

SURVEY

IsiXhosa in the Digital Age: Navigating Language Preservation and Innovation: A Systematic Scoping Review

ABAYOMI AGBEYANGI^{ID} AND NOBERT JERE^{ID}

Department of Business and Application Development, Walter Sisulu University, Buffalo City Campus, East London 5200, South Africa

Corresponding author: Abayomi Agbeyangi (aagbeyangi@wsu.ac.za)

ABSTRACT Preserving and innovating indigenous languages is crucial for maintaining cultural heritage and facilitating community development in this digital age. Understanding the digital landscape of IsiXhosa language digitization efforts, however, remains underexplored. This systematic scoping review aims to map the existing literature on digitalization efforts for IsiXhosa language preservation and innovation. A scoping review was conducted, guided by PRISMA-ScR, on relevant literature pertaining to isiXhosa digitization, preservation, and innovations. Two databases were searched (Scopus and Web of Science), and additional articles were identified through a grey literature search (Google Scholar) to identify other relevant literature. Data were extracted in terms of title, year, country, language, digitization efforts, and summary of main findings and results related to isiXhosa language preservation and innovation. A total of 85 unique articles were included from 479 records, leading to the identification of five themes under the digitization efforts. Most studies were conducted in South Africa, accounting for about 78% of the total articles. As such, significant efforts were identified in grammar and morphology, speech recognition, and machine translation. There are unresolved gaps and challenges (such as large and high-quality datasets) that must be addressed. Nevertheless, the efforts demonstrate an increased acceptance of the significance of protecting and improving indigenous African languages, such as isiXhosa, in this era of digital technology.

INDEX TERMS Digitization, IsiXhosa language, IsiXhosa digitization, language preservation, language innovation, Xhosa.

I. INTRODUCTION

Indigenous languages have a vital role in preserving cultural history and are essential for passing down knowledge, identity, and values from one generation to another [1], [2]. Stories, traditions, and beliefs are transmitted from ancestors to future generations through the complex network of language, establishing robust links between the past, present, and future. By protecting indigenous languages, we not only protect the origins of varied cultures but also guarantee that their distinct viewpoints and knowledge will continue to enhance the globe for future generations. Language serves as the medium through which the fundamental nature of a community is conveyed, resonating throughout history as evidence of strength, ingenuity, and the enduring determination of mankind. In an earlier study focusing on

the speech community, Gumperz [3] expressed the viewpoint that language is the most potent and adaptable means of communication, and it is present in all known human societies. Unlike other methods of communication, the verbal system has the ability to refer to a wide range of objects and concepts by refining its grammatical and semantic structure.

In this digital era, the convergence of language preservation and technology has gained significant prominence, presenting both unparalleled prospects and daunting obstacles [4]. IsiXhosa, as one of the South Africa's eleven official languages [5], encompasses this intricate interplay. Prominent and widely spoken in the Eastern Cape region of South Africa, IsiXhosa faces the twin challenge of keeping its traditional roots while embracing current technical improvements. As attempts are made to digitize and promote the language, it is vital to establish a balance between innovation and legacy to ensure its survival for future generations. The rich linguistic past and cultural value of the language

The associate editor coordinating the review of this manuscript and approving it for publication was Vivek Kumar Sehgal^{ID}.

emphasise the need to ensure its ongoing life in the digital realm.

One way to think of digitalization as a kind of standardised language translation is to consider it as the binary coding of knowledge, information, and social processes, as defined by Turner [8]. In particular, Turner [8] asserted that since our ability to conceptualise is influenced by the knowledge and skills we acquire through digital methods, digitization can be understood as an epistemic and ontological shift. Hence, it can be regarded not just as a technical process but also as a paradigm shift in our perception of and engagement with everything around us. This shift in perspective enables novel approaches to thinking, problem-solving, and information exchange, ultimately influencing how we adapt to the rapidly changing terrain of the digital era. By adopting digitalization as a means of language translation, we can more effectively comprehend its ability to close divides, facilitate interactions, and improve our collective perception of the world.

