



A Comprehensive Review of the *Yorùbá* Numbering Scheme: Linguistic Structure, Computational Modelling, and Digital Preservation

R.K. Adeeyo
Department of Intelligent
System Engineering
Obafemi Awolowo University
Ile-Ife, Nigeria

O.A. Odejobi
Department of Intelligent
System Engineering
Obafemi Awolowo University
Ile-Ife, Nigeria

A.B. Olorunlomeye
Department of Intelligent
System Engineering
Obafemi Awolowo University
Ile-Ife, Nigeria

J.A. Hassan
Department of Intelligent
System Engineering
Obafemi Awolowo University
Ile-Ife, Nigeria

ABSTRACT

The *Yorùbá* numeral system is one of the most complex indigenous numeral systems in the literature, with a vigesimal (base 20), decimal, and quinary structure and the usage of addition, subtraction, multiplication and division (algebra) in the computation of numbers. With the recent advances in computational linguistics and natural language processing (NLP) technologies in low-resource African languages, there has been a surge of research on the morphology, translation and computational Modelling of *Yorùbá* numerals. This paper is based on a systematic literature review (SLR) of different studies either directly or indirectly related to *Yorùbá* numeral from 1970 to 2026, which focus on the linguistic structure, computational modelling, and educational application of the *Yorùbá* number system, as well as comparative studies of the role of number systems in other computing fields. The reviewed works are summarised in five thematic areas: (i) morphological and linguistic analysis, (ii) computational and machine translation systems, (iii) educational technology and language preservation tools, (iv) implications of the vigesimal system and of computing architectures, and (v) number systems in other areas of computing. The synthesis reveals that although the linguistic structure of *Yorùbá* numerals is relatively well documented, computational tools that can effectively process large numerals, context-sensitive translation and integration with modern NLP pipelines are still in their infancy and are fragmented. The paper ends with a research agenda for a standardized, data driven computational framework for processing *Yorùbá* numerals.

General Terms

Computational linguistics, natural language processing, language preservation, machine translation.

Keywords

Yorùbá numbering scheme, linguistic structure, computational modelling, digital preservation.

1. INTRODUCTION

One of the oldest cognitive and linguistic technologies created by human societies are number systems, which not only

represent quantity but also cultural logic, trade practice, and worldview. The *Yorùbá* system of numbers from southwestern Nigeria is computationally vigesimal, decimal, and quinary, which is based on multiplication, addition, and uniquely subtraction to create numerals for unit, tens, hundreds, and thousands. Unlike decimal numeral, which denotes “55” as “ $50 + 5$ ” and “300” as “ 3×100 ”, the *Yorùbá* language forms these numbers as recursive algebraic expressions as in “ $55 = 20 \times 3 - 10 + 5$ ” and “ $300 = 400 - 100$ ” which must be parsed and evaluated to determine their value. This structural richness is what makes the *Yorùbá* numeral system a rich object of linguistic interest and a difficult object for computational Modelling. Indigenous numeral systems like that of the *Yorùbá* are under pressure in the digital age. The use of numbers in English through digital devices, applications and education materials with younger generations may lead to the loss of the indigenous numeral system. Meanwhile, developments in computational linguistics, rule-based translation, and machine learning have opened new avenues to document, formalize, and archive the *Yorùbá* numeral system through software, Web-based applications, mobile applications, and text-to-speech and text-conversion systems. In the last five decades, a body of research has grown concerning the linguistic description of *Yorùbá* numerals, computational methods of numeral translation and conversion, and the application of educational technology in numeral instruction and language preservation. This literature, however, is spread out in the fields of linguistics, computer science and education, and no synthesis, yet brings the various threads together into a coherent view of the state of the art. This gap inspires the present systematic literature review (SLR) and corresponding Research Questions (RQ); What linguistic and morphological properties of *Yorùbá* numerals have been reported in the literature? (RQ1), (RQ2) What computational methods have been suggested for translating, generating or converting *Yorùbá* numerals? (RQ3) What tools and systems are available for the educational use and cultural preservation of the *Yorùbá* numeral system? and (RQ4) What are the gaps between the linguistic documentation of *Yorùbá* numerals and the maturity of computational tools that process them? The remainder of this paper is organised as follows. The review methodology used for this study is detailed



in Section 2. Quantitative analysis and evaluation of the reviewed corpus is presented in Section 3. The cross-cutting findings and research gaps are discussed in Section 4. Future research directions are outlined in Section 5, and the paper is concluded in Section 6 with future research direction.

