Dev and Test sets.

3.2 MODEL TRAINING

We utilize the Hugging Face Transformers library for its enhanced capabilities and seamless model integration with state-of-the-art Seq2Seq modeling.

The training process has been carefully designed to optimize model performance, using the following key steps:

- **Hyper-parameter tuning:** We perform experiments with different hyper-parameters like the number of epochs, batch sizes, and learning rate, within the limits of the data to identify an appropriate configuration for the Po Tangle translation.
- **Loss function:** An appropriate loss function, the cross-entropy with label smoothing, is selected to guide model learning to produce translations.
- **Regularization:** Considering the size of the dataset, we implemented a dropout regularization technique to avoid over-fitting and improve model generalization.
- **Evaluation and refinement:** Realizing the limitations of using a small dataset for evaluation, we monitor the performance of the model throughout training using established metrics such as the BLEU score(bilingual evaluation understudy), an algorithm for the evaluation of the quality of text that has been machine-translated from one natural language to another, and human evaluation to assess translation quality, fluency, and cultural appropriateness.

Our methodology was to create a functional Po Tangle-English MT system by fine-tuning a pre-trained M2M100 model with a limited dataset. Despite the amount of data, our system has shown promising results. These indicate its ability to produce translations. Therefore, the fine-tuning of a pre-trained model proves effective for the translation of Po Tangle, even with a modest dataset of 1150 sentences. It can be scaled. Using larger datasets, such as 5,000-15,000 sentences of corpora, will greatly improve its accuracy.

4 RESULTS AND DISCUSSION

The model was trained using distributed training with a single GPU. The training process took approximately 16 minutes and 59 seconds for 10 epochs. The training loss decreased from 4.3968 to 0.6254 during the training process. The evaluation metrics after 10 epochs are as follows:

- **Evaluation BLEU Score**: 6.7604
- **Evaluation Generation Length**: 51.8087
- **Evaluation Loss**: 4.6902
- **Evaluation Runtime**: 1 minute and 15.04 seconds
- **Evaluation Samples**: 115
- **Evaluation Samples per Second**: 1.532
- **Evaluation Steps per Second**: 0.386

The model also underwent prediction with the following metrics:

- **Prediction BLEU Score**: 6.0101
- **Prediction Generation Length**: 51.0565
- **Prediction Loss**: 4.7735
- **Prediction Runtime**: 2 minutes and 2.82 seconds
- **Prediction Samples**: 230
- **Prediction Samples per Second**: 1.873
- **Prediction Steps per Second**: 0.472

These results indicate a reasonable performance of the model, especially in terms of the BLEU score.

BLEU is a common metric for evaluating the quality of machine translation models. Therefore, our Our English-to-Po Tangle machine translation (MT) system achieved very promising results with only 1150 parallel sentence corpora, demonstrating its ability to overcome language barriers and promote linguistic inclusivity.

4.1 DISCUSSION OF RESULTS

The results of our study demonstrate the feasibility and the potential of the fine-tuning of a pre-trained M2M100 model for the machine translation of Po Tangle-English with a limited amount of data. The model achieved promising results despite the size of the dataset. This is demonstrated by the evaluation BLEU score of 6.7604 and the predictive BLEU score of 6.0101. These scores indicate that, especially with more data to learn from, the model can produce translations that may be reasonably close to human translations. This is particularly true given the complexity and linguistic diversity of Tangle.

Several factors may explain the relatively low BLEU scores. Firstly, the ability of the model to fully learn the nuances of the Tangle language and its translation into English was limited by the small size of the dataset. Secondly, improving the quality of the data by following the orthography and diacritising all the relevant alphabets appropriately, thirdly, the M2M100 model, although versatile, may not be specifically optimized for the Tangle language, which could affect its performance compared to models trained on larger, more diverse datasets.

Our study is an important step towards developing a robust machine translation system for Po Tangle. By fine-tuning the pre-trained models and expanding our dataset, we aim to create a system that can provide accurate and culturally sensitive translations, which will ultimately be of benefit to both the Tangle-speaking community and the thousands of other neglected or underserved languages that are seriously endangered.

4.2 KEY FINDINGS

Performance Measures: The model developed achieved a BLEU score of 6.7604 during evaluation, indicating its accuracy, and a predicted BLEU score of 6.0101 on unseen data, highlighting its reliability.

Translation quality: Despite the limited data set, our translations tend to be fluent and clear, demonstrating the soundness and effectiveness of our approach.

Cultural sensitivity: Our machine translation system prioritizes the preservation of cultural nuances. This facilitates a deeper understanding and appreciation of languages.