Lately, there have been efforts in the study of IsiXhosa that have concentrated on creating digital tools and platforms with the aim of advancing language preservation and revitalization [6], [7], [8]. These efforts seek to enable IsiXhosa speakers to actively participate in their language through novel and creative methods, guaranteeing its significance and durability in contemporary society. Fundamentally, we can understand the importance of preserving indigenous languages through their immense impact on shaping cultural heritage and fostering a feeling of inclusion. This encapsulates the narratives of victories and challenges, the wisdom of the older generation, and the principles that have supported communities through centuries of transformation [9]. Thus, by understanding the need to preserve and revive our language, we enable future generations to continue the legacy of their forebears while also honoring the accomplishments of the past. This helps us maintain the linguistic diversity of the language and safeguard its distinct worldviews to avoid the danger of disappearing over time.

Global initiatives to digitise languages have resulted in the documentation and online accessibility of numerous languages for a worldwide audience. The World Economic Forum's 2021 report [10] assessed the global digital competitiveness of 140 countries based on mindset and digital ecosystem dimensions, as well as the change in ranks between 2018 and 2020. According to the report, China, followed by Saudi Arabia and Brazil, were identified as the top three countries that experienced significant improvement in digital competitiveness among the G20 nations. The efforts towards African languages have not matched the significance and immediacy of safeguarding these varied linguistic legacies [11]. Groups such as WikiAfrica¹ have taken advantage of this to help disseminate knowledge in many African languages that are not widely used online. Hence, publishing written and spoken content online in a given language contributes to the corpus of publicly

available information in the language. Maseko [5] submits that indigenous African languages will remain sentimental in their usage, serving as languages of socio-cultural importance within the home and limited to primary schooling. However, Maseko suggests that meaningful and deliberate intervention is necessary to accelerate the process of intellectualising these languages within the framework of national development initiatives. It is imperative that African languages be given the same attention and resources as other highly-resourced languages worldwide. Specifically, African languages are still being categorised as low-resource languages.

Based on the data from Web Technology Surveys² (a web technology survey service that provides information about the usage of various web technologies on the internet), the English language is utilised by 50.8% of all websites, while IsiXhosa is only used by approximately 0.1%. Specifically, it was asserted by Turner [8] under the subject of decolonization through digitalization that digitising isiXhosa entails expanding and diversifying its online presence to establish its complete validity as a South African language across all social domains.

This systematic scoping review aims to investigate the efforts made to digitalize the IsiXhosa language. The review seeks to identify and analyse existing literature as well as new trends, obstacles, and gaps that are impeding progress in IsiXhosa language digitization. This review offers an in-depth analysis of the current status by combining insights from many sources in the literature, focusing on initiatives, academic research, and technological developments. The main goal is to provide information and support strategic actions that could promote collaboration and encourage new methods to empower the IsiXhosa language to succeed in the digital era while preserving its cultural heritage.

The review's selection of isiXhosa is indicative of its significance as one of South Africa's recognised languages (the second highest spoken language)³ and as a model for other African languages and many other languages around the world. The review further seeks to bring attention to the broader concerns facing African languages in the digital era, emphasising progress, obstacles, and prospective opportunities. The goal is to encourage the development of other indigenous languages across the continent by emphasizing IsiXhosa, a language that holds great historical and cultural importance. The review is guided by the following research questions: (1) What is known from the literature about digitalization efforts implemented to preserve and promote the IsiXhosa language and culture? (2) What are the gaps in digitalizing IsiXhosa, including technological, socio-cultural, and linguistic factors?

II. METHOD

A. TRANSPARENCY STATEMENT

The systematic scoping review was conducted in accordance with the PRISMA-ScR guidelines (Preferred Reporting Items

¹<https://www.wikiafrica.org/>

²<https://w3techs.com/>

³<https://southafrica-info.com/arts-culture/11-languages-south-africa/>

TABLE 1. Eligibility criteria.

Criterion	Inclusion	Exclusion	Justification
Population	Studies focusing on IsiXhosa-speaking communities or initiatives related to IsiXhosa language	Studies unrelated to IsiXhosa language	Ensures relevance to the target population and alignment with the scope of the review
Concept	Studies examining digitalisation efforts specifically aimed at preserving and innovating IsiXhosa	Studies unrelated to digitalisation	Aligns with the primary objectives of the review and ensures a focused exploration of relevant initiatives
Context	Studies conducted in South Africa or other regions with a focus on isiXhosa digitization.	Studies from unrelated contexts	Ensures the findings are applicable to IsiXhosa and enhances the contextual relevance of the review
Language	Studies available in either English or IsiXhosa	Studies not available in English or IsiXhosa	Ensures accessibility to the target audience and facilitates comprehensive coverage of relevant literature
Geographic area	Studies conducted in South Africa or other regions with a focus on isiXhosa digitization	Studies not related to isiXhosa from any regions	Enhances the contextual relevance of findings to IsiXhosa-speaking communities and ensures alignment with the scope of the review
Time	Studies published from January 2004 to the present	Studies published before January 2004	Enables a comprehensive analysis of digitisation initiatives in a modern context while incorporating recent work
Type of source	Peer-reviewed journal articles, conference papers, and grey literature	Studies from unreliable sources	Provides diverse perspectives and insights, ensuring methodological rigor and relevance to the research goals

for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews). The Scopus and Web of Science databases were searched for relevant literature. A grey literature search was also conducted in Google Scholar to identify other relevant literature. The paper is an accurate, concise, and unambiguous representation of the study.

B. SEARCH STRATEGY

The search was conducted from March 16 to March 21, 2024. The focus was on literature published between 2004 and 2024 (the last two decades) to ensure relevance to current digitalization and preservation initiatives. Scopus and Web of Science were selected due to their extensive coverage of peer-reviewed literature across various fields. Google Scholar was incorporated to obtain a wider variety of literature, including conference papers, theses, unindexed items in Web of Science and Scopus, and grey literature. The search parameters were formulated based on available literature and expert guidance to ensure the search’s thoroughness and pertinence. The search terms used are the following:

- (“isixhosa” OR “xhosa”) AND language OR digital OR initiatives OR preservation OR digitisation) - Scopus
- ((“isixhosa” OR “xhosa”) OR language digital initiatives, preservation and digitisation) - Web of Science and Google Scholar

The terms “isiXhosa” and “Xhosa” are used interchangeably in the search; as *Xhosas* call their language “isiXhosa”, it is usually referred to as “Xhosa” in English and is sometimes interchanged in the literature. Duplicates were eliminated from the search results to ensure the accuracy and reliability of the findings. The Google Scholar search only includes the first 100 entries to guarantee relevancy to the topic of the review.

C. SEARCH ELIGIBILITY CRITERIA

The studies included in this systematic scoping review have to meet particular criteria to ensure their relevance and completeness. First, the article needed to be available in either English or IsiXhosa, which are the main languages of interest for this review. Also, only research published between January 2004 and the present (2024) was considered. This allowed for a comprehensive analysis of digitization and preservation efforts in a modern context while incorporating recent research. Furthermore, the content focused on digitalization efforts explicitly aimed at conserving and advancing the IsiXhosa language and culture. This aligns with the main goals of this review. Finally, a wide range of literature was incorporated, including peer-reviewed journal articles, conference papers, and grey literature (theses), to present various viewpoints and valuable insights. Studies conducted in South Africa and other regions were included, focusing on isiXhosa digitization efforts.

The criteria were chosen to enable a thorough investigation of efforts to digitise and preserve the IsiXhosa language while also ensuring that the research methods were rigorous and relevant to the research goal. Additional information on this can be found in Table 1.

D. DATA EXTRACTION

The data extraction process follows the specified eligibility criteria to guarantee the inclusion of relevant research studies. These criteria determine which papers are eligible for inclusion and ensure that the review remains focused and relevant. In addition, the criteria establish a structure for the methodical selection and retrieval of information from the literature, improving the correctness and accuracy of the review process. The entries for the form in Table 2 were pilot-tested using three studies to extract the information. One reviewer carried out the extraction procedure, while a second reviewer verified it. Discrepancies or disagreements

TABLE 2. Key information extracted from the searched sources.