2. REVIEW METHODOLOGY

This review is conducted in a manner similar to a systematic literature review, with the following characteristics; transparent identification of the studies included, explicit inclusion criteria, and thematic synthesis of the results. Several studies were found between 1970 and 2026. The corpus includes works directly addressing the *Yorùbá* numeral system from a linguistic, morphological or computational point of view, as well as closely related works on the numeral system of other ethnic groups in Nigerian languages, such as Hausa and Igbo, which use similar approaches, and a collection of comparative studies on the use of number systems in the main computing areas (machine learning, operating systems, data science, robotics, financial computing, and artificial intelligence). The latter group was added to offer a comparative context for assessing the potential for indigenous numeral systems to be placed within and benefit from existing computing paradigms. The author(s), year of publication and title of the study were noted along with the problem statement, methodology used, and results reported. The studies were then thematically analysed based on the main contribution made, resulting in five clusters namely; morphological and linguistic analyses, computational and machine translation systems, educational technology and language preservation tools, vigesimal system and computing architecture implications, and number systems in other computing fields. The methodologies found in the corpus encompass qualitative morphological analysis, ethnographic fieldwork, rule-based computational Modelling, empirical software development and evaluation, construction morphology, descriptive linguistics and comparative technical review. This methodological variety is due to the interdisciplinary character of the subject matter which covers the fields of linguistics, computer science and education.

2.1 Morphological and Linguistic Analyses of Yorùbá Numerals

Much of the literature reviewed here consists of descriptions of the internal structure of *Yorùbá* numerals by means of morphological and linguistic analysis. Èkúndayò [3] gave an early and fundamental description of the vigesimal derivational morphology of *Yorùbá* numerals, describing the compounding, suffixation, and subtractive patterns that produce numbers greater than twenty, and presenting this description as a foundation for further computational treatment. Zaslavsky [5] placed the *Yorùbá* numeral system in a wider ethnomathematical context of the *Yorùbá* people and their neighbours in southern Nigeria, and ethnographically demonstrated the influence of cultural context, trade and social organisation on the evolution of the number system. Also, Babárinde [11] provided a detailed descriptive account of the counting system, describing its derivational patterns to enrich the *Yorùbá* numeral database for research purpose. Omachonu [22] describes *Yorùbá* numeral structure typologically and highlight the uniqueness in its formation rule. Ajiboye [20] presents the *Yorùbá* numeral as a rule-based that is vigesimal in nature, built on addition and subtraction and Ajibádé [9] used a construction morphology approach, where numerals were classified based on the arithmetic operations they share, to show that the *Yorùbá* numerals exhibit systematic form,

meaning correspondences within numeral classes. More recently, Ayanwoye & Aiyede [24] frames *Yorùbá* numeral grammar as tied to cognition and cultural identity. This tradition was continued by Adeniyi *et al* [1] who added a computational perspective to the morphological description, examining the numeral multiplicands of the *Yorùbá* language in terms of compounding and inflectional morphology, and explicitly observing the lack of a simplified and systematic approach to the formation of large numbers as a barrier to a full computational analysis.

Together these studies confirm that the formation of numerals in *Yorùbá* is under identifiable, rule-like processes, which in principle can be formally and algorithmically described, compounding, subtraction, multiplication and suffixation. What is observed in this cluster, though, is that there is no agreed and widely accepted full systematic and computation-ready grammar of *Yorùbá* numeral formation, especially for very large numbers, yet. What is observed, however, in this cluster, from Èkúndayò [3] to Adeniyi *et al* [1], is that a full systematic and computation-ready grammar of *Yorùbá* numeral formation, for very large numbers, has not yet been agreed upon or widely accepted.

2.2 Computational and Machine Translation Systems

A second group of studies has taken a step from the linguistic description to the computational implementation of the machine translation process, dealing explicitly with the problem that multiple valid *Yorùbá* renderings are possible for a single English number, depending on the context. Isaac [2] designed a rule-based machine translation system for the translation of English numeral text to *Yorùbá*. They relied on pre-defined translation rules to choose the translation of the numerals from the set of numeral translations, which demonstrates the feasibility and complexity of context-sensitive numeral translation. Akínadé & Odejóbi [7] worked on computational modelling approach in a number-to-text conversion, which involved the automatic generation of *Yorùbá* numeral text from numeric input, while Babarinde [6] adopted an empirical system development approach, which involved the practical implementation of *Yorùbá* numeral translation algorithms in computational platforms. Similar approaches have been used for other related Nigerian languages. Eludiora [10] created a number transcription system for Hausa language, which translates numbers from Hindu-Arabic to Hausa and vice versa using a mixture of rule-based phonetic and orthographic rules and machine learning methods, and Rhoda *et al*. [12] developed a number-to-text conversion system for Igbo language, using both rule-based and algorithmic approaches, validated by experts for accuracy. Jimoh *et al* [18] conducted a systematic analysis of 105 primary studies between 2014 and 2024 to identify challenges, resources, techniques and applications of *Yorùbá* NLP, and discovered that the most popular techniques used were rule-based, statistical, deep learning and transfer learning, with various *Yorùbá* speech and text datasets and Orife [19] transformed Automatic Diacritic Restoration into a machine translation problem and employed two attentive Sequence-to-Sequence neural models to restore diacritics to undiacritized text, with the results showing diacritization error rates lower than 5%. More recently, Adewole *et al*. [15] worked on rule-based Arabic numeral to *Yorùbá* translator.

The review explicitly states that *Yorùbá* is a tonal and morphologically rich language with limited use of NLP, and that the lack of resources is a persistent challenge, as there is



scarcity of annotated corpora, the limited availability of pre-trained language models, and the linguistic challenges of tonal complexity and diacritic dependency.

This shows that most of the current computational treatment of the Nigerian indigenous numeral systems are rule-based and hybrid rules/machine learning approaches. It also shows a common weakness; evaluations are often based on small expert panels or on a small number of test sets, with no large, standardized benchmark corpus, limiting the claims that can be made about the generalizability and robustness of the systems, especially for numerals outside the most commonly used ranges.

2.3 Educational Technology and Language Preservation

A third group focuses on the use of technology in teaching, reinforcing and retaining knowledge of the *Yorùbá* numeracy system, especially for younger speakers. Eludiora [8] developed an arithmetic multimedia learning system for *Yorùbá* language based on UML design and Python/PyQt implementation and conducted an evaluation of the system using questionnaires from the learners, which revealed that a significant number of learners had difficulty producing correct numeral expressions, indicating that numeral acquisition is a pedagogical challenge. To address the potential for numeral knowledge to be lost in school-age speakers, Elesemoyo [23] and Jimoh *et al.* [14] created a mobile application that uses Android Studio and an SQLite database to incorporate *Yorùbá* numerals into a mobile assisted language learning environment, finding that this resulted in increased learner engagement and retention. Agbeyangi [13] explored another educational technology by developing a Web-based *Yorùbá* numeral translation system through a case study of numeral-integrated educational software, while Asahiah [16], Asahiah *et al.* [17] discussed how diacritic restoration tools support digital preservation of correctly tone-marked *Yorùbá* text, this enables digital text processing of *Yorùbá* numeral.

The studies all show that educational software and mobile applications can have measurable impact on the engagement of learners in the learning of *Yorùbá* numerals, but also that the tools are largely isolated, small-scale prototypes and not part of an interoperable, shared digital infrastructure for *Yorùbá* language technology.

2.4 Vigesimal System and Computing-Architecture Implications

Ekundayò [3] documented *Yorùbá* numeral as a number system that uses vigesimal (base 20) only, then explained the potential of the *Yorùbá* base 20 system for computational contexts, suggesting that vigesimal structures may provide alternative methods of data representation and processing that differ from the binary and decimal systems that are prevalent in contemporary computing. Similarly, Verran [21] presented deep context on how the vigesimal (base 20) functions conceptually and functionally. Ajiboye [20] discussed the ways additive and subtractive vigesimal rules complicate direct positional numeral computing approaches. Although largely speculative and exploratory, this cluster suggests that culturally based number systems, other than the decimal system, might provide new insights into how numbers are represented in software for multilingual and cross-cultural applications.

2.5 Number Systems in Broader Computing Domains

The last cluster is a comparison of the use of number systems in specific computing areas that are already well developed, included here to provide a context for the study of *Yorùbá* numerals in the context of a broader number system research in computer science. This has the least contribution in literature. Omachonu [22] English/German comparison could be used as a bridge point in broader numeral system. Operating systems depend on number systems for resource management and number representation is used in data science and visualization. None of the literature is specifically on *Yorùbá* numerals, but they do show that number system design is a fundamental and pervasive issue in computing. A well-designed computational representation of the *Yorùbá* numeral system might plausibly be used in a variety of technical domains, including natural language interfaces, embedded systems, and educational systems.

Table 1. Thematic Distribution of Reviewed Studies

Themes	Reviewed Corpus
Morphological and Linguistic Analyses of Yorùbá Numerals	Ekundayò [3], Zaslavsky [4], [5], Babárinde [11], Omachonu [22], Ajiboye [20], Ajíbádé [9], Ayanwoye & Aiyede [24], Adeniyi <i>et al</i> [1].
Computational and Machine Translation Systems	Isaac [2], Akínadé & Odejobí [7], Babarinde [6], Eludiora [10], Rhoda <i>et al.</i> [12] Adewole <i>et al.</i> [15], Jimoh <i>et al</i> [18], Orife [19].
Educational Technology and Language Preservation	Eludiora [8], Elesemoyo [23], Jimoh <i>et al.</i> [14], Agbeyangi [13], Asahiah [16], Asahiah <i>et al.</i> [17].
Vigesimal System and Computing Architecture Implications	Ekundayò [3], Verran [21], Ajiboye [20].
Number Systems in Broader Computing Domains	Omachonu [22]

Table 1 above presents the distribution of the reviewed corpus over the five thematic clusters presented in Section 2 and the most common method used in each theme. The distribution is relatively well-balanced in terms of the themes of morphology, computation, and education, and the latter is included mainly for comparative purposes rather than as a part of *Yorùbá* numeral studies.

3. Quantitative Analysis and Evaluation of the Reviewed Corpus

This section aims to present a more rigorous and reproducible account, and therefore a quantitative analysis of the corpus is undertaken. All values are calculated directly from the 24 studies in the reference list and their thematic assignment in Table 2, no external data are used. The studies cover decades of years and are therefore descriptive in nature and are designed to describe the structure and maturity of the evidence base, not to test hypotheses.

3.1 Corpus Profile and Temporal Trends

The corpus output in different years is shown in Fig. 1. The evidence base is heavily skewed towards recent years. 19 of 24 studies (79.2%) appeared in 2013 or later and 9 (37.5%) in 2021 or later, whereas only four studies (16.7%) predate 2010 [3], [4], [5], [21]. The number of studies per year reaches a maximum of three in 2014, 2017 and 2025, and the total number of studies triple between 5 in 2012 and 15 in 2018.

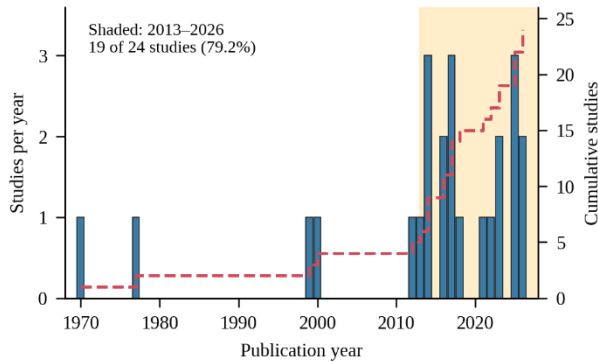


Fig. 1. Number of reviewed studies per publication year (bars, left axis) and cumulative number of studies (dashed line, right axis). The shaded region marks 2013-2026.

The corpus is broken down by theme and period in Fig. 2. The earliest period has only one cluster (morphological and linguistic analysis) with three studies prior to 2000, and there is no study in 2018-2022. Themes B and C emerge only in 2013-2017, with the earliest studies dated 2014 [7], [16]. The median publication year is 2013 for Theme A, 2017 for Theme C and 2019 for Theme B, and the computational literature is on median 6.5 years younger than the linguistic literature it builds upon; the earliest linguistic study (1970 [5]) is 44 years differ from the earliest computational study (2014 [7]). Themes D and E are small (3 and 1 studies), no study from Theme D is published after 2016 [20] and only one study from Theme E dates back to 2012 [22].

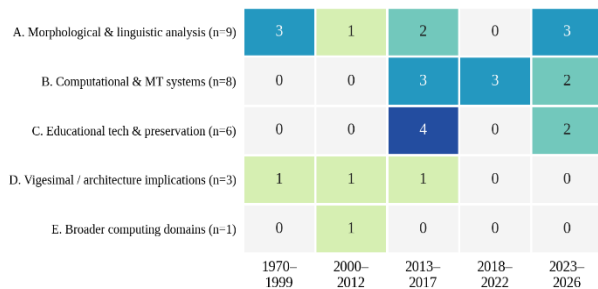


Fig. 2. Heat map of reviewed studies by thematic cluster and publication period (multi-label counts; 27 assignments over 24 studies).

The main descriptive indicators are summarised in Table 3 and the composition of the corpus by publication type and scope is represented in Fig. 3. Journal articles dominate (17 of 24; 70.8%), followed by book chapters (3; 12.5%) and preprints (2; 8.3%). Two-thirds of the corpus (16 studies; 66.7%) directly concern Yorùbá numerals, a further four studies (16.7%) concern Yorùbá NLP more broadly (diacritic restoration and an NLP survey) and are relevant as enabling technology. There are only two studies for the numerals of other Nigerian languages [10] and [12] and two comparative studies [4] and [22]. The

thinness of the cross-lingual evidence restricts the scope of comparison of methods across Nigerian languages.

Table 2. Summary Indicators for the Reviewed Corpus

Indicator	Value
Studies in corpus	24
Publication span	1970-2026
Published in 2013 or later	19 (79.2%)
Published in 2021 or later	9 (37.5%)
Journal articles	17 (70.8%)
Studies on Yorùbá numerals directly	16 (66.7%)
Studies on other Nigerian-language numerals	2 (8.3%)
Median year, Theme A / B / C	2013 / 2019.5 / 2017
System-oriented studies (Themes B, C; excl. [18])	13
With a numeric error metric	1 of 13 (7.7%)
With no evaluation procedure identified	7 of 13 (53.8%)
Reporting a shared corpus, annotation scheme or API	0 of 24

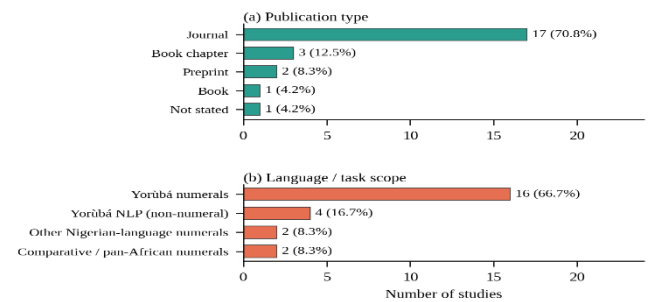


Fig. 3. Composition of the corpus by (a) publication type and (b) language or task scope (n = 24; percentages are shares of the corpus).

3.2 Audit of Evaluation Practice

How the reviewed systems were evaluated is the main methodological question that arises in a review of computational work. Out of the 14 studies in Themes B and C, one of them [18] is a secondary study (a systematic review), and 13 are system-oriented studies. These are classified in Table 3 and Fig. 4 by the strongest evaluation evidence that can be recovered from their characterisation in Section 2.

Table 3. Evaluation Evidence in the 13 System-Oriented Studies

Evidence category	Studies	n (%)
Numeric error metric reported	[19]	1 (7.7%)
Expert validation	[12]	1 (7.7%)
Learner questionnaire / engagement	[8], [14], [23]	3 (23.1%)
Case study	[13]	1 (7.7%)
No evaluation procedure identified	[2], [6], [7], [10], [15], [16], [17]	7 (53.8%)

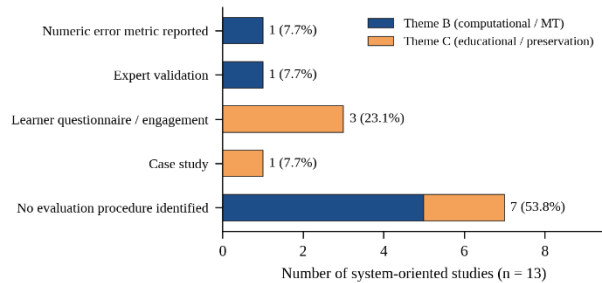


Fig. 4. Evaluation evidence in the 13 system-oriented studies, split by theme (B: computational and machine translation; C: educational technology and preservation).

Three observations follow. The first is that only one study ([19] with less than 5% numeric error rates) reports a numeric error measure, and that for an enabling task (restoration of diacritics) and not for numeral generation or numeral translation; none of the numeral conversion or numeral translation systems [2], [7], [10], [12], [15] is described with a numeric accuracy figure. Second, the evidence that does exist is not of the same type across the themes, expert validation for the computational theme [12] and learner questionnaires or engagement and retention outcomes for the educational theme [8], [14], [23] so results cannot be compared across studies even in principle. Third, over half of the system-oriented studies (7 out of 13; 53.8%) do not have an evaluation procedure identified in the synthesis, and none of the studies in the corpus report a shared corpus, annotation scheme or application programming interface. The percentages are conservative estimates and details of the evaluation are reported only in the full texts of individual studies, which may be under-represented in the review.

3.3 Evaluation against the Research Questions

The evidence base is assessed against the four research questions in Section 1 and the number of studies supporting each question, the main finding, and the maturity assessment are presented in Table 4. The assessment is qualitative with reference to the quantities reported above.

Table 4. Evidence Base and Maturity by Research Question

RQ	Evidence	Principal finding	Maturity
RQ1	Theme A: 9 studies (1970-2026)	Numerical formation follows rule-like processes (compounding, suffixation, subtraction, multiplication) [1], [3], [9]; no agreed computation-ready grammar for large numerals	Documented, but not formalised
RQ2	Theme B: 8 studies (2014-2026)	Predominantly rule-based or hybrid systems; 1 of 13 system-oriented studies report a numeric error metric	Emerging; fragmented
RQ3	Theme C: 6 studies (2014-2025)	Mobile and web prototypes report engagement and retention benefits [8], [14], [23]; tools are isolated and small-scale	Prototype stage
RQ4	All 24 studies	Computational and educational studies (14) outnumber linguistic studies (9), yet evaluation is non-comparable and no shared benchmark exists	Gap in maturity, not volume

As per Table 4, the gap found in RQ4 is not one of research volume, as Theme B and C have 14 studies, while Theme A has 9. The deficiencies are in formalization of language knowledge, comparability of assessment, and reusing of resources.

3.4 Proposed Minimum Evaluation Framework

Since there is no common basis for comparison in the audit, we suggest a minimum evaluation framework (Table 5) that future *Yorùbá* numeral systems should report. The framework divides the numeral space into range bands (1-20, 21-100, 101-1000, 1001-10000 and above 10000) because the subtractive and multiplicative structure outlined in Section 2 becomes more complex in the higher range bands, and an overall accuracy figure would mask problems with large numerals. Inter-annotator agreement should be reported in human evaluation by native speakers, and pre-test and post-test gains should be reported in educational tools, not engagement. The framework is completed with a reporting checklist that includes test-set size and construction, range bands covered, number of annotators, and release of data and code.

Table 5. Proposed Minimum Evaluation Framework for Yorùbá Numeral System

Dimension	Metric	Protocol
Correctness	Exact-match accuracy per range band	Numeral to Yorùbá text (and the reverse) on a held-out test set stratified by the five range bands
Orthography and tone	Character error rate, tone-mark accuracy	Comparison with a fully diacritized native-speaker reference
Round-trip consistency	Round-trip accuracy	Integer to Yorùbá to integer must recover the input
Context sensitivity	Acceptable-variant coverage	Share of valid alternative renderings of the same number that are accepted or produced
Human judgement	Adequacy and fluency	Native-speaker Likert ratings with inter-annotator agreement reported
Pedagogical impact	Learning gain, effect size	Controlled pre-test and post-test with a comparison group
Robustness	Error rate on malformed or out of range input	Report failure behaviour on invalid inputs

4. DISCUSSION AND RESEARCH GAPS

The 24 reviewed studies follow a path from descriptive linguistic documentation (1970s - 2010s) to increasingly computational and applied approaches to the Yorùbá numeral system (2010s - 2020s). This synthesis yields several cross-cutting observations.

First, there is imbalance between the depth of linguistic documentation and the level of development of computational tools. Although there are several studies that provide detailed morphological accounts of numeral derivation, such as Èkúndayò [3], Babárinde [11] and Ajíbádé [9], the computational systems built on these accounts, like those of Isaac [2], Akínadé and Odéjóbí [7] have generally been limited to rule sets that do not fully operationalise the richer morphological patterns documented in the linguistic literature, especially for large or compound numerals.

Second, the evaluation practices in computational and educational studies are still inconsistent and are generally small in scope. The reported evaluations are based on expert judgment, questionnaires for limited groups of users [7] and [8] [11] rather than standardized benchmark datasets, which makes it hard to compare systems or test robustness of the systems in the entire range of Yorùbá numeral expressions.

Third, unlike the machine learning and neural network approaches that currently dominate the mainstream of natural language processing, most current computational systems are

rule-based, whereas the numeral representation is readily applicable to machine learning. This indicates that modern NLP methods like sequence-to-sequence models or transformer-based architectures could be employed to enhance the robustness of numeral variety and context dependency in the numeral generation and translation task for Yorùbá.

Fourth, there are pedagogical tools for preservation that have been shown to have educational value, but are still isolated prototypes [7], [16], [18]. No study in the reviewed corpus reports a shared corpus, a standard numeral annotation scheme, or an interoperable application programming interface (API) that would enable translation systems, educational applications and linguistic research to share a common technical basis. Lastly, the vigesimal system studies [15] and [19] indicate conceptual significance of investigating other numeral systems for computation than the decimal system, which has not yet been linked to actual system implementations.

The study reveals that Yorùbá numeral is not only based on vigesimal as documented by Èkúndayò [3] but also include quinary (base 5) and decimal (base 10).

4.1 Limitations of the Reviewed Corpus

The following limitations of the literature reviewed must be recognized. The system designs and qualitative results presented in many of the computational studies reviewed here, such as [2, 5, 6] are not accompanied by quantitative accuracy measures like translation precision, recall, or word-error rate, which are common in the evaluation of machine translation systems. This renders it difficult to make absolute judgements of the effectiveness of the existing Yorùbá numeral translation systems in comparison with similar systems in higher resource languages. Furthermore, some of the studies on educational technology [7], [16], [18] have small samples of participants, usually within a single school or institution, and the results of engagement and retention reported in these studies are not generalizable to the wider population of Yorùbá speaking learners.

Another restriction is the consistency of the corpus in terms of time and terminology. The reviewed studies cover a span of almost half a century and come from a variety of disciplinary backgrounds (descriptive linguistics, ethnomathematics, software engineering, and education), which makes it difficult to directly compare findings across studies, such as when considering the description of numeral structure, e.g., "multiplicand" in Èkúndayò's [3] and "compounding unit" in Adeniyi *et al* 's [1] are of related phenomena. It would be helpful for future systematic reviews in this area to have a shared, explicit terminology agreed upon by the linguistics and computing communities who deal with Yorùbá numerals.

5. FUTURE RESEARCH DIRECTIONS

Based on the gaps identified above, four research directions are suggested. First, a standardized, machine-readable grammar of Yorùbá numeral formation is needed that formalizes the patterns of numeral formation described in [1], [3], and [9] into a computation ready specification that can be used with arbitrarily large numerals.

Second, future computational systems should go beyond the rule-based approach to translation to hybrid or data-driven systems, using annotated numerals to facilitate machine learning-based translation and generation as in [2], [5] and [6].

Third, a shared open numeral corpus and evaluation benchmark (like the standard benchmarks in mainstream NLP) would

enable systematic and reproducible comparison of *Yorùbá* (and related Hausa and Igbo) numeral processing systems [8, 11].

Fourth, there is a need for integration with a common back-end numeral processing translation and validation service, so that the improvements in the accuracy of numeral processing can be cascaded down to the downstream applications such as those in the field of education and preservation as in [7], [16] and [18].

Lastly, there is a need to integrate positional number for *Yorùbá* numerals to enhance familiarity and computational usage as used in decimal.

Table 6. Phased Research Roadmap and Success Indicators

Phase	Direction	Success indicator
Short term	Formal numeral grammar, shared annotation scheme and reporting checklist	Grammar validated against native-speaker judgements in every range band
Short term	Open benchmark corpus stratified by range band	Public release, independent systems evaluated on the same benchmark
Medium term	Hybrid and neural numeral generation and translation, including tone-aware models	Outperform rule-based baselines in every range band on Table 5 metrics
Medium term	Shared numeral-processing API integrated into learning applications	Existing prototypes [8], [13], [14], [23] consume the same service
Long term	Extension to Hausa, Igbo, other Nigerian languages and other vigesimal systems	Cross-lingual results reported under the same protocol
Long term	Positional (abacus-style) representation and speech interfaces for numerals	Round-trip consistency between spoken, written and positional forms

5.1 THREATS TO VALIDITY

Like all literature reviews, there are a number of threats to validity for the present study. The comparative studies of number systems in mainstream computing areas (Section 2) [12] - [14], [20] - [24] were chosen to give the reader background information, but not to directly contribute to the findings of this work on *Yorùbá* numerals, the reader is advised to take this section as background and not as a contribution to the findings of this paper. The internal validity threat is that the methodological details reported in the corpus are not uniform, some studies, especially older ones like [3] and [4] do not report the details of experiments as they are now, so some elements of their methodology had to be inferred from the problem statements and results reported.

The external validity threat relates to the generalizability of the thematic clusters identified here to other vigesimal or non-decimal numeral systems in other indigenous or low-resource languages beyond *Yorùbá*, Hausa, and Igbo, and the patterns observed (rich linguistic documentation outpacing computational maturity) are plausible for other vigesimal or non-decimal numeral systems, but this review does not empirically test the generalizability of these patterns. Lastly, a conclusion validity threat is the failure to have standardized

quantitative outcome measures across studies in the underlying corpus, which limited this review to qualitative, rather than meta-analytic, synthesis of reported results.

6. CONCLUSION AND FUTURE WORK

This paper summarized 24 studies on the linguistic structure, computational modelling and educational applications of the *Yorùbá* numeral system and added a quantitative analysis and an assessment of the evidence base to the thematic synthesis. In the case of RQ1, the morphological and derivational structure of *Yorùbá* numerals is well documented (9 studies), although there is no agreement on a grammar that can be used for computation. In terms of RQ2, the computational systems that translate, generate and convert numerals are primarily rule-based or hybrid and are disjointed; only one of thirteen system-oriented studies provides a numeric error metric, and 7 do not have an identifiable evaluation procedure. With respect to RQ3, the findings indicate that although the tools are successful in engaging and retaining learners, they are still prototypes with limited applicability. For RQ4, the difference between documentation and tooling is not in quantity, but in maturity. 79.2% of the corpus has been published since 2013 and computational and educational studies (14) now outnumber linguistic studies (9), but results cannot be compared due to the heterogeneity of the evaluation and the lack of a shared benchmark, annotation scheme or interface. The main practical results of this paper are the audit of the evaluation practice (Table 3, Fig. 4), the minimum evaluation framework (Table 5) and the phased research roadmap (Table 6). A common evaluation protocol would enable rule-based, hybrid and neural approaches to be compared fairly within numeral range bands and provide a foundation for assessing pedagogical impact for educators and developers. There are three horizons for future work. In the near future, the research community should establish a formal numeral grammar, a common annotation scheme and an open range-stratified benchmark corpus. In the medium term, the models should be trained and evaluated against the rule-based baselines on this benchmark and exposed through a common numeral-processing service that current mobile and web applications can use. In the long term, the framework should be extended to other Nigerian languages such as Hausa and Igbo and to other vigesimal numeral systems, to speech technologies (text-to-speech and automatic speech recognition for numerals), and to positional or abacus-style representations that make the subtractive structure of *Yorùbá* numerals explicit and computable. Longitudinal studies of learning outcomes and studies of open resources, controlled by the community, are also required to ensure that the technical progress will lead to the long-term digital preservation and accessibility of the *Yorùbá* number system.

7. REFERENCES

- [1] Adeniyi S. K., *et al.* (2025), A morphological and computational analysis of *Yorùbá* numeral multiplicands: Linguistic structure and cultural significance.
- [2] Isaac, O. (2022). Machine translation system for numeral in English text to *Yorùbá* language. *Ife Journal of Information. Communication. Technology*, 6, 26-37.
- [3] Èkúndáyò S. A. (1977), Vigesimal numeral derivational morphology: *Yorùbá* grammatical competence epitomized, *Anthropological Linguistics*, 19(9):436-453, <http://www.jstor.org/stable/30027551>
- [4] Zaslavsky C. (1999), *Africa Counts: Number and Pattern in African Cultures*, Lawrence Hill Books, 3rd edition.



- [5] Zaslavsky, C. (1970). Mathematics of the Yorùbá people and of their neighbors in southern Nigeria. *The Two-Year College Mathematics Journal*, 1(2), 76-99.
- [6] Babarinde, O. (2014). Linguistic analysis of the structure of Yorùbá numerals. *Language Matters*, 45(1), 127-147.
- [7] Akínadé, O. O., & Odéjobí, O. A. (2014). Computational modelling of Yorùbá numerals in a number-to-text conversion system. *Journal of Language Modelling*, 2(1), 167-211.
- [8] Eludiora, S. (2017). Development of a Yorùbá Arithmetic Multimedia Learning System. *Universal Journal of Educational Research*, 5(5), 862-873.
- [9] Ajibádé, M. A. (2023). A construction morphology approach to Yorùbá numerals. *Constructions*, 15(2).
- [10] Eludiora, S. I., & Abubakar, M. A. (2021). A Hindu-Arabic to Hausa number transcription system. *Malaysian Journal of Computing (MJoC)*, 6(1), 727-744.
- [11] Babárindé O. (2013), "The linguistic analysis of the structure of the Yorùbá numerals," *Journal of Literature, Languages and Linguistics*, Vol 1, 78-88.
- [12] Rhoda, Í. A. (2017). Computational analysis of Igbo numerals in a number-to-text conversion system. *Journal of Computer and Education Research*, 5(10), 241-25.
- [13] Agbeyangi, A. O., Eludiora, S. I., & Popoola, O. A. (2016). Web-based Yoruba numeral translation system. *IAES International Journal of Artificial Intelligence (IJ-AI)*, 5(4), 127-134.
- [14] Jimoh, E. R., Akinlolu, A. O., Akanbi, M. B., & Olojede, O. A. (2023). An android-based Yorùbá numeral translation system. *FUOYE Journal of Pure and Applied Sciences (FJPAS)*, 8(3), 38-46.
- [15] Adewole, L. B., Adewole, V. T., Owofe, M. C., Mese, F. M., & Abiola, O. B. (2026). Development of a rule-based Arabic numerals to Yorùbá translator. *ABUAD International Journal of Natural and Applied Sciences*, 6(1), 60-69.
- [16] Asahiah, F. O. (2014). Development of a Standard Yorùbá digital text automatic diacritic restoration system. *Signature*, 22(02).
- [17] Asahiah, F. O., Odejobi, O. A., & Adagunodo, E. R. (2017). Restoring tone-marks in standard Yorùbá electronic text: improved model. *Computer Science*, 18(3).
- [18] Jimoh, T. A., De Wille, T., & Nikolov, N. S. (2025). Bridging Gaps in Natural Language Processing for Yorùbá: A Systematic Review of a Decade of Progress and Prospects. *arXiv preprint arXiv:2502.17364*.
- [19] Orife, I. (2018). Attentive Sequence-to-Sequence Learning for Diacritic Restoration of Yorùbá Language Text. *arXiv preprint arXiv:1804.00832*.
- [20] Ajiboye, O. (2016). The Yorùbá Numeral System¹. *The Numeral Systems of Nigerian Language*, 1.
- [21] Verran, H. (2000). Accounting mathematics in West Africa: some stories of Yoruba number. In *Mathematics across cultures: The history of non-western mathematics* (pp. 345-371). Dordrecht: Springer Netherlands.
- [22] Omachonu, G. S. (2012). Comparative analysis of the numeral systems of Igala, Yoruba, German and English. *Linguistik online*, 55(5), 57-73.
- [23] Elesemoyo, I. O. (2025). An Android-Based Number Conversion System to Yorùbá Language Numeral Forms. *FUOYE Journal of Pure and Applied Sciences (FJPAS)*, 10(1), 102-113.
- [24] Ayanwoye, O. K., & Aiyede, E. I. (2026). Numeral Grammar: Structure and Cognition of Yoruba Numeral-Preserving African Identity and History. In *Preserving Identity and History with African Languages* (pp. 321-350). IGI Global Scientific Publishing.