IT Journal Research and Development (ITJRD)
Vol.10, No.2, March 2026, E-ISSN : 2528-4053 | P-ISSN : 2528-4061
DOI : 10.25299/itjrd.2025.21277

A Systematic Review on Machine Translation for Low Resource Nigerian Languages

Tijani M. Abdulmusawir [1], A. F. Donfack Kana [2], Amina H. Abubakar [3]
Department of Computer Science, School of Technology, Federal Polytechnic Idah, Kogi state, Nigeria[1]
Department of Computer Science, Faculty of Physical Sciences, Ahmadu Bello University, Zaria, Kaduna state, Nigeria[2,3]
musaritijani@gmail.com[1], donfackkana@gmail.com[2], aminatgusau@yahoo.com[3]

Article Info

Article history:

Received Feb 03, 2025
Revised Jul 21, 2025
Accepted Nov 25, 2025

Keyword:

Artificial Intelligence
Domain Adaptation
Machine Translation
Low-Resource

ABSTRACT

Nigeria ranks among Africa's most linguistically diverse countries with over 500 indigenous languages, yet machine translation (MT) research remains severely limited for these low-resource languages. This systematic review examines the current state of MT research for Nigerian languages, identifies persistent challenges, and analyzes methodological trends. A systematic literature search was conducted across eleven databases including PubMed, Web of Science, and Scopus from January 2010 to August 2025. Search terms combined machine translation approaches with Nigerian language terms. Studies were screened using PRISMA guidelines requiring original research with evaluation metrics. From 51 papers, 25 duplicates were removed, 7 excluded for selection criteria, and 3 for lack of contribution, resulting in 16 studies. Only 11 Nigerian languages (2.2% of over 500 languages) were covered, creating a 97.8% research gap. Yoruba led with 4 studies, followed by Igala (3), Igbo and Nigerian Pidgin (2 each). Methods evolved from rule-based (4 studies, 2014 to 2021) through SMT (2 studies, 2016 to 2019) to NMT dominance (10 studies, 2018 to 2025). Idiomatic expression handling was the most persistent challenge (16.7%), followed by complex sentences, data scarcity, and domain specificity (each 9.5%). Nigerian MT research shows severe underrepresentation with persistent challenges in idiomatic expressions and data scarcity across all approaches. Neural method adoption reflects global trends but doesn't address resource constraints. Coordinated national approaches prioritizing parallel corpora creation and institutional partnerships are needed to prevent digital divides and support language preservation.

Corresponding Author:

Tijani M. Abdulmusawir
Department of Computer Science
School of Technology, Federal Polytechnic Idah
Kogi state, Nigeria
Email : musaritijani@gmail.com

1. INTRODUCTION

Machine Translation (MT) represents a branch of computational linguistics that leverages artificial intelligence to translate text automatically from a source language to a target language using computer algorithms [1]. Translation facilitates national and international collaboration, supporting social, economic, philosophical, and political development across linguistic boundaries.

Contemporary MT systems still face significant challenges in accuracy and fluency. Accuracy is defined as the model's ability to maintain the meaning of input text during translation, while fluency represents the natural flow of translation resembling native speaker output [2]. The development of transformer models in 2017 marked the most recent breakthrough in addressing translation quality challenges [3].

High-resource languages have benefited tremendously from MT system evolution. Language pairs such as English-German, English-French, Chinese-English, and English-Arabic have achieved significant success. However, low-resource African languages, particularly Nigerian languages, remain significantly underserved [4,5].

Nigeria ranks among Africa's most linguistically diverse countries with over 200 million people and more than 500 indigenous languages, representing approximately 7% of global languages. The three major ethnic groups are: Hausa: Spoken by over 50 million people in northern regions, serving as a lingua franca across West Africa, Yoruba: Spoken by approximately 45 million people in western Nigeria, Igbo: Spoken by about 44 million people in southeastern Nigeria [5]. The objective of the review is to conduct a systematic review of machine translation research for Nigerian low-resource languages, analyze current approaches, identify challenges, and propose future research directions with particular focus on Neural Machine Translation (NMT).