Publication Details	Title	Corres. Author	Publication Year	Source Type	Digitalisation Efforts	Key Findings	Language	Geographic Area	Relevance
Article 1	Using Linked Data Techniques for Creating an IsiXhosa Lexical Resource-a Collaborative Approach	Eckart, T.	2018	Conference Proceedings	Proposes collaborative approach for creating an isiXhosa lexical resource using Linked Data techniques	Linked Data allows for creating a more interoperable and expandable isiXhosa lexical resource.	English	Germany	Yes
Article 2	A Xhosa language translation of the CORE-OM using South African university student samples	Erasmus, T.	2017	Journal	Focuses on translation of CORE-OM (psychological assessment tool) into isiXhosa for universities	Successful translation highlights potential of digitisation for isiXhosa mental health resources.	English	South Africa	Yes
Article 3	Enriching and Increasing the Usability of Lexicographical Data for Less-Resourced Languages	Goldhahn, D.	2020	Conference Proceedings	Focuses on using crowdsourcing to enrich existing lexicographic data for under-resourced Bantu languages spoken in southern Africa.	Crowdsourcing can be a valuable tool for enriching digital lexicographic resources for less-resourced languages.	English	Germany	Yes

encountered during the extraction were resolved through discussion and consensus. Key information extracted from each study included:

- Study title
- Name of author(s).
- Year of publication
- Publication type
- Name of Journal/book or conference
- Description of digitalisation initiatives related to IsiXhosa language preservation and innovation
- Summary of main findings/results related to IsiXhosa language preservation and innovation
- Language(s) of the study
- Geographic area(s) covered in the study
- Time period covered in the study
- Alignment with the objectives and scope of the review.

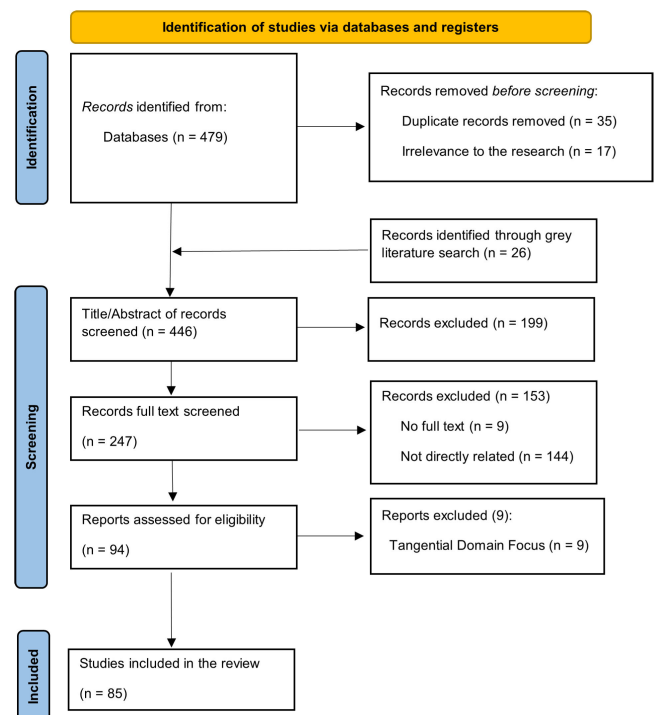
E. DATA SYNTHESIS

After extracting the data from the selected studies, an interpretive approach [12] was used. At first, the studies were classified according to their sources (Scopus, Web of Science, and Google Scholar). The categories were merged into a single entity for analysis. This merging enables an appropriate presentation of the results from a single data source. The process of synthesising the data was carried out and structured using MS Excel.

III. RESULTS

A. SEARCH RESULTS AND DEMOGRAPHIC CHARACTERISTICS

The search across bibliographical databases and grey literature yielded a vast amount of material on IsiXhosa language preservation. Scopus and Web of Science provided initial sets of 583 and 1250 records, respectively. This was filtered based on publication dates (2004 to the present), language (English or IsiXhosa), and keywords associated with language preservation and IsiXhosa. The targeted

**FIGURE 1.** Preferred reporting items for systematic reviews and meta-analyses (PRISMA) flow diagram selected.

search of grey literature on Google Scholar also offers a wider search, identifying roughly 12,000 potentially relevant articles as additional documents. This was also reduced to the first 100 entries to identify only relevant studies. Combining all the sources resulted in 505 records. However, 52 articles that are irrelevant ($n = 17$) and duplicates ($n = 35$) were excluded, reducing the total to 446. Figure 1 shows the records and the flow diagram for the review.

The screening stage involved a meticulous review of titles and abstracts based on a stricter criterion: digitalization efforts explicitly aimed at conserving and advancing the

TABLE 3. Demographic characteristics of sources.

	Sources mentioning isiXhosa digitisation efforts (n=94)	Sources used for the digitisation efforts (n=85)
Nationality of corresponding authors (top 3)	South africa - 74 Germany - 6 USA - 3 Norway - 2 Sweden - 2 Belgium 1 China - 1 Egypt - 1 France - 1 Japan - 1 Netherlands - 1 United kingdom - 1	South africa - 66 Germany - 5 USA - 3 Norway - 2 Sweden - 2 Belgium 1 China - 1 Egypt - 1 France - 1 Japan - 1 Netherlands - 1 United kingdom - 1
Publication year	2005 - 1 2006 - 1 2007 - 2 2008 - 5 2009 - 1 2010 - 3 2011 - 2 2012 - 1 2014 - 3 2015 - 4 2016 - 6 2017 - 3 2018 - 9 2019 - 17 2020 - 11 2021 - 10 2022 - 9 2023 - 5 2024 - 1	2005 - 1 2006 - 1 2007 - 1 2008 - 4 2009 - 1 2010 - 3 2011 - 2 2012 - 1 2014 - 3 2016 - 6 2015 - 4 2017 - 3 2018 - 9 2019 - 15 2020 - 10 2021 - 10 2022 - 7 2023 - 3 2024 - 1
Source type	Journal article - 57 Conference paper - 32 Book chapter - 1 Thesis - 2 Data paper - 2	Journal article - 49 Conference paper - 31 Book chapter - 1 Thesis - 2 Data paper - 2

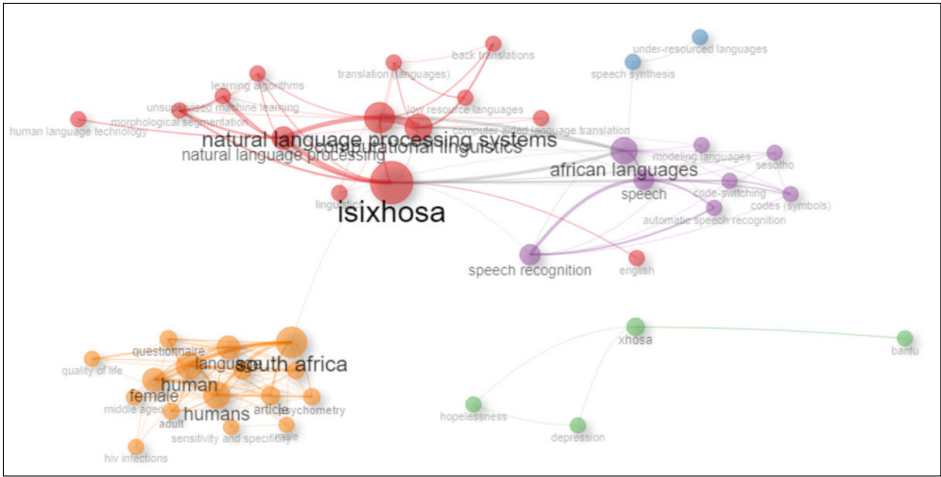


FIGURE 2. Bibliometric Map of Publications on “IsiXhosa digitisation” (tools: rstudio, method: keyword Co-occurrenceNetwork).

IsiXhosa language and culture. This in-depth analysis resulted in the selection of 247 articles for full-text examination. In the full-text examination of these 247 articles, 153 were excluded based on no access to full-text (n = 9) and those whose focus fell outside the core areas of IsiXhosa digitization (n = 144).

The final screening for eligibility excluded articles that have a tangential domain focus (n = 9), as they addressed related topics but did not directly contribute to digital preservation efforts. It’s important to note that after the screening, only two papers written in isiXhosa itself were included. The final screening process for eligibility yielded a

collection of 85 articles that formed the basis for the thematic analysis of the review. These studies were published in the last two decades (between 2004 and 2024) with corresponding authors from 12 countries. This included journal articles, conference papers, book chapters, theses, and data papers. The majority, 49 out of 85, were journal articles, accounting for 57.65%. Table 3 presents the breakdown.

B. KEYWORDS ANALYSIS

The analysis of the keywords was conducted using the software RStudio biblioshiny package to determine the frequency of term usage in relation to each other. Fig. 2 presents the visualisation of the results, enabling us to identify the thematic area of interest in the digitization efforts.

The constructed map highlights four areas describing different aspects of the isiXhosa language digitization, innovation, and preservation efforts. The first theme identified from the cluster is lexicon development and digitization; the second is grammar and morphology analysis; the third is speech synthesis technologies; the fourth is machine translation; and the fifth is digitization and language preservation tools. Keywords such as “lexicon”, “digitization”, and entries related to “human language technology” and “natural language processing systems” represent the theme of lexicon development and digitization in the network. Within the network, keywords such as “morphological segmentation”, “linguistics”, and “computational linguistics” correspond to the theme of grammar and morphology analysis. Speech synthesis technologies correspond to the keywords “speech”, “speech recognition”, “speech synthesis”, and “automatic speech recognition”. The keywords “machine translation”, “translation (languages)”, “back translation”, and “low resource languages” reflect the machine translation theme. The digitization and language preservation theme corresponds to keywords like “digitization”, “language preservation”, “cultural adaptation”, and various health-related terms in the network.

C. ISIXHOSA DIGITIZATION, INNOVATION, AND LANGUAGE PRESERVATION IN THE DIGITAL AGE

The 85 articles examined explore various aspects of IsiXhosa digitization, innovation, and language preservation. Five main themes emerged, underscoring current efforts to ensure IsiXhosa’s preservation in the digital era:

- Lexicon Development and Digitization
- Grammar and Morphology Analysis
- Speech Synthesis Technologies
- Machine Translation
- Digitization and Language Preservation Tools

1) LEXICON DEVELOPMENT AND DIGITIZATION

Language is a dynamic entity, constantly evolving and adapting to the needs of its users. However, as the world increasingly relies on digital technology, it becomes crucial to preserve and document languages in a manner that is easily accessible and sustainable [8]. In this context, lexicon development and digitalization activities are critical. Imagine a

linguistic system without a lexicon, a collection of synonyms, or any easily accessible repository of its lexemes. Acquiring, teaching, and even employing such a language would be extremely difficult. Lexicon creation and digitization narrow this divide by constructing extensive databases of terms, their definitions, and their usage. These digital resources serve as the basis for language learning materials, machine translation tools, and the creation of speech recognition software. By carefully recording a language’s words and phrases, the process of developing and digitising its lexicon ensures that it remains vibrant and easily accessible in the digital era. This applies not only to IsiXhosa but to any language that aims to flourish in the modern age.

The articles in this category (Table 4) explore diverse approaches to lexicon development and digitization, ranging from manual compilation to automatic extraction from textual sources (see Table 4). Thus, they provide an in-depth overview of several efforts aimed at improving the language’s linguistic resources and digital presence. Du Toit and Puttkammer [13] work on generating annotated corpora and advancing machine-learned linguistic technologies, which are required for advanced language processing tools. Also, Gaustad and Puttkammer [53] make a valuable contribution by offering a linguistically annotated dataset, which assists in advancing Natural Language Processing (NLP) research specifically for isiXhosa. van der Westhuizen and Niesler [14] created a speech corpus from South African soap operas, which provides important data for the study of code-switching and word usage. In the field of specialized lexicons, Alberts [29] creates bilingual and multilingual political dictionaries. These dictionaries help to improve communication in political contexts by establishing standardised political terminology in isiXhosa. Gillis-Webber [32] converts the English-Xhosa Dictionary for Nurses into a flexible digital tool that facilitates the spread of medical terminology. Packham and Suleman [33] employ an innovative strategy by utilizing crowdsourced translations via an online game, highlighting a resourceful technique for collecting multilingual content.

Mzamo et al. [50] highlight the advancement of computational tools by designing a probabilistic graphical lemmatizer. This tool improves the accuracy of isiXhosa’s identification of lemmas. Eckart et al. [87] employ linked data techniques to synchronise multilingual data, hence improving the compatibility of isiXhosa lexical resources. In a similar manner, Eckart et al. [91] developed a lexical resource in RDF format that is both highly interoperable and machine-processable. This resource makes a valuable contribution to the future advancement of language tools and applications. Deumert and Masinyana’s [57] examination of multilingual communication patterns, which offers insights into the use of isiXhosa digital vocabulary in text messages, is another notable contribution. Parker et al. [63] translated and validated the Brief Pain Inventory into isiXhosa, thereby improving the language’s medical vocabulary and digital healthcare resources. Bosch et al. [66] compile and employ

TABLE 4. Efforts on IsiXhosa lexicon development and digitization.

Source	Publication Year	Digitization Effort	Summary/Key Findings
du Toit and Puttkammer [13]	2021	The curation and annotation of corpora for isiXhosa and other Nguni languages, the development of machine-learned core linguistic technologies such as lemmatizers, part-of-speech taggers, and morphological analyzers, and the creation of easily accessible linguistic resources that are publicly available through a local resource agency.	The paper contributes to isiXhosa lexicon development and digitization by providing annotated corpora and machine-learned core linguistic technologies specifically tailored for isiXhosa, including lemmatization and morphological analysis, which can enhance the quality and accessibility of lexical resources for the language.
van der Westhuizen and Niesler [14]	2018	the compilation of a speech corpus from South African soap operas, where digital video recordings were used to extract mono audio samples for transcription and annotation.	The study contributes to isiXhosa lexicon development and digitization by providing a dataset that includes code-switched speech involving isiXhosa, enabling researchers to analyze the language interactions, code-switching patterns, and vocabulary usage in isiXhosa within a multilingual context.
Mgijima [23]	2021	It focuses on the use of translanguaging techniques to improve text prediction skills for multilingual learners.	The study contributes to the development of the isiXhosa lexicon by exploring the use of translanguaging techniques to enhance text prediction skills for Grade 4 bilingual isiXhosa and English learners.
Alberts [29]	2024	It focuses on the compilation of bilingual and multilingual political dictionaries	The paper contributes to isiXhosa lexicon development and digitization by defining and translating core political terms into isiXhosa as part of the broader effort to promote multilingualism and enhance communication in the political and related fields within Southern Africa.
Gillis-Webber [32]	2018	Converting the English-Xhosa Dictionary for Nurses into a Linguistic Linked Data framework.	The Study contributes to isiXhosa lexicon development and digitization by digitizing the English-Xhosa Dictionary for Nurses, thereby transforming it into a bidirectional and diachronic resource that can potentially aid in the development of future language resources for isiXhosa.
Packham and Suleman [33]	2015	It involves crowdsourcing translations through an online competitive game to gather original and relevant multilingual content.	The paper contributes to isiXhosa Lexicon Development and Digitization by proposing a novel approach to sourcing original and relevant multilingual content through crowdsourcing translations.
Steele and Edwards [36]	2008	It involves translating and validating the Beck inventories into Xhosa.	The paper contributes to isiXhosa lexicon development and digitization by highlighting the challenges and strategies involved in translating and validating psychological inventories into Xhosa, thereby expanding the lexicon and digitising mental health assessment tools in the isiXhosa language.
Marquard and Suleman [47]	2023	The use of focused web crawling to automatically generate a corpus for IsiXhosa.	The study contributes to IsiXhosa lexicon development and digitization by showcasing how focused web crawling can be utilised to automatically build a corpus for the language.
Maseko [5]	2011	Focused on developing terminologies and glossaries in isiXhosa	The study contributes to isiXhosa lexicon development and digitization by advocating for the intellectualization of African languages, particularly isiXhosa, through the creation of terminologies and glossaries in various disciplines to support education and language advancement in South Africa.
Mzamo et al. [50]	2015	The development of XGL, a lexicalised probabilistic graphical lemmatiser for isiXhosa	The paper's contribution to isiXhosa lexicon development and digitization lies in the creation of XGL, a machine learning-based lemmatiser that enhances the accuracy of identifying lemmas from word forms in isiXhosa, thereby advancing the digitization and development of the language's lexicon.
Eiselen and Bukula [52]	2022	The development of Named Entity Recognition (NER) resources for the isiXhosa language, including annotated data sets and a NER system	The study contributes to isiXhosa lexicon development and digitization by providing a detailed linguistic analysis of named entity annotated data for the isiXhosa language, identifying morphosyntactic features relevant to named entities, and developing a conditional random field-based named entity recognizer to enhance the digitization and processing of isiXhosa text data.

TABLE 4. (Continued.) Efforts on IsiXhosa lexicon development and digitization.

Gaustad and Puttkammer [53]	2022	The creation of a linguistically annotated dataset for four South African languages to facilitate language-specific and cross-language development and evaluation of Natural Language Processing technologies.	The study contributes to isiXhosa lexicon development and digitization by providing a linguistically annotated dataset for isiXhosa, which includes annotations for morphology, part-of-speech, and lemmas.
Deