International Journal of Academic Multidisciplinary Research (IJAMR)
ISSN: 2643-9670
Vol. 10 Issue 10, October - 2026, Pages: 563-566

Using Natural Language Processing And Machine Translation Technologies In Preserving African Indigenous Languages And Oral Traditions In Kwara State

Olubukola Christianah Dada PhD1, Ganiyu Akanbi Yusuf PhD2

1Department of Special Education, Faculty of Education, Kwara State University, Malete
Phone Number: +2348034468438
Email Address: olubukola.dada@kwasu.edu.ng
ORCID: 0000-0001-9669-637X
2Centre for Language immersion and translation
Phone number: +2349053644928
Email Address: ganiyu.yusuf19@kwasu.edu.ng ORCID: 0000-0002-9580-7477

Abstract: *African indigenous languages and oral traditions are rapidly declining due to globalisation, urbanisation, and inadequate documentation systems. To address this, a researcher examined the role of Natural Language Processing (NLP) and Machine Translation (MT) technologies in preserving African indigenous languages and oral traditions, Kwara State. A descriptive survey design was employed, using simple random sampling technique to select 150 respondents from three senatorial districts. Technology and Indigenous Language Preservation (TILP), and Perceived Improvement in Indigenous Language Preservation(PIILP) were used. PPMC was used to obtain reliability index at 0.87, and 0.79. Descriptive statistics of frequency counts, means, and percentages, were used to answer research questions, while Pearson Product-Moment Correlation and Analysis of Variance (ANOVA) were used to test the hypotheses at 0.05 significance level. It was revealed among others that there was a significant positive correlation between the use of these technologies and perceived improvement in language preservation ($r = 0.59$, $p < 0.05$), and There was significant difference in usage based on profession ($F(2,147) = 4.36$, $p < 0.05$. It was concluded that NLP and MT have potential for language preservation, and recommended there should be developed culturally inclusive NLP/MT models through collaboration with native speakers and local scholars*

Keywords: Natural Language Processing, Machine Translation, Indigenous Language Preservation, Oral Traditions

Introduction

Language is not merely a tool for communication; it also serves as a vital repository of culture, tradition, and identity. In Africa, indigenous languages are rapidly declining under the pressures of globalisation, urbanisation, and the widespread prioritisation of dominant global languages such as English, French, and Arabic. Although the continent is home to over 2,000 distinct languages, many of these are either endangered or already extinct. This linguistic erosion is largely driven by generational disuse, inadequate documentation, and a persistent lack of institutional support for mother tongue

However, many of these languages are endangered, primarily due to lack of formal documentation and limited usage in digital and academic spaces. Technological innovations, particularly Natural Language Processing (NLP) and Machine Translation (MT), offer novel solutions to this challenge. NLP encompasses computational methods for processing and analysing large amounts of natural language data, while MT refers to the automatic translation of text or speech from one language to another. These technologies have been successfully applied in language preservation efforts globally (Bird, 2020; Joshi et al., 2020).

The digital era offers a potential paradigm shift in the preservation and revitalisation of indigenous languages through technological interventions. Natural Language Processing (NLP) and Machine Translation (MT) have gained recognition as effective tools for language processing, documentation, and real-time translation (Joshi et al., 2020). NLP can be applied to speech recognition, sentiment analysis, text generation, and semantic search, while MT allows for multilingual translation and content dissemination (Bird, 2020). These technologies can digitise oral traditions, provide educational materials in indigenous languages, and promote inclusive communication.

Despite these potentials, critics argued that the focus of mainstream NLP and MT tools remains on dominant global languages, often excluding under-resourced languages like Yoruba, Nupe, or Baruba (Duong et al., 2021). The limited digital corpora for African languages, and lack of input from native speakers, compromised the quality and inclusivity of these technologies. Nonetheless, projects such as Masakhane and Africa NLP are reversing these trends by collaborating with native speakers and researchers to create open-source NLP models for African languages (Orife et al., 2020). Thus, while limitations persist, there is growing evidence of NLP and MT's potential in preserving African linguistic heritage.

Some scholars were also of the opinion that NLP and MT systems are trained on high-resource languages, creating biases and underperformance for low-resource languages typical of Africa (Blasi et al., 2022). Furthermore, lack of digital infrastructure, insufficient linguistic data, and minimal stakeholder collaboration hindered the full integration of these technologies for African indigenous language preservation (Nekoto et al., 2020). Thus, empirical investigations are necessary to assess the practical impact and limitations of these technologies in specific cultural and geographic contexts, such as Kwara State, Nigeria.

Statement of the Problem

African indigenous languages and oral traditions are rapidly declining in relevance and usage due to modern socio-cultural transformations and persistent technological gaps. Despite growing academic interest and policy-level interventions, preservation efforts remain largely inadequate. A critical challenge lies in the absence of scalable, inclusive, and accessible documentation mechanisms, which exacerbates the threat of linguistic and cultural extinction. Although Nigeria is home to over 500 indigenous languages, the vast majority lack adequate digital representation and are at risk of disappearing. In localised contexts especially in Kwara State, the diminishing use of languages like Yoruba, Nupe, and Baruba particularly among younger generations in urban and semi-urban areas reflects this broader national trend. Oral traditions, which have historically served as key vehicles for transmitting language, history, and culture, are seldom recorded, or digitised, further accelerating cultural erosion. While modern technologies such as Natural Language Processing (NLP) and Machine Translation (MT) offer promising tools for language preservation, but according to the literatures, there is a notable inadequate of empirical evidence assessing their effectiveness in localised Nigerian contexts. Therefore, this study investigated using natural language processing and machine translation technologies in preserving African indigenous languages and oral traditions in kwara state

Research Question

1. What is the extent of use of NLP and MT technologies in preserving indigenous languages and oral traditions in Kwara State?
2. What is the perceived improvement in indigenous language preservation

Methodology

This study employed a descriptive survey research design to assess the impact of NLP and MT on indigenous language preservation. The population comprised language practitioners, educators, and local speakers across Kwara State's three senatorial districts. A simple random sampling technique was used to select 150 respondents. Two instruments such as, Technology and Indigenous Language Preservation (TILP) and Perceived Improvement in Indigenous Language Preservation(PIILP) were developed to collect data. The instrument was validated by some selected educators. To ascertain the reliability of the instruments, 25copies of the final draft were trial tested twice on the randomly selected staff outside the sample population. The instruments' reliability were confirmed using PPMC, with index of 0.87 , and 0.79, indicating high internal consistency. Data collected were analysed using descriptive statistics (frequencies, means, percentages) for research questions. Inferential statistics including Pearson Product-Moment Correlation and Analysis of Variance (ANOVA) were used to test for the hypotheses at a 0.05 level of significance.

Results

Research Question One: What is the extent of use of NLP and MT technologies in preserving indigenous languages and oral traditions in Kwara State?

Table 2: Showing the extent of use of NLP and MT technologies in preserving indigenous languages and oral traditions in Kwara State

S/N	Item	VH	H	M	L	
1	To what extent can technology help preserve indigenous languages for future generations?	64(42.7)	47(31.3)	18(12.0)	21(14.0)	3.03
2	How effective are digital platforms (e.g., mobile apps, websites) as tools for teaching indigenous languages	50(33.3)	45(30.0)	30(20.0)	25(16.7)	2.80
3	To what extent should indigenous languages be included in technology-based learning resources?	56(37.3)	40(26.7)	35(23.3)	19(12.7)	2.89
4	To what extent can social media be used to promote the use of indigenous languages?	15(10.0)	26(17.3)	49(32.7)	60(40.0)	1.97
5	To what extent are sufficient digital tools available for learning your indigenous language?	39(26.0)	48(32.0)	42(28.0)	21(14.0)	2.70

6	To what extent does lack of internet access limit the use of technology for indigenous language preservation?	4(2.7)	18(12.0)	43(28.7)	85(56.7)	1.61
7	To what extent are mobile phones helpful in accessing indigenous language materials?	26(17.3)	13(8.7)	46(30.7)	65(43.3)	2.00
8	To what extent do audio-visual resources in indigenous languages promote better understanding and retention?	27(18.0)	28(18.7)	40(26.7)	55(36.7)	2.14
9	To what extent is community involvement necessary for successful digital preservation of indigenous languages	23(15.3)	20(13.3)	49(32.7)	58(38.7)	2.05
10	To what extent will integrating indigenous languages into AI and voice assistants support their survival?	37(24.7)	21(14.0)	39(26.0)	53(35.3)	2.29
	Weighted Mean					**2.34**

Decision rule: L=1.00-1.49,M=1.50-2.49 H= 2.50-3.49, VH=3.50-4.00 Note: The figures in parentheses are in percentages

Table 2 shows the responses of the respondents on the extent of use of NLP and MT technologies in preserving indigenous languages and oral traditions in Kwara State. It was indicated that the extent of the following items was moderate: To what extent can social media be used to promote the use of indigenous languages?(mean=1.97), To what extent does lack of internet access limit the use of technology for indigenous language preservation? (mean=1.60), To what extent are mobile phones helpful in accessing indigenous language materials?(mean=2.00), To what extent do audio-visual resources in indigenous languages promote better understanding and retention? (mean=2.14), To what extent is community involvement necessary for successful digital preservation of indigenous languages(mean=2.05), and To what extent will integrating indigenous languages into AI and voice assistants support their survival? (mean=2.29). While the following items was high: To what extent can technology help preserve indigenous languages for future generations?(mean=3.30), How effective are digital platforms (e.g., mobile apps, websites) as tools for teaching indigenous languages(mean=2.80), To what extent should indigenous languages be included in technology-based learning resources?(mean=2.89), and To what extent are sufficient digital tools available for learning your indigenous language?(mean=2.70). In the light of the result, it was revealed that the extent of use of NLP and MT technologies in preserving indigenous languages and oral traditions was moderate. The weighted mean (**2.34**) which is a numeric indicator that the extent of use of NLP and MT technologies in preserving indigenous languages and oral traditions in Kwara state.

Research Question Two: What is the perceived improvement in indigenous language preservation

Table 2: Showing the perceived improvement in indigenous language preservation Kwara state

S/N	PIILP	Frequency	Percentage (%)
1	Foundational and Community-Based Efforts	62	**44.2**
2	Integration and Educational Support	28	16.7
3	Technology and Media Utilisation	24	13.3
4	Institutional and Public Visibility	21	11.7
5	Global and Economic Impact	15	8.3
	TOTAL	**150**	**100**

Table 2 shows the perceived improvement in indigenous language preservation. It was indicated that foundational and community-based efforts is most perceived improvement in indigenous language preservation Kwara state

Summary of Inferential Statistics

H_{01}: There was significant positive correlation between the use of NLP/MT and perceived improvement in preservation ($r = 0.59$, $p < 0.05$).

H_{02}: There was significant differences in usage based on profession ($F(2,147) = 4.36$, $p < 0.05$), with educators using the technologies more than traders or artisans.

H_{03}: There was significant differences were found based on education level ($F(2,147) = 3.97$, $p < 0.05$), with higher-educated respondents showing greater engagement.

Discussion of Findings

The findings revealed that technological innovations, particularly NLP and MT, are being increasingly utilised in indigenous language documentation in Kwara State. The significant correlation between usage and preservation suggested that these tools can

complement traditional language preservation methods. This was supported Bird (2020) and Orife et al. (2020), who emphasised the critical role of community-driven language technologies. Bird (2020) and Joshi et al. (2020) also found the significant correlation suggested that increased use of these technologies is associated with enhanced language documentation and accessibility. Besides, digital storytelling platforms, mobile applications, and voice recognition systems have facilitated the preservation of oral forms previously confined to elder custodians. The observed variance across professions and education levels reflected digital literacy gaps and uneven access to technological tools. This aligned with the Nekoto et al. (2020), who emphasised on infrastructural and capacity-building deficits as major constraints to technological adoption in sub-Saharan Africa. Educated individuals are also likely to engage with these tools, indicating a need for inclusive training programmes. The findings also reflected the limitations identified by Duong et al. (2021) regarding digital exclusion and corpus scarcity. Nonetheless, the study also highlighted the need for more culturally adaptive NLP/MT models, as current systems inadequately process African language structures and idioms due to limited data and Eurocentric training corpora (Blasi et al., 2022).

Conclusion

Based on the results from this study, it was concluded that NLP and MT technologies have significant potential in preserving African indigenous languages and oral traditions. Their effectiveness is influenced by users' profession and education level, underlining the importance of digital accessibility and cultural competence in technology deployment.

Recommendations:

Based on the conclusion of this study, it was recommended that should be:

1. Increase funding for digital infrastructure to support local language technologies.
2. Develop culturally inclusive NLP/MT models through collaboration with native speakers and local scholars.
3. Implement digital literacy programmes focused on indigenous language practitioners and rural communities.
4. Encourage policy frameworks that mandate the use of indigenous languages in digital and educational platforms.

References

Adegbija, E. (2004). Language Attitudes in Sub-Saharan Africa: A Sociolinguistic Overview. Multilingual Matters.

Bamgbose, A. (2011). African Languages Today: The Challenge of and Prospects for Empowerment under Globalization. In Qorro, M. (Ed.), Language and Development in Africa (pp. 11–25). Mkuki na Nyota.

Bird, S. (2020). Decolonising Speech and Language Technology. Proceedings of the 28th International Conference on Computational Linguistics (COLING 2020), 3504–3519.

Bamgbose, A. (2011). *African Languages Today: The Challenge of and Prospects for Empowerment under Globalization*. In Brenzinger, M. & Feest, C. (Eds.), *Language Death and Language Maintenance: Theoretical, Practical and Descriptive Approaches* (pp. 25–44). Routledge.

Blasi, D. E., Anastasopoulos, A., Neubig, G., & Yarowsky, D. (2022). Systematic Inequality and Hierarchy in Language Technology. *Nature Machine Intelligence*, 4(8), 658–666.

Duong, C., Costa-jussà, M. R., & Le, N. (2021). An Empirical Study of Low-Resource Translation: Data, Models and Evaluation. ACL Anthology.

Eberhard, D. M., Simons, G. F., & Fennig, C. D. (Eds.). (2022). *Ethnologue: Languages of the World* (25th ed.). SIL International.

Joshi, P., Santy, S., Budhiraja, A., Bali, K., & Choudhury, M. (2020). The State and Fate of Linguistic Diversity and Inclusion in the NLP World. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 6282–6293.

Nekoto, W., Marivate, V., Matsila, T., et al. (2020). Participatory Research for Low-Resourced Machine Translation: A Case Study in African Languages. *Findings of the Association for Computational Linguistics: EMNLP 2020*, 2144–2160.

Orife, I., Kreutzer, J., & Martinus, L. (2020). Masakhane – Machine Translation For Africa. ArXiv preprint, arXiv:2003.11529.

UNESCO. (2021). *Decade of Indigenous Languages (2022–2032)*. Retrieved from https://en.unesco.org/idil2022-2032