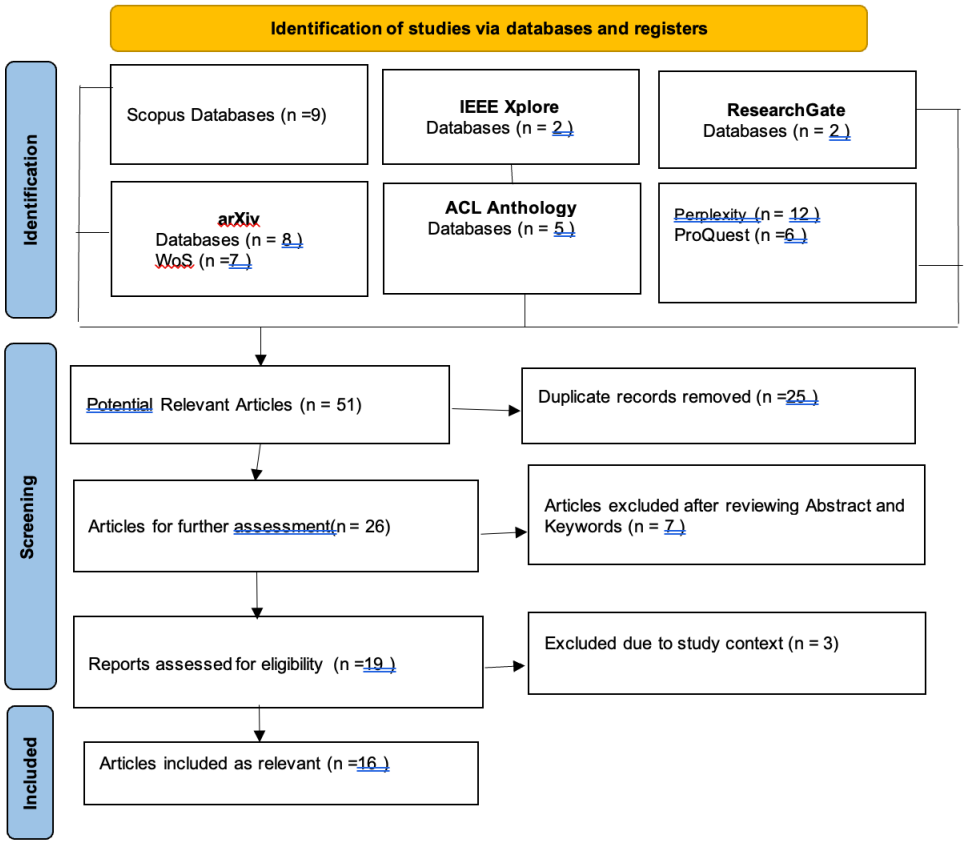

Figure 1. The selection process of the studies included in the present systematic review. Identification: In this step, authors started to search the 9 databases to identify the relevant articles for the present study. Screening: In this step, the authors screened the initially obtained studies by title, abstract, and keywords, focusing on reasons to exclude the studies. Eligibility: In this step, the authors considered the criteria that were employed to select articles for the present study based on the research objectives. Included: In this step, authors finally selected the articles, used for the present study, based on the research objectives.

1.2 Chronological Evolution of MT Approaches

The foundation of modern MT was established by Warren Weaver in 1947 during World War II for code-breaking applications. Georgetown University's 1954 experiment demonstrated machine

translation of 250 words using 6 grammar rules. However, the 1966 ALPAC report led to reduced government support after $20 million investment showed limited progress [6].

Second Generation (Rule-Based Systems): Despite setbacks, rule-based systems emerged focusing on linguistic rules and dictionaries. Notable systems included Météo (1976) for weather forecasting, ATLAS2 (1978) by Fujitsu for Korean-Japanese translation, and SHARP's DUET system (1982). Statistical methods emerged at IBM in the 1980s, with Google Translate adopting SMT in 2005 [7].

Third Generation (Neural Methods): Neural approaches gained momentum in the 2010s. Bahdanau et al. (2014) introduced attention mechanisms addressing encoder-decoder limitations [2]. Vaswani et al. (2017) developed the transformer architecture, establishing current state-of-the-art NMT approaches [3]. Methods column in Table 2 shows categories that Nigeria MT work belongs and Figure 2 displays how MT work on Nigerian languages progresses through this approaches generation.

2. METHOD

2.1 Search Strategy

Eleven databases, including PubMed, Web of Science, Scopus, The Cochrane Library, and ProQuest, were used to obtain the articles for the present review. Keywords such as "machine translation" OR "neural machine translation" OR "statistical machine translation" OR rule-based machine translation) AND ("Nigerian languages" OR "Hausa" OR "Yoruba" OR "Igbo" OR "low-resource languages") were employed to search paper.