In the future, we plan to expand our training data to 5,000 or even up to 15,000 parallel sentences. We will also build the necessary parallel corpora and apply domain adaptation techniques, most especially with the right funding. We will also involve local stakeholders. This will ensure that our system meets the unique needs of the Tangle community, to foster a harmonious organized effort from all the required channels, working together to achieve our collective definite chief aim.

Social impact: Our work is about empowering Po Tangle speakers, enabling them to participate in the global digital economy, and preserving their linguistic heritage for future generations, alongside the technological digitization and advancements of similar languages across the African continent.

Our research embodies inclusivity, equality, and the vision of a linguistically diverse Africa. One that fosters mutual understanding and collective progress. We are committed to empowering marginalized communities, championing the preservation of neglected languages, and creating a future where every voice is heard, regardless of the language.

5 CONCLUSION

Our research successfully demonstrates the feasibility of fine-tuning a pre-trained M2M100 model for Po Tangle-English machine translation, even with a limited dataset. Despite this constraint, the model achieved promising results, as reflected in the evaluation BLEU score of 6.7604 and the prediction BLEU score of 6.0101. These scores indicate the model's ability to generate translations.

Our findings suggest that significant performance gains can be achieved by expanding the dataset to between 5,000 and 15,000 parallel sentences. Future research will explore the suitability of other pre-trained models for the Tangle language, applying in-context learning with Generative AI, and also investigate the inclusion of domain adaptation techniques to further enhance translation accuracy.

Ultimately, this research aims to empower the entire Tangle-speaking community by providing access to accurate and culturally sensitive machine translation tools. It also demonstrates the feasibility of digitizing any African language, regardless of the number of speakers. This work is an important step towards fostering intercultural understanding and promoting linguistic diversity, contributing to a more inclusive and connected global society.

6 ACKNOWLEDGEMENT

Special commendation goes to Data Science Nigeria (DSN) for their impactful efforts to nurture 1 million AI talents in Nigeria, directly or indirectly impacting Tangale communities by 2020. DSN's successful execution of the annual AI invasion in both Billiri and Kaltungo underscores their commitment, as well as their invaluable support to individuals currently honing their AI skills. We thank our lead mentor and convener, Dr Olubayo Adekanmi, for his pivotal role in making it all possible. We also thank Mrs Toyin Adekanmi for her innovative support in the area of staff welfare, which was crucial in overcoming significant challenges. We would also like to thank the DSN Research and Innovation team, led by Mr Anthony Soronnadi, whose inspiration and guidance were instrumental in getting this paper accepted for the AfricaNLP workshop at the ICLR conference.

We would also like to express our deepest gratitude to Dr David I. Adelani for his invaluable guidance and unwavering support throughout this project. Dr. Adelani's expertise, mentorship, and the original code from his research paper "A few thousand translations go a long way!" have contributed to the success of our research from its inception to its completion, his dedication to advancing the field of machine translation and commitment to training and mentoring the next generation of researchers has been an inspiration to us, and his insights and contributions have greatly enriched our work and helped achieve our results.

We extend our deepest gratitude to the esteemed translators whose dedication and expertise have been essential to the success of this research. Through their meticulous work, they have not only translated words and sentences but also preserved the essence and cultural richness of the Tangle language. Their names will be forever inscribed in the annals of this project as a testament to their commitment to linguistic diversity and cross-cultural understanding. Their contributions have not only facilitated the advancement of machine translation but have also empowered the Tangle-speaking community by providing access to information and ideas in their mother tongue. They are listed below:

(a) Prof. Nebath Tanglang
(b) Dr. Malata A. Zakayo
(c) Mr. Sokomi Sariel Ankruma
(d) Mr. Ankale Yelyel Tiling
(e) Mr. Jibrin Dawa Butak
(f) Rev. Barnabas Kano
(g) Mr. Yusuf Ahmed Danhausa
(h) Mr. Obed Tarkie K.
(i) Mrs. Briska Keftin Amuga
(j) Mrs. Nita Priscilla Bello
(k) Revd Inspector Ankayo Iliyasu John
(l) Mr. Danladi London
(m) Mr. Emmanuel Umaru
(n) Mr. Mallum Amos
(o) Mr. Kela Andrew Mela
(p) Mr. Stephen Laiko
(q) Mr. Naaman Lamela
(r) Mr. Abdullahi Fredrick
(s) Mr. Dulyamba Bagauda Alkeria
(t) Mrs. Esther Durami
(u) Mrs. Briska Bulus
(v) Zaphaniah Buba
(w) Mrs. Betty Elisabeth Bako

(x) Mr. Isa Ezra Yunusa

(y) Mrs. Angela Butack

We honour their invaluable role in this endeavour, and We express our deepest appreciation for their unwavering support and outstanding contributions. Their work is a beacon of excellence and a source of pride for the Tangle community and all those who value the preservation of languages and cultures.

Several people deserve special recognition for their extraordinary contributions to the success of this project. Mrs Asabe Mwalin Bulack's fundamental tools were crucial in laying the groundwork without which this project would not have been possible. Professor Nebath Tanglang's guidance and leadership in several aspects were invaluable. His experience, including collaborating with three experienced elders from the Tangle community who had previously translated the Tangle Bible, namely Mr. Ankale Yelyel Tiling, Mr Jibrin Dawa Butak, and Rev. Barnabas Kano, and the support of the ACETEL institution, contributed greatly to the success of the project. Dr Malata Andrew Zakayo's role as Chief Linguist was crucial. Her guidance, linguistic reviews, and experience from co-authoring the Po Tangle - English Dictionary and the Po Tangle Orthography were invaluable. Mr Sokomi Sariel's contributions have been immense, particularly in providing high-quality translations and valuable lessons for the whole team. His ability to attract and utilize a high-quality team from the outset of the project was also remarkable. Dr Amina Sambo's tireless efforts were crucial to the overall success of the project, her invaluable contributions cannot be quantified and are deeply appreciated. Not forgetting Mr. Yusuf Ahmed Danhausa who has always been there with the Linguistic professionalism that contributed in shaping the work and Mr. Emmanuel Okose who played a crucial role on a daily basis through encouragement, expert advice and directions among other things.

Lastly, we would like to thank the members of the Tangle community who generously gave their time, expertise, and support to this project. Their commitment and dedication are invaluable at every stage of this research endeavour. We are particularly grateful to those who contributed before the project, during the project, and especially to those who are dedicated to expanding the dataset from 1150 to between 5000 and 15000 ideal parallel sentences to provide a sufficient corpus for full machine translation and datasets for other natural language processing tasks. This is because MT thrives on the active participation of the community, including the linguists of Tangle origin, who play a crucial role in preserving and enriching the Tangle language.

While the list is extensive, we wish to express our gratitude to all individuals, including but not limited to: Asabe Mwalin Bulack, Pheetami Alexandra, Bilyminu Babadidi, Rambi Ayala, Danlami Arabs Rukuijei, Stephen Garba Chikasoro, Yusufu Maimuruchi, Abner Ibrahim Longi, Shedrack Napoleon Abimaje, David Mele Ankama, Esther Wada, Musa Dawa, Gaius George, Bedan Kwaras, Rita Begel, Ann George, Sambo Raymond, Joel James Radakson, Gabrial George, Mahmood Mustapha, Jerry Charles, Mohammed Mustapha, Maikwada Maikenti, Mark Dawa, Iliya Charles Adamu, John Jack, Ibrahim Babaji Babadidi, Liatu Caleb Kapau, Emmanuel Ehud, Isaiah Yusuf, Ezekiel Bako, Yusuf Babadidi, Ayuba Maikaho, Kabati Ankale, Zainab Sabo Mustapha, Andrew Balarabe Barnabas, Emmanuel Maisamari, Ibrahim Pane, Alfred John, Lilian Moses Ngbale, Theresa Tibangs Sani, Suleiman Silas, Peter Tibangs, Daniel Yakubu Molmela, Timza Dantata, Benjamin Dangoma, Isaac Abraham, Mwalin Abdu Buba, Emmanuel Ibrahim, Naomi Ibrahim, Edna Danladi Natti, Musa Ankama, Tafida Sabulu, Paul Yarida, Mela Richard Wada, Molta Ibrahim Lakajang, Molmela John Panguru, Musa Kwangs, Ruben Nahaya, Tinah Shehu, Bore Baras, Hassan Hussaini, Joseph Miller, Alice Stephen Bulack, Mikah Ruben, Mela Ankale, Mela Maddo, Ibrahim Dauda Karau, Abubaker Jimeta, Samuel Pane, Samuel Silas, Yakong Madugu, Kabrang Nayako, Timodok Ankale, AbdulRasheed Abdulaziz, Khamis Adamu, Sadiq Ibrahim Mailafiya, Lydia Maddo, Ishaku David Dantata, Serah Stephen, Umbi Henry Wada, Umar Ibrahim Dikko, Vicky Solomon, Yilla Maddo, Shuaibu Jimeta, Wada D Wada, Samuel Peter, Lydia Kure, Gabrial Mabudi, Caleb Abigail Kapau, Ankale Manzo Memucan, Yohanna Angela Butack, Luka Elkana, Solomon Simon, Joram Jonathan, Ezra Madaki, Yila John Tadi, Yakubu James, Samaila Stephen, Dora Nebath, Susanna Stephen, Joshua Ankama, and the 300 members from the whatsApp group.

In one way or another, and different in capacities, each of these individuals has played or is playing a vital role in advancing our understanding and appreciation of the Tangle language and its

contributions. We are very grateful for their support and proud to acknowledge their invaluable contributions.

Special thanks to Google Africa for awarding the conference travel grant to participate in the ICLR conference and the AfricaNLP workshop 2024. This scholarship not only makes it possible to attend this prestigious event but also reflects Google's commitment to supporting the development of AI research in Africa. This opportunity is truly honoured and appreciated.

Thank you

REFERENCES

David Ifeoluwa Adelani, Jesujoba Oluwadara Alabi, Angela Fan, Julia Kreutzer, Xiaoyu Shen, Machel Reid, Dana Ruiter, Dietrich Klakow, Peter Nabende, Ernie Chang, et al. A few thousand translations go a long way! leveraging pre-trained models for african news translation. *arXiv preprint arXiv:2205.02022*, 2022.

Benjamin Akera, Jonathan Mukiibi, Lydia Sanyu Naggayi, Claire Babirye, Isaac Owomugisha, Solomon Nsumba, Joyce Nakatumba-Nabende, Engineer Bainomugisha, Ernest Mwebaze, and John Quinn. Machine translation for african languages: Community creation of datasets and models in uganda. 2022.

Moussa Koulako Bala Doumbouya, Baba Mamadi Diané, Solo Farabado Cissé, Djibrila Diané, Abdoulaye Sow, Séré Moussa Doumbouya, Daouda Bangoura, Fodé Moriba Bayo, Ibrahima Sory Condé, Kalo Mory Diané, et al. Machine translation for nko: Tools, corpora and baseline results. *arXiv preprint arXiv:2310.15612*, 2023.

Ethnologue. Languages of the world. `https://www.ethnologue.com/`, a. Accessed on April 8, 2024.

Ethnologue. Tangale: A language of nigeria. `https://www.ethnologue.com/language/tan`, b. Accessed on April 8, 2024.

Ignatius Ezeani, Paul Rayson, Ikechukwu Onyenwe, Chinedu Uchechukwu, and Mark Hepple. Igbo-english machine translation: An evaluation benchmark. *arXiv preprint arXiv:2004.00648*, 2020.

Wilhelmina Nekoto, Vukosi Marivate, Tshinondiwa Matsila, Timi Fasubaa, Tajudeen Kolawole, Taiwo Fagbohungbe, Solomon Oluwole Akinola, Shamsuddeen Hassan Muhammad, Salomon Kabongo, Salomey Osei, et al. Participatory research for low-resourced machine translation: A case study in african languages. *arXiv preprint arXiv:2010.02353*, 2020.

Wilhelmina Nekoto, Julia Kreutzer, Jenalea Rajab, Millicent Ochieng, and Jade Abbott. Participatory translations of oshiwambo: Towards sustainable culture preservation with language technology. In *3rd Workshop on African Natural Language Processing*, 2022.

Ebelechukwu Nwafor and Anietie Andy. A survey of machine translation tasks on nigerian languages. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pp. 6480–6486, 2022.

Kelechi Ogueji and Orevaoghene Ahia. Pidginunmt: Unsupervised neural machine translation from west african pidgin to english. *arXiv preprint arXiv:1912.03444*, 2019.

Iroro Orife. Towards neural machine translation for edoid languages. *arXiv preprint arXiv:2003.10704*, 2020.

Iroro Orife, Julia Kreutzer, Blessing Sibanda, Daniel Whitenack, Kathleen Siminyu, Laura Martinus, Jamiil Toure Ali, Jade Abbott, Vukosi Marivate, Salomon Kabongo, et al. Masakhane–machine translation for africa. *arXiv preprint arXiv:2003.11529*, 2020.

Machel Reid, Junjie Hu, Graham Neubig, and Yutaka Matsuo. Afromt: Pretraining strategies and reproducible benchmarks for translation of 8 african languages. *arXiv preprint arXiv:2109.04715*, 2021.

Nerus Yerima Tadi and Malata Andrew Zakayo. Po tangle - english dictionary and vocabulary. *Po-Tangle Committee*, 2008a.

Nerus Yerima Tadi and Malata Andrew Zakayo. Po tangle orthography. *Nigerian Educational Research and Development Council*, 2008b.

Allahsera Auguste Tapo, Bakary Coulibaly, Sébastien Diarra, Christopher Homan, Julia Kreutzer, Sarah Luger, Arthur Nagashima, Marcos Zampieri, and Michael Leventhal. Neural machine translation for extremely low-resource african languages: A case study on bambara. *arXiv preprint arXiv:2011.05284*, 2020.

Pavanpankaj Vegi, J Sivabhavani, Biswajit Paul, KR Prasanna, and Chitra Viswanathan. Anvita-african: A multilingual neural machine translation system for african languages. In *Proceedings of the Seventh Conference on Machine Translation (WMT)*, pp. 1090–1097, 2022.