NATURAL LANGUAGE PROCESSING

# YorùbáNER-X: An Explainable Hybrid Transformer–Linguistic Framework for Robust Named Entity Recognition in Low-Resource Yorùbá Text

Jelili Olalekan Amoo

Department of Computer Science, School of Secondary Education (Science Programmes), Federal College of Education (Special), Oyo, Nigeria

Published November 2026

**Abstract**

Named Entity Recognition (NER) is a key activity in Natural Language Processing; it facilitates information extraction, search, answering questions, building knowledge bases, and other downstream applications. Yorùbá NER is still hindered by a lack of labelled data, domain variation, orthographic inconsistency, and variations between standard and regional versions, despite recent advancements in African-language NLP. The empirical foundation for Yorùbá NER has been significantly strengthened by recent resources, especially MasakhaNER 2.0 and the multi-domain YoNER dataset. Nevertheless, recent data also indicate that transformer models suffer performance loss when transported across domains. In this paper, we propose YorùbáNER-X, an explainable hybrid framework that combines explicit Yorùbá linguistic features with transformer contextual representations and Association Rule Mining (ARM). The framework represents orthographic, character-level, lexical, contextual, and entity-indicative information alongside contextual transformer embedding; ARM is used as a supplementary evidence mechanism for recurring linguistic and contextual patterns rather than as a substitute for neural prediction; the framework is assessed using a controlled experimental design that includes standard NER evaluation, diacritic-variation tests, cross-domain evaluation, component ablation, statistical comparison, error analysis, and explainability analysis. The main baselines are XLM-R, an Africa-centric transformer, and OyoBERT. The study offers a repeatable methodological framework for examining whether pre-trained contextual representations in low-resource African-language NER can be enhanced by explicit language-specific information and statistical connections. In this publication, numerical experimental results are purposefully left out and will be added after implementation and assessment.

**Keywords:** Low-resource NLP; Named Entity Recognition; OyoBERT; Transformer; XLM-R

## 1 Introduction

Natural Language Processing (NLP), a major Artificial Intelligence field, aims to help computers understand, analyze, and synthesize human language (Fanni et al., 2023). Its uses include machine translation, information retrieval, question answering, sentiment analysis, and automated information extraction. Named Entity Recognition (NER), which identifies and classifies people, firms, places, dates, and other relevant expressions in textual data, is a major NLP challenge. NER is used to structure unstructured text for knowledge discovery and other language processing applications. Despite NLP advances, these tools are still unevenly available across languages. Annotated datasets, computational tools, linguistic resources, and language-specific models are still challenges for many African languages, although high-resource languages gain from massive digital corpora, pretrained language models, and evaluation criteria.

West African languages like Yorùbá demonstrate this disparity. Despite its cultural, historical, and communicative relevance, Yorùbá's computational resources have not improved as swiftly as those of other highly investigated languages (Saadu and Oyèwọlé, 2025). Recent evaluations of Yorùbá NLP research have identified code-switching, a lack of annotated data, and poor digital representation. Deep learning and transfer learning have replaced rule-based and statistical approaches in the business, yet reliable and adaptable language systems are still lacking. These problems highlight the necessity for study methods that accommodate for Yorùbá's language and low-resource language processing limits.

Enhancing African language NLP resources yields good results. MasakhaNER 2.0 included 20 languages and provided hand-annotated named entity recognition datasets for several African languages. These projects demonstrated the importance of language-specific evaluation materials and cross-lingual transfer methods (Adelani et al., 2024). Additionally, their findings suggest that transfer learning success depends on the source-target language relationship. Thus, information transfer from English or another high-resource language may not necessarily benefit African languages, especially when their orthographic standards and linguistic frameworks differ.

Transformer-based multilingual language models enable NLP with little resources. In addition to improving language understanding and sequence-labeling tests, models like the Cross-lingual Language Model–RoBERTa (XLM-R) provide strong multilingual representations (Conneau et al., 2020). However, target-language training data, pre-training coverage, and linguistic factors can alter their effectiveness. African-focused pre-trained models enhance this method. AfriBERTa, which used text from eleven African languages, including Yorùbá, performed well in text categorization and named entity recognition (Ogueji et al., 2021). These findings suggest that multilingual transfer and Africa-centric pre-training can further NER in under-resourced languages.

Recent research strengthens Yorùbá NER's resources. The MasakhaNER Yorùbá component and Yorùbá-specific annotated corpora are available. Falola et al. (2026) presented YoNER, a multi-domain Yorùbá NER dataset containing over 5,000 phrases and 100,000 tokens from the Bible, blogs, movies, radio broadcasts, and Wikipedia. This paper also introduced OyoBERT, a Yorùbá language representation model. However, cross-domain results reveal that model performance can suffer when evaluated on text from domains other than training data, particularly in blogs and movies. This highlights a major disadvantage of current methods: a model that works well in a well-known domain may not work in many real-world communication scenarios.

Another major problem for Yorùbá NER is orthographic variance. Diacritical markings are used in the language to indicate specific vowel and tonal contrasts, although they are often missing or inconsistently rendered in digital writing (Asahiah et al., 2017). Additionally, Yorùbá text may use code-switching, informal spellings, and English expressions. As a result, models that were primarily trained on standardised, fully diacritical text may find it difficult to handle informal messages, social media posts, digital cultural assets, and other naturally occurring language data. The impact of removing diacritical marks on recognition performance and the possible advantages of adding orthographic variation to training have been noted in recent Yorùbá NER research. Addressing this challenge requires approaches that can exploit contextual information while remaining sensitive to language-specific spelling and character patterns. Together, these advances demonstrate that the main research topic goes beyond transformer models for Yorùbá text. It involves constructing NER systems that can forecast with clear explanations and withstand orthographic variation, cross-domain distribution shifts, and a lack of training data. Transformer architectures, multilingual

transfer, specialized pretraining, and data augmentation have advanced low-resource natural language processing (NLP), but more research is needed to integrate learned contextual representations with statistical patterns and explicit linguistic knowledge. This integration may be useful when labeled examples are few and language-specific knowledge helps improve pretraining representations.

YorùbáNER-X, an explainable hybrid framework for Named Entity Recognition, is proposed in this study to address these challenges. The system combines transformer-based language model contextual representations with association-rule knowledge and Yorùbá-specific linguistic traits. While association rule mining finds recurrent correlations between linguistic properties and entity labels, the linguistic component captures useful orthographic, character-level, lexical, and contextual patterns. An explainability component is added to investigate individual predictions and how contextual representations, linguistic patterns, and acquired relationships affect entity categorization. Merging these alternative information sources may improve recognition robustness compared to transformer representations alone.

The proposed framework will be compared to neural and transformer-based baselines using entity-level precision, recall, and F1-score. Cross-domain evaluation, diacritical representation differences, and reduced training-data performance will be tested. Ablation experiments will distinguish association-rule knowledge, Yorùbá-specific linguistic features, and transformer representations. Explainability analyzes will determine if the model's predictions are supported by significant language and contextual data. These studies will establish whether the hybrid architecture improves predictive performance and whether factors support its efficacy and robustness. This research should help create more reliable and understandable NLP systems for Yorùbá and other African languages with low resources. Improved Yorùbá NER could improve information retrieval, digital archiving, instructional technologies, conversational systems, cultural heritage preservation, and multilingual information extraction. The research also examines how to integrate transformer-based learning with explicit linguistic knowledge and explainability to support larger efforts to create language technologies that represent the diversity, orthographic complexity, and resource limitations of languages underrepresented in mainstream NLP research.

### 1.1 Statement of the problem

Despite recent advances in Yorùbá Named Entity Recognition (NER), such as benchmark datasets and pretrained language models, the cross-domain evaluation of YoNER illustrates how challenging it is to transfer NER performance across different types of Yorùbá text. Generalization to unexpected situations may not be guaranteed by benchmark dataset performance. Domain, vocabulary, annotation techniques, and labeled training data can affect transformer-based models (Falola et al., 2026).

Babatunde et al. (2026) reported that the lack of Yorùbá linguistic characteristics in transformer-based contextual models is a major limitation. Transformers can learn language patterns from training data, but they do not provide orthographic, character-level, lexical, and contextual properties for direct analysis or controlled inclusion. In Yorùbá, diacritical markings affect linguistic interpretation, and their inconsistent use in digital communication may hamper entity recognition.

Traditional NER assessment relies on precision, recall, and F1-score, which measure predictive efficacy but don't explain specific predictions. Without explainability frameworks, determining whether a model relies on significant linguistic features or accidental patterns in its training data can

be difficult. This issue is especially important in resource-constrained contexts where labeled data are sparse and error analysis is essential for model reliability.

This study aims to address the necessity for a systematically evaluated Yorùbá NER framework that integrates transformer-based contextual representations, explicit linguistic attributes, association-rule-based evidence, and explainability. To assess whether this integration can improve entity recognition efficacy, resilience to orthographic variation, generalization across domains, and interpretability in scenarios with constrained training data, the research proposes YorùbáNER-X.

### 1.2 Research gap

A unified Yorùbá NER framework that jointly integrates transformer representations, explicit Yorùbá-specific linguistic features, association-rule-based evidence, and explainability with systematic evaluation under low-resource, orthographic-variation, and cross-domain conditions is the specific research gap. Although Yorùbá NER has been improved by previous research using pre-trained language models and dataset building, it is still unclear how much these complimentary methods can work together to increase recognition accuracy, robustness, and interpretability.

### 1.3 Aim and objectives of the study

The aim of this study is to develop and evaluate YorùbáNER-X, an explainable hybrid framework for robust Named Entity Recognition in low-resource Yorùbá text.

The specific objectives of the study are to:

1. Develop a Yorùbá text representation method that captures diacritic patterns, lexical features, contextual information, and relevant linguistic variations to improve Named Entity Recognition (NER).
2. Design and develop the YorùbáNER-X hybrid framework by integrating transformer-based contextual representations, Yorùbá-specific linguistic features, association rule mining, and explainable AI techniques.
3. Assess the performance and robustness of YorùbáNER-X in comparison with established NER models using appropriate evaluation metrics under limited training data, orthographic variations, and cross-domain settings.
4. Perform ablation, explainability, and error analyses to examine the contributions of individual framework components, interpret model predictions, and identify the strengths and limitations of YorùbáNER-X.

### 1.4 Research questions

The following research questions are formulated in line with the four specific objectives of the study:

1. How can Yorùbá-specific linguistic features, including diacritic patterns, lexical characteristics, contextual information, and orthographic variations, be effectively represented for Named Entity Recognition?
2. How can transformer-based contextual representations, Yorùbá-specific linguistic features, association-rule mining, and explainability be integrated into the proposed YorùbáNER-X hybrid framework?

3. To what extent does YorùbáNER-X improve Named Entity Recognition performance and robustness compared with established NER models under limited training-data conditions, orthographic variations, and cross-domain settings?
4. What contributions do the individual components of YorùbáNER-X make to its performance and interpretability, and what strengths and limitations are revealed through ablation, explainability, and error analyses?

## 2 Literature review

### 2.1 Low-resource African language NLP

Scarce-resource NLP is distinguished by limited computational resources, annotated datasets, pretrained representations, assessment benchmarks, and linguistic tools, in addition to a small number of speakers. These restrictions particularly impact African languages. MasakhaNER was presented by Adelani et al. (2021) as a manually annotated NER benchmark for African languages, proving that extensive comparative analysis is feasible. MasakhaNER 2.0 studied Africa-centric transfer learning and extended the benchmark to twenty African languages, demonstrating that performance can be greatly impacted by the choice of transfer language (Adelani et al., 2022).

These studies establish two principles that are pertinent to the current work: first, transfer performance is not uniform across languages, so general multilingual models should not be considered universally optimal for low-resource African languages; second, benchmark resources are necessary for dependable progress because they enable methods to be evaluated under comparable conditions.

### 2.2 Transformer models for multilingual and African NLP

In current NLP, transformer architectures are fundamental as self-attention facilitates contextual representations that may express relationships among tokens during quite lengthy sequences. Multilingual BERT and XLM-R revealed the promise of multilingual pretraining for knowledge transfer among languages. XLM-R, utilising large-scale multilingual training data, became a key reference point for cross-lingual sequence tagging and other related tasks (Conneau et al., 2020).

African-language-specific models have also arisen. AfriBERTa was created to explore the potential of acquiring beneficial pretrained representations for under-resourced African languages with limited data availability (Ogueji et al., 2021). Recently, language-specific tools like OyoBERT have offered a more robust Yorùbá-focused representation for subsequent tasks. The current research thus regards transformer design as a robust basis while exploring if explicit linguistic and associative data might yield further advantages.

### 2.3 Yorùbá NLP and NER

A decade's worth of Yorùbá NLP research has generated resources for tasks including machine translation, sentiment analysis, speech processing, language modelling, and information extraction. Adeboje et al. (2025) examined a decade's worth of Yorùbá NLP research and underscored ongoing difficulties concerning resources, pre-trained models, linguistic traits, diacritics, and extensive coverage. The critique endorses the necessity for studies that progress past individual task performance to include thorough resource and durability analysis.

Specifically for NER, Yorùbá Global Voices, MasakhaNER, MasakhaNER 2.0, and YoNER furnish an expanding array of materials. Advancing from news-focused data to multi-domain evaluation is vital as named-entity distributions, writing styles, language, and contextual elements can differ markedly by domain.

### 2.4 Dialectal and orthographic variation

Yorùbá represents a dialect continuum instead than a wholly homogeneous language framework. Ahia et al. (2024) created YORULECT spanning Standard Yorùbá, Ifẹ̀, Ilàjẹ, and Ijẹ̀bú varieties and found major performance differences between standard and regional variants. Although their study focused on translation and speech-related tasks rather than NER, its findings suggest that NLP systems created primarily for Standard Yorùbá should not automatically be assumed to generalize across dialects.

Orthographic diversity holds significance as well. Yorùbá orthography employs tonal and under-dot diacritics to differentiate lexical variations. In digital communication, users may forgo certain or all diacritical marks. A resilient NER system ought to be assessed under multiple orthographic conditions instead of supposing uniformly standardised input.

### 2.5 Explainable NLP

Contemporary neural architectures can generate remarkably precise forecasts, however they remain challenging to decipher (Kong et al. 2025). Techniques for explainability like SHAP and Integrated Gradients can be employed to examine feature attribution and predictive justification. In sequence labelling, interpretative techniques can assist in pinpointing tokens, contextual cues, or clusters of features linked to a prediction. In the context of low-resource languages, interpretability is especially advantageous for diagnosing errors as it can indicate whether the model depends on significant linguistic evidence or on coincidental correlations within a restricted training dataset.

The current research does not consider an attribution score as evidence of causal inference. Explanations provide as analytical proof that enhances quantitative performance and qualitative error assessment.

### 2.6 Empirical review and research gap

The practical advancement of Yorùbá NER can be encapsulated as a transition from restricted news-domain assets to African-language transfer standards, to multi-domain assessments and language-specific pretrained frameworks. This advancement has significantly fortified the discipline, yet it simultaneously uncovers a fresh array of enquiries.

Table 1 summarises the main Yorùbá NER resources and approaches and their limitations relevant to this study.

**Table 1.** Yorùbá NER resources and approaches.

| Resource/ Approach | Main contribution | Limitation relevant to this study |
|---|---|---|
| Global Voices NER | Early manually annotated Yorùbá NER resource | Primarily news-domain |
| MasakhaNER | African-language NER benchmark | Limited domain diversity |

| Resource/ Approach | Main contribution | Limitation relevant to this study |
|---|---|---|
| MasakhaNER 2.0 | Expanded African benchmark and transfer-learning study | Does not directly solve linguistic robustness or explainability |
| YORULECT | Evidence of regional Yorùbá variation | Not primarily an NER benchmark |
| YoNER | Multi-domain Yorùbá NER; 5,000 sentences and 100,000 tokens | Cross-domain degradation remains; three principal entity types |
| OyoBERT | Yorùbá-specific pretrained representation | Transformer-only representation does not explicitly test hybrid linguistic/association evidence |
| YorùbáNER-X | Proposed hybrid and robustness framework | Requires empirical validation |

The ensuing research void is consequently not characterised by a lack of Yorùbá NER datasets. Four critical deficiencies are present: constrained controlled low-resource assessment; insufficient systematic incorporation of explicit Yorùbá linguistic characteristics with robust transformers; inadequate exploration of association-rule evidence as supplementary data; and a lack of focus on explainability and error mechanisms. This research tackles these deficiencies via a cohesive experimental framework.

### 2.7 Conceptual framework

YorùbáNER-X uses a multi-branch representation framework. Given a Yorùbá input sequence X, the transformer branch produces contextual representations H, the linguistic branch produces explicit features L, and the association branch produces rule-based evidence R. The three representations are fused before final token classification.

$$H = \text{Transformer}(X) \tag{1}$$

$$L = \text{LinguisticFeatures}(X) \tag{2}$$

$$R = \text{AssociationEvidence}(X) \tag{3}$$

$$Z = [H; L; R] \tag{4}$$

$$P(y_i|X) = \text{softmax}(WZ_i + b) \tag{5}$$

The architecture intentionally regards ARM as supplementary evidence instead of a definitive decision-making tool. This enables the trials to ascertain if repetitive linkages enhance predictions while preventing errant rules from superseding contextual representations.

## 3 Methodology

### 3.1 Research design

A quantitative experimental framework will be employed. The research evaluates baseline models alongside the suggested hybrid model in both standard and robustness scenarios. The primary datasets consist of MasakhaNER 2.0 Yorùbá and YoNER. Experiments using incompatible label inventories will be restricted to common entity types rather than silently combining disparate annotation techniques.

### 3.2 Data preparation

The data will be transformed into a uniform BIO sequence-labeling format. Data preprocessing will encompass the elimination of erroneous records, verification of duplicates, normalisation of Unicode, tokenisation, alignment of annotations, and detection of code-switched segments. We shall maintain the original diacritics to facilitate experimental assessment of orthographic robustness instead of discarding them during preprocessing.

### 3.3 Baseline models

The essential family will consist of a traditional sequence-labelling model where suitable, XLM-R as a general-purpose multilingual transformer, an Africa-centric transformer that supports Yorùbá, and OyoBERT as the key Yorùbá-specific transformer baseline. This baseline determination is purposely stringent: the proposed model must illustrate its utility beyond effective language-specific contextual representations.

### 3.4 Linguistic features

The linguistic domain will illustrate diacritic occurrence and configurations, letter n-grams, prefixes and suffixes, token morphology and length, lexical identity, adjacent tokens, punctuation, and contextual patterns indicative of entities. Feature selection and normalisation will be conducted exclusively with training and validation datasets to avert test-set leaks.

### 3.5 Association rule mining

Apriori will initially be used to identify frequent associations among linguistic and contextual feature items. For a rule $A \rightarrow B$, support is $\text{Support}(A \rightarrow B) = \text{Count}(A \cup B)/N$ and confidence is $\text{Confidence}(A \rightarrow B) = \text{Count}(A \cup B)/\text{Count}(A)$. Lift will also be considered to identify associations stronger than chance baseline frequency. Minimum support and confidence will be selected on validation data. Very rare rules will be pruned to reduce instability.

### 3.6 Hybrid fusion

The principal fusion mechanism will concatenate projected transformer, linguistic, and association representations. For token $i$, $Z_i = [P_H(H_i); P_L(L_i); P_R(R_i)]$, followed by a trainable projection and softmax classifier. A gated fusion variant may be examined if validation results justify its additional complexity.

### 3.7 Explainability

SHAP, Integrated Gradients, or a suitable attribution technique will be employed to examine chosen predictions. Explanations will be scrutinised for both accurate and inaccurate predictions, especially in the context of domain shift and orthographic variation. The research will clearly differentiate attribution evidence from causal explanation.

### 3.8 Evaluation protocol

Performance will be measured using entity-level precision, recall, and F1. Macro and micro/entity-level scores will be reported where appropriate. The study will additionally report performance by entity type, training-data fraction, orthographic condition, and test domain. Multiple random seeds will be employed, and confidence intervals, effect sizes, and suitable paired statistical tests will accompany major comparisons.

### 3.9 Robustness experiment

Three categories of robustness will be examined. Initially, the training data will be diminished to 100%, 50%, 25%, 10%, and 5%, contingent upon dataset size. Subsequently, the evaluation data will be assessed under fully diacritized, partially diacritized, and non-diacritized scenarios. Finally, models will undergo evaluation across various domains utilising cross-domain and, where applicable, leave-one-domain-out methodologies.

### 3.10 Ablation

The core ablation sequence will compare: transformer only; transformer plus linguistic features; transformer plus ARM; and the whole transformer-plus-linguistic-plus-ARM architecture. The objective is to identify the contribution of each suggested component.

### 3.11 Experimental setup

The training process will employ Python alongside PyTorch and Hugging Face Transformers, utilizing suitable datasets, tokenization, evaluation, and explainability libraries. We will determine hyperparameters, including learning rate, batch size, maximum sequence length, dropout, weight decay, number of epochs, ARM support, and ARM confidence, via a validation-driven search. We will document specific software versions, model identifiers, random seeds, and hardware configurations to ensure reproducibility.

The proposed hyperparameter search space is given in Table 2.

**Table 2.** Proposed hyperparameter search space.

| Parameter | Proposed search/setting |
|---|---|
| Learning rate | 1e-5, 2e-5, 3e-5, 5e-5 |
| Batch size | 8, 16, 32 |
| Epochs | 3–10 with early stopping |
| Maximum sequence length | 128, 256, 512 |
| Dropout | 0.1, 0.2, 0.3 |
| Training proportions | 100%, 50%, 25%, 10%, 5% |
| ARM support | 0.01–0.10 |
| ARM confidence | 0.50–0.90 |
| Random seeds | Multiple independent runs |

## 4 Results and reporting framework

Due to the fact that the experiments have yet to be conducted, this research framework does not generate numerical results. The ultimate submission must substitute the specified placeholders with empirical findings, confidence ranges, significance assessments, and effect magnitudes.

Table 3 shows the template for reporting the main results.

**Table 3.** Template for reporting the main results (TBD = to be determined).

| Model | Precision | Recall | F1 | Runtime |
|---|---|---|---|---|
| XLM-R | TBD | TBD | TBD | TBD |
| African-centric model | TBD | TBD | TBD | TBD |
| OyoBERT | TBD | TBD | TBD | TBD |
| YorùbáNER-X | TBD | TBD | TBD | TBD |

Additional tables should report low-resource performance, diacritic robustness, cross-domain performance, and ablation results. The discussion must be based only on the observed results and should report both improvements and cases in which the proposed method does not outperform a baseline.

## 5 Discussion

The anticipated analytical contribution of this research is to ascertain if explicit language information adds supplementary value to transformer representations. A favorable outcome would bolster the assertion that language-specific characteristics can retain their use despite the presence of a robust pre-trained model. A null or negative outcome would also be enlightening since it would suggest that the chosen explicit features or association mechanism fail to yield adequate supplementary information under the examined conditions.

The low-resource experiment holds significant importance as benchmark performance alone could obscure the rapid decline of a model when annotations are limited. Should the hybrid model retain a greater percentage of its full-data efficacy as supervision diminishes, the outcome would bolster the suggested resilience theory. If the drop is similar to or greater than the baseline decline, the study should report that finding and analyze the underlying error patterns.

Cross-domain evaluation provides another essential test. YoNER illustrates that domain transfer is problematic for Yorùbá NER, notably for blogs and movie material. A good hybrid system should therefore be judged according to stability across domains rather than merely its greatest in-domain F1. Domain-specific errors should be investigated to assess whether linguistic and association features capture reusable patterns or only encode training-domain regularities.

The diacritic experiment necessitates meticulous analysis. Removing diacritics modifies the information accessible to the model and may generate uncertainty. The purpose is not to show that non-diacritized text is identical to ordinary written Yorùbá, but to measure how sensitive a NER system is to a controlled and realistically meaningful input perturbation.

Finally, explainability and error analysis should connect quantitative performance to model behaviour. If attribution analysis regularly identifies meaningful contextual or linguistic information, it might assist researchers grasp the model. Conversely, explanations that continually highlight irrelevant or corpus-specific aspects would indicate an essential problem even if aggregate F1 were high.

## 6 Limitations

The study depends on the coverage, annotation quality, and label definitions of available Yorùbá datasets.

- YoNER covers five domains and three main entity categories and therefore does not represent every possible Yorùbá communication setting.
- Controlled diacritic removal is a simulation and may not reproduce every naturally occurring orthographic pattern.
- Association rules can become sparse or corpus-specific when training data are limited.
- Explainability methods provide attribution evidence and should not automatically be interpreted as causal explanations.
- Computational requirements may increase when transformer, linguistic, ARM, and explainability components are combined.
- Results obtained on Standard Yorùbá should not automatically be generalized to all regional dialects.

## 7 Conclusion

This research showcased YorùbáNER-X, an interpretable hybrid framework for resilient Named Entity Recognition in under-resourced Yorùbá literature. The paradigm caters to a transforming research landscape in which emerging resources like MasakhaNER 2.0 and YoNER allow for more thorough evaluations, while ongoing research continues to uncover issues linked to topic shift, linguistic diversity, and resource restrictions.

The outlined framework fuses transformer contextual representations with overt Yorùbá linguistic elements and association-rule insights. Its review goes beyond typical in-domain F1 by encompassing reduced-data trials, systematic diacritic variations, cross-domain testing, ablation, statistical analysis, error assessment, and clarity of explanation.

The fundamental methodological thesis of the research suggests that low-resource NER must be regarded as a robustness challenge alongside a benchmark-accuracy concern. The final assessment of YorùbáNER-X's worth will thus rely on empirical evidence illustrating whether the added linguistic and associative information leads to consistent and practically relevant gains over powerful transformer models. Empirical research articles necessitate the acquisition of actual numerical results before submission for publication.

The approach is likely extendable to several African languages with insufficient annotated data, notably in situations where explicit linguistic expertise and orthographic differences are essential. Future studies should examine regional Yorùbá dialects, code-switched Yorùbá-English text, parameter-efficient adaptation, entity linkage, and bigger models for African languages.

## 8 Recommendations for future research

- Develop larger, multi-domain, manually annotated Yorùbá NER resources with documented annotation guidelines.
- Extend NER evaluation to regional Yorùbá varieties and dialect-specific resources.
- Investigate Yorùbá-English code-switched NER in social and conversational media.

- Compare association-rule evidence with graph-based and neuro-symbolic knowledge representations.
- Evaluate parameter-efficient fine-tuning to reduce the computational cost of language-specific adaptation.
- Release code, configurations, preprocessing scripts, and model checkpoints where licensing permits.
- Develop standardized robustness benchmarks covering orthographic variation, domain shift, and limited supervision.

## References


[1] Adeboje OA, et al. (2025). Bridging gaps in natural language processing for Yorùbá: A systematic review of a decade of progress and prospects. *Natural Language Processing Journal*, 13, 100194. https://doi.org/10.1016/j.nlp.2025.100194

[2] Adelani DI, Abbott J, Neubig G, D'Souza D, Kreutzer J, Lignos C, et al. (2021). MasakhaNER: Named entity recognition for African languages. *Transactions of the Association for Computational Linguistics*, 9, 1116–1131.

[3] Adelani DI, Neubig G, Ruder S, Rijhwani S, Beukman M, Palen-Michel C, Lignos C, Alabi JO, Muhammad SH, Nabende P, et al. (2022). MasakhaNER 2.0: Africa-centric transfer learning for named entity recognition. *Proceedings of the* 2022 Conference on Empirical Methods in Natural Language Processing, 4488–4508.

[4] Adelani DI, Liu H, Shen X, Vassilyev N, Alabi J, Mao Y, et al. (2024). SIB-200: A simple, inclusive, and big evaluation dataset for topic classification in 200+ languages and dialects. In *Proceedings of the* 18th Conference of the European Chapter of the Association for Computational Linguistics (Volume 1: Long Papers) (pp. 226-245).

[5] Ahia O, Aremu A, Abagyan D, Gonen H, Adelani DI, Abolade D, Smith NA, Tsvetkov Y (2024). Voices unheard: NLP resources and models for Yorùbá regional dialects. *Proceedings of the* 2024 Conference on Empirical Methods in Natural Language Processing, 4392–4409. https://doi.org/10.18653/v1/2024.emnlp-main.251

[6] Asahiah FO, Odejobi OA, Adagunodo ER (2017). Restoring tone-marks in standard Yorùbá electronic text: improved model. *Computer Science*, 18(3).

[7] Babatunde AN, Ajao JF, Kayode AA, Babatunde RS, Isiaka SO, Ogundokun RO, Lee CC (2026). Sentiment classification in the Yoruba language using multilingual Bidirectional Encoder Representations from Transformers (mBERT). *International Journal of Speech Technology*, 29(1), 9.

[8] Conneau A, Khandelwal K, Goyal N, Chaudhary V, Wenzek G, Guzmán F, Grave E, Ott M, Zettlemoyer L, Stoyanov V (2020). Unsupervised cross-lingual representation learning at scale. *Proceedings of the* 58th Annual Meeting of the Association for Computational Linguistics, 8440–8451.

[9] Falola PB, Alabi J, Akinola SO, Ogunajo FT, Alabi EO, Adelani DI (2026). YoNER: A new Yorùbá multi-domain named entity recognition dataset. *Proceedings of the Fifteenth Language Resources and Evaluation Conference*, 7137–7148. https://doi.org/10.63317/4o6m2hzpy7gs

[10] Fanni SC, Febi M, Aghakhanyan G, Neri E (2023). Natural language processing. In *Introduction to artificial intelligence (pp*. 87-99). Cham: Springer International Publishing.

[11] Kong X, Chen Z, Liu W, Ning K, Zhang L, Muhammad Marier S, et al. (2025). Deep learning for time series forecasting: a survey. *International Journal of Machine Learning and Cybernetics*, 16(7), 5079–5112.

[12] Ogueji K, Zhu Y, Lin J (2021). Small data? No problem! Exploring the viability of pretrained multilingual language models for low-resourced languages. Proceedings of the 3rd Workshop on Multilingual Representation Learning, 116–126.

[13] Saadu T, Oyèwọlé JA (2025). Enhancing Yorùbá language education with technology. *Ondo Journal of Arts (OJA)*, 1(1), 177–198.