2.2 Eligibility Criteria

Studies explored based on the above strategies had to further come across the following eligibility criteria to be incorporated in the present review: must be written in English and published between January 2010 to August 2025; be original research on MT for Nigerian languages, peer-reviewed publications, theses, conference papers, not included reviews without original research, studies without Nigerian language focus, publications without evaluation metrics, abstracts without full papers, were not eligible.

2.3. Included studies

Having use different database search strategies including the use of perplexity search engine, just 51 potential papers were identified. After the initial screening, 25 papers were deleted because of their duplication across databases such Scopus, ProQuest and Web of Science. Out of the remaining 26 papers, 7 were excluded as they did not meet the selection criteria (the title, abstract, and keywords). Of the remaining 19 papers, 3 were eliminated after reviewing study context, design, and literature review, hence a lack of significant contribution. Thus, 16 papers that met the inclusion criteria were selected for the final analysis. Figure 1 reports the extensive selection process, and Table 1 reports all the potential papers, their languages pairs, methods adopted as well as the potential challenges explored by authors, while Table 3 shows Nigerian Languages with MT research coverage.

2.4. Ethical consent

Ethical consent was not called for the present review as no human participants are involved in the present study. The present study reviews and explores the challenges in previously published papers regarding researchers' intention to adopt the automatic machine translation system based on the objectives of the present study.

3. RESULTS

3.1. Major findings

After carrying out systematic analysis of machine translation research on Nigerian languages, a total of 15 distinct limitations were identified, and these 15 limitations appeared 42 times in total across all studies. The comprehensive list of the total number of limitations, their sources, and the frequencies of these limitations appeared in the review are reported in the Table 2. From the frequency analysis, we can find that "Inability to handle idioms" is reported as the major challenge in Nigerian language machine translation, appearing in 16.7% of all incidents (7 out of 42) [1,8,9,10,11,12,13]. "Inability to handle nuances and complex sentences," "Idioms, colloquialisms, and figurative language are difficult to handle," "Absence of named entity recognition," and "Domain specificity and scarcity of data" each emerged in 9.5% of all incidents (4 out of 42) [1, 8,9,10], [14,15,16,17], [13,1,8,9], and [14,15,16,17] respectively, standing 2nd, 3rd, 4th, and 5th. "Lack of flexibility," "Languages have different rules," and "Only focused on specific language/task (limited scope)" each appeared in 7.1% of all occurrences (3 out of 42) [1, 8,9], [1, 8,9], and [18,19,20] respectively, standing 6th, 7th, and 8th. "Language complexity," "Poor long sentence translation," and "Small datasets" each appeared in 4.8% of all incidents (2 out of 42) [11, 13], [12,21], and [13,17] respectively. The remaining five limitations each appeared in 2.4% of all incidents (1 out of 42), including system fluency and contextual accuracy issues [10], reliance on predefined rules affecting contextual subtleties handling [10], data quality concerns from internet sources [21], poor topic classification performance [22], and general system limitations.

3.2 Nigerian Languages with MT Research Coverage

Table 3 shows the analysis of machine translation research coverage across Nigerian languages reveals a total of 22 MT studies distributed among only 11 major Nigerian languages, representing approximately 2.2% of Nigeria's over 500 ethnic languages and dialects. Yoruba demonstrates the highest research attention with five studies, followed by Igbo with four studies, and Hausa with three studies. Nigerian Pidgin, despite its widespread use across all Nigerian zones with 30 million speakers, has received moderate attention with two studies. The remaining seven languages Edo, Esan, Isoko, Urhobo, Nupe, Ebira, and Igala have each been the focus of 1 to 2 studies. This coverage represents a significant research gap, as over 489 Nigerian languages (97.8%) remain without any documented machine translation research. Geographically, the South-West zone (Yoruba) and South-East zone (Igbo) have received the most research attention, while the North-West and North-East zones, despite hosting Hausa with 48 million speakers, have relatively fewer studies. The Middle-Belt region shows mixed coverage with languages ranging from 800,000 to 1.8 million speakers receiving 1 to 2 studies each, and the South-South zone's Edoid languages have received minimal but consistent single-study attention across four different dialects. Figure 2 displays the progression from rule-based approach to NMT approach over the years.

Table 1. Details of Studies Included in This Review

Study	Language Pair	Method	Challenges
Ayegba et al.[1]	English-Igala	Rule-based	Inability to handle idioms, nuances and complex sentences; lack of flexibility; languages have different rules
Akinwale et al.[10]	English-Yoruba	Rule-based	Inability to handle idioms, nuances and complex sentences; lack of flexibility; languages have different rules
Ezeani et al.[13]	Igbo (diacritic restoration)	SMT	Idioms, colloquialisms, and figurative language are difficult to handle; domain specificity and scarcity of data
Nguyen and Chiang[15]	Hausa-English (among 8 language pairs)	NMT	Idioms, colloquialisms, and figurative language are difficult to handle; domain specificity and scarcity of data
Onyenwe et al.[14]	Igbo (POS tagging)	SMT	Idioms, colloquialisms, and figurative language are difficult to handle; domain specificity and scarcity of data

Ahia and Ogueji[16]	English-Nigerian Pidgin	NMT	Idioms, colloquialisms, and figurative language are difficult to handle; domain specificity and scarcity of data
Butryna et al.[29]	Nigerian Pidgin (speech corpus development)	NMT	Solely relied on internet data that may not guarantee quality
Gutkin et al.[31]	Yoruba (speech dataset development)	NMT	Only focused on Yoruba speech and may not be usable for other tasks
Hedderich et al.[17]	Multiple African languages (NER)	NMT	Did not produce good results for topic classification
Orife[18]	Edoid languages-English (Edo, Esan, Urhobo)	NMT	Focused on limited languages of Edoid group only
Eludiora and Ajibade[11]	English-Yoruba	Rule-based	Inability to handle idioms, nuances and complex sentences; lack of flexibility; languages have different rules
Abdulmusawir et al.[12]	English-Ebira	Rule-based	System lacks fluency and contextual accuracy; struggles to convey idiomatic expressions; inability to handle complex sentence structure; reliance on predefined rules makes it unable to handle nuances and contextual subtleties
Adelani et al.[32]	English-Yoruba	NMT	Limited to only Yoruba language and may not be applicable to other languages
Ayegba[19]	English-Igala	NMT	Complexity: inability to translate idiomatic expressions due to language complexity
Umar B.[20]	Nupe-English	NMT	Only works for English-Nupe translation; poor long sentence translation and idiomatic expressions
Emmanuel Makoji; Felix Sani[33]	English-Igala	NMT	Complexity: inability to translate idiomatic expressions due to language complexity; small datasets; absence of named entity recognition

Table 2. Frequency of Challenges

Challenges	Reference Numbers	Frequency (n=42)
Inability to handle idioms/idiomatic expressions	[1], [8], [9], [10], [11], [12], [13]	7
Inability to handle nuances and complex sentences	[1], [8], [9], [10]	4
Idioms, colloquialisms, and figurative language are difficult to handle	[14], [15], [16], [17]	4
Absence of named entity recognition	[13] [1], [8], [9]	4
Domain specificity and scarcity of data	[14], [15], [16], [17]	4
Lack of flexibility	[1], [8], [9]	3
Languages have different rules	[1], [8], [9]	3
Only focused on specific language/task (limited scope)	[18], [19], [20]	3
Language complexity	[11], [13]	2
Poor long sentence translation	[12], [10]	2
Small datasets	[13], [17]	2
System lacks fluency and contextual accuracy	[10]	1
Reliance on predefined rules makes it unable to handle contextual subtleties	[10]	1
Solely relied on internet data that may not guarantee quality	[21]	1
Did not produce good results for topic classification	[22]	1

Table 3. Nigerian Languages with MT Research Coverage

Dialect	Origin	No. of Speakers	Zone	MT Studies
Yoruba	Niger-Congo	39.5M	South-West	4
Igala	Niger-Congo (Yoruboid languages)	800TH	MIDDLE-BELT	3
Igbo	Niger-Congo	27M	South-East	2
Nigerian Pidgin	English Creole	30M	All	2
Hausa	Afro-Asiatic	48M	North-West and North-East	1
Edo	Niger-Congo (Edoid languages)	1.6M	South-South	1
Esan	Niger-Congo (Edoid languages)	300TH	South-South	1
Urhobo	Niger-Congo (Edoid languages)	546TH	South-South	1
Nupe	Niger-Congo (Nupoid languages)	1.5M	MIDDLE-BELT	1
Ebira	Niger-Congo (Nupoid languages)	1.8M	MIDDLE-BELT	1

3.3 Time Period Analysis

Figure 2 demonstrates the clear evolution from early reliance on Rule-Based and Statistical methods to the complete dominance of neural approaches in recent years. From 2014 to 2017 equal distribution with 2 Rule-Based and 2 Statistical studies were recoded; from 2018 to 2021, NMT dominance with 7 studies, plus 2 Rule-Based studies were noticed and lastly, from period of 2022 to 2025 exclusively neural approaches with 3 studies have been recoded so far. The patterns shows how SMT approaches had a brief period of adoption but were quickly overtaken by neural methods, while Rule-Based approaches persisted longer but eventually disappeared from recent research.

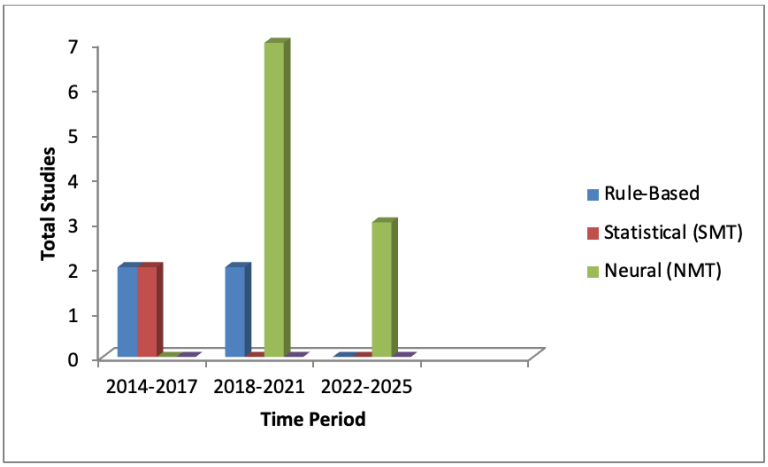

Figure 2. Shows a clear progression shift from rule-based approaches in the early period to NMT methods becoming dominant from 2018 onwards, reflecting the broader trends in machine translation research.

4. DISCUSSION

The present systematic review found that idiomatic expression handling constitutes the most significant barrier to machine translation implementation for Nigerian languages, appearing in 16.7% of all incidents across rule-based, statistical, and neural approaches [1, 8, 9, 10,11,12, 13]. This finding indicates that MT systems consistently struggle with the cultural and linguistic nuances embedded in Nigerian languages, where idiomatic expressions carry deep contextual meanings that cannot be directly translated through literal word-for-word conversion. The persistence of this

challenge across different methodological approaches suggests that current MT architectures lack the sophisticated cultural knowledge representation needed to handle the rich figurative language patterns characteristic of Nigerian linguistic communities. This aligns with findings from other African language MT research, where similar challenges have been reported for Swahili and Amharic translation systems [23, 24].

Data scarcity and domain specificity emerged as equally critical challenges, each affecting 9.5% of all incidents, highlighting the fundamental resource constraints facing Nigerian language MT development [14], [15], [16], [17]. The limited availability of high-quality parallel corpora for Nigerian languages creates a cascading effect that impacts both statistical and neural approaches, as these methods require substantial amounts of training data to achieve acceptable performance levels. This scarcity is particularly pronounced for the over 489 languages (97.8%) that have received no MT research attention, creating a digital divide that threatens the preservation and technological advancement of Nigeria's linguistic heritage. Similar patterns have been observed in other low-resource African contexts, where successful systems like South African multilingual transformers achieved 25 to 35% BLEU scores only after significant corpus development investments [25].

The methodological evolution from rule-based approaches (4 studies, from 2014 to 2021) to neural dominance (10 studies, from 2018 to 2025) reflects broader trends in global MT research, yet the persistence of fundamental challenges suggests that technological advancement alone is insufficient without addressing underlying resource constraints [2,3]. Rule-based systems demonstrated consistent limitations in handling complex sentences and nuances [1, 8, 9, 10], while neural approaches, despite their theoretical advantages, continue to struggle with the same idiomatic expression challenges that plagued earlier methodologies [11, 12, 13]. This pattern contrasts with successful African MT implementations in South Africa and Kenya-Tanzania joint initiatives, where government support and coordinated resource development enabled more substantial progress [23, 25].

The distribution of MT research across Nigerian languages reveals a critical underrepresentation of the country's linguistic diversity, with only 11 languages receiving research attention out of over 500 ethnic languages and dialects, highlighting a massive research gap of 97.8%. This stark disparity underscores the challenges facing low-resource language preservation and digital inclusion in Nigeria's multilingual context. The research concentration on major languages like Yoruba (five studies), Igbo (four studies), and Hausa (three studies) reflects a natural focus on high-population languages but simultaneously neglects hundreds of smaller ethnic communities whose languages may be at risk of digital extinction. Most notably, Hausa, with 48 million speakers representing the largest linguistic community, has received disproportionately limited research attention compared to Yoruba (39.5 million speakers) and Igbo (27 million speakers), suggesting that factors beyond demographic considerations such as research infrastructure, institutional capacity, or academic interest drive research priorities. The inclusion of smaller languages like Igala (over 800,000 speakers), Nupe (over 1.5 million speakers), and various Edoid languages demonstrates some awareness of linguistic diversity, yet this represents merely the tip of the iceberg considering Nigeria's vast linguistic landscape. The minimal attention to Nigerian Pidgin, despite its unique position as a lingua franca, and the complete absence of MT research for the remaining over 489 languages highlight the urgent need for systematic, comprehensive research planning that balances demographic significance with cultural preservation priorities [1, 8, 9, 11,12,13, 14,15,17, 18, 19, 20, 21, 22]. These findings call for innovative approaches to low-resource language MT development, potentially through transfer learning, multilingual models, and community-based research initiatives to bridge the enormous gap between Nigeria's linguistic richness and its digital language technology infrastructure.This pattern may reflect historical research infrastructure, institutional capacity, or funding availability rather than strategic language preservation planning. Successful African models demonstrate that coordinated government support and cross-linguistic transfer learning can overcome resource limitations, as evidenced by Moroccan Darija systems benefiting from Arabic language proximity [22].

Technical limitations such as complex sentence handling, lack of flexibility in rule-based systems, and contextual accuracy issues demonstrate that current MT approaches require significant

architectural improvements to address the morphological complexity and syntactic variations present in Nigerian languages [1, 8, 9, 10]. The prevalence of these challenges across different methodological periods suggests that fundamental advances in multilingual model architectures, transfer learning techniques, and cross-lingual representation learning are necessary to overcome the persistent barriers identified in this review. Additionally, system-level challenges including data quality concerns from internet sources [21], absence of named entity recognition capabilities [13], and limited scope in language coverage [18, 19, 20] indicate that comprehensive solutions require both technological innovation and systematic resource development.

The comparative analysis with other African MT contexts reveals that Nigeria's linguistic diversity presents unique scaling challenges that require innovative approaches beyond traditional bilateral translation models. While individual Nigerian languages show performance patterns similar to other African languages, the sheer number of languages requiring attention (over 500 vs. 11 to 25 in most other African countries) necessitates fundamentally different resource allocation strategies and potentially novel multilingual architectures that can leverage cross-linguistic relationships within Nigeria's language families [22, 23,24, 25]. The findings underscore the need for comprehensive national language technology policies that balance demographic considerations with cultural preservation priorities, potentially through collaborative research initiatives, community based data collection programs, and innovative low-resource language MT approaches that can extend coverage to Nigeria's vast linguistic landscape.

4.1 Way Forward

Addressing the multifaceted challenges in Nigerian language machine translation requires a comprehensive, coordinated approach. Quality parallel corpora creation must be the foundational element for sustainable MT development. Data scarcity affects 9.5% of all documented challenges [14], [15], [16], [17]. This scarcity fundamentally underlies the persistent idiomatic expression handling difficulties that affect 16.7% of all incidents across all methodological approaches [1, 8, 9, 10,11,12, 13]. A systematic corpus development program should establish standardized collection protocols. These protocols must involve native speakers, linguists, and cultural experts. This ensures comprehensive coverage of figurative language, contextual expressions, and domain-specific terminology that current systems consistently fail to handle.

The corpus development strategy must adopt a multi-tiered approach. This approach should address both high-resource and low-resource language needs. Robust quality assurance mechanisms must be implemented to overcome data quality concerns identified in previous efforts [21]. For major languages like Hausa, Yoruba, and Igbo, large-scale professional translation projects should create comprehensive multilingual datasets. These datasets should span news, literature, technical documents, and conversational text. Community based collection programs should be developed for the over 489 languages currently without research coverage. These corpora should specifically target the complex sentence structures and nuances that challenge rule-based systems [1, 8, 9, 10]. They should also provide sufficient diversity to train neural models capable of handling language complexity issues that persist in current NMT approaches [11, 13].

Technical advancement must parallel corpus development through innovative architectural solutions. These solutions should address the flexibility limitations of traditional approaches. They must leverage the growing NMT research momentum evidenced by the 62.5% share of studies using NMT methods. Cross-linguistic transfer learning techniques should be developed to maximize the utility of limited resources across Nigeria's language families. This could address the limited scope challenges [18,19,20] by creating multilingual models. These models can benefit multiple related languages simultaneously. Specialized modules for named entity recognition [13] and contextual subtlety handling [10] should be integrated into MT systems. This will address specific technical gaps identified across different approaches.

Institutional coordination represents a critical success factor. This requires partnerships between Nigerian universities, government agencies, technology companies, and international development organizations. These partnerships must establish sustainable funding mechanisms and technical infrastructure. The research concentration disparity is significant. Only 11 languages receive attention among more than 500 ethnic languages. This necessitates national language

technology policies that mandate systematic coverage. These policies must balance demographic significance with cultural preservation priorities. Learning from successful African models such as South African multilingual transformers [25] and Kenyan-Tanzanian joint initiatives [23] is essential. Nigeria should establish coordinated research centers that can implement standardized methodologies. These centers should share resources and ensure consistent quality across different linguistic communities. This comprehensive approach would transform the current fragmented landscape into an integrated ecosystem. It would address the full spectrum of challenges while establishing a sustainable foundation. This foundation would preserve and advance Nigeria's rich multilingual heritage through modern translation technologies.

5. CONCLUSION

This systematic review reveals significant potential for Nigerian language MT development alongside substantial challenges. While recent neural approaches show promise, success requires coordinated efforts through analysis of 16 studies spanning 2014-2025, covering 11 major Nigerian languages out of over 500 ethnic languages, several critical findings emerge that demand immediate attention from researchers, policymakers, and technology developers. Addressing critical priorities including large-scale parallel corpus development, standardized evaluation frameworks appropriate for tonal languages, community-engaged research methodologies, cross-linguistic transfer learning strategies, and sustainable funding mechanisms.The evolution from rule-based (2014 to 2021) to NMT approaches (2018 to 2025) demonstrates growing research maturity, but the field requires systematic coordination to achieve impact comparable to high-resource language MT systems.

REFERENCES

[1] S. F. Ayegba, O. E. Osuagwu, and N. D. Okechukwu, (2014) "Machine translation of noun phrases from English to igala using the rule-based approach" *West African Journal of Industrial and Academic Research,* vol. 11, pp. 18-28

[2] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate." *In Proceedings of the 3rd International Conference on Learning Representations (ICLR). San Diego, CA.* 2015; *https://arxiv.org/abs/1409.0473*

[3] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need." *In Proceedings of the 31st International Conference on Neural Information Processing Systems,* 2017, *pp. 5998-600, https://doi.org/10.5555/3295222.3295349*

[4] X. Wang, Y. Tsvetkov, G. Neubig "Balancing Training for Multilingual Neural Machine Translation" Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 2020, pp. 8526-8537. DOI: 10.18653/v1/2020.acl-main.754F. Stahlberg, "Neural machine translation: A review". *Journal of Artificial Intelligence Research, vol. 70, pp.1-30. 2020. https://doi.org/10.1613/jair.1.12007*

[5] D. S. Doris, "Languages spoken in Nigeria as of 2021, by number of speakers" *statista*. Available P. Koehn, "Statistical Machine Translation" *Cambridge University Press*, ISBN-13 978-0-511-69132-4. 2010.

[6] *W. Yonghui, M. Schuster, Z. Chen, Q. V. Le, and N. Mohammad, "Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation." In: arXiv:1609.08144v2 [cs.CL], pp 1–23, Oct 2016.* https://www.statista.com/statistics/1285383/population-in-nigeria-by-languages-spoken.2024

[7] F. Stahlberg "Neural machine translation: *A review. Journal of Artificial Intelligence Research, 70, 1-30.*2020. *https://doi.org/10.1613/jair.1.12007*

[8] O. I. Akinwale, A. O. Adetunmbi, O. O. Obe, A. T. Adesuyi. " Web-Based English to Yoruba Machine Translation". International Journal of Language and Linguistics. Vol. 3, No. 3, 2015, pp. 154-159. doi: 10.11648/j.ijll.20150303.17

[9] S. I. Eludiora, and B. A. Ajibade. "Design and Implementation of English To Yorùbá Verb Phrase Machine Translation System." *Koozakar Festschrift*, 2021. Available: https:/api.semanticscholar.org/CorpusID:233204633.

[10] T. M. Abdulmusawir, S. F. Ayegba, Y. M. Kayode, and E. C. Christian, "A system for machine translation from English to Ebira using the rule-based approach*" Journal of Scientific Research and Reports,* 2021; *vol. 27 no. 11, pp. 137-148. https://doi.org/10.9734/jsrr/2021/v27i1130465*

[11] S. F. Ayegba, "Development of English-to-Igala machine translation system using neural machine translation with attention mechanism," International Journal of Research and Development. Vol . 8, no. 1, pp. 1-10, 2023.

[12] B. U. Umar, "Nupe-English Neural Machine Translation Using Sequence to Sequence Model With Attention Mechanism."Natural Language Processing Journal."Available:https://www.researchgate.net/publication/38227542, 2024.

[13] E. Makoji; F. Sani. "Development and Evaluation of an English-to Igala Neural Machine Translation System using Deep Learning." *International Journal of Innovative Science and Research Technology*, 10(5), 914-919. 2025 https://doi.org/10.38124/ijisrt/25may556

[14] I. Ezeani, P. Rayson, I. Onyenwe, C. Uchechukwu, M. Hepple, "Automatic Restoration of Diacritics for Igbo Language". Transactions of the Association for Computational Linguistics. https://eprints.whiterose.ac.uk/id/eprint/117833/1/TSD_2016_Ezeani.pdf

[15] I. E. Onyenwe., M. Hepple., U. Chinedu, & E. Barnard. "Toward an Effective Igbo Part-of-Speech Tagger". ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP). https://aclanthology.org/2019.nsurl-1.18.pdf

[16] T. Nguyen, & D. Chiang, "Improving Rare Word Translation with Dictionaries and Attention Masking." 2018. arXiv preprint. https://arxiv.org/pdf/2408.09075.pdf

[17] O. E. Ahia, and K. E. Ogueji, "Toward Supervised and Unsupervised Neural Machine translation Baseline for Nigerian Pidgin" arXiv:2003.12660v1[cs.CL]. 2020 DOI: https://doi.org/10.48550/arXiv.2003. 12660

[18] A. Gutkin, I. Demirashin, O. Kjartansson, C. E. Rivera, K. Tubosun , " Developing an Open-Source Corpus of Yoruba Speech" *Proceedings Interspeech , International Speech Communication Association.* pp. 404-408, 2020.

[19] I. Orife, "Neural machine translation for Edoid languages." *Journal of Intelligent Information Systems*, vol. 56. no. 2, pp. 299-315, 2020.

[20] D. I. Adelani, D. Ruiter, J. O. Alabi, D. Adebonojo, A. Ayeni, M. Adeyemi, A. Awokoya, and C. Espana-Bonet, "MENYO 20K: A multilingual parallel corpus for under-resourced languages." *In Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics:* Main Volume. pp. 3450-3463, 2021.

[21] Butryna, A., et al "Google Crowdsourced Speech Corpora and Related Open-Source Resources for Low-Resource Languages and Dialect*." Proceedings of the 12th Language Resources and Evaluation Conference,* pp.3424-3433, 2020. https://doi.org/10.48550/arXiv.2010.06778

[22] M. A. Hedderich, L. Lange, H. Adel, J. Strötgen, and D. Klakow. "A Survey on Recent Approaches for Natural Language Processing in Low-Resource Scenarios." In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp 2545–2568, Online. Association for Computational Linguistics.

[23] W. Nekoto. Participatory Research for Low-resourced Machine Translation: A Case Study in African Languages. 2020. arXiv:2010.02353v2, https://doi.org/10.48550/arXiv.2010.02353.

[24] S. Abate, M. Woldeyohannis, M. Tachbelie, M. Meshesha, S. Atnafu, W. Gewe, Y. Assabie, H. Abera, B. Seyoum, T. Abebe, W. Tsegaye, A. Lemma, T. Andargie, S. Shifaw (2018) Parallel corpora for bi-lingual English–Ethiopian languages statistical machine translation. In: Proceedings of the 27th international conference on computational linguistics, Santa Fe, New Mexico, USA, pp 3102–3111

[25] J. T. Sefara, V. Marivate, S. G. Zwane, N. Gama, H. Sibisi, P. N. Senoamadi. Transformer-based Machine Translation for Low-resourced Languages embedded with Language Identification.2021 Conference on Information Communications Technology and Society. IEEE, 2021. DOI: 10.1109/ICTAS50802.2021.9394996

[26] Y. Moukafih, N. Sbihi, M. Ghogho, K. Smaili. (2022). Improving Machine Translation of Arabic Dialects Through Multi-task Learning. In: Bandini, S., Gasparini, F., Mascardi, V., Palmonari, M., Vizzari, G. (eds) AIxIA 2021 – Advances in Artificial Intelligence. AIxIA 2021. Lecture Notes in Computer Science(), vol 13196. Springer, Cham. https://doi.org/10.1007/978-3-031-08421-8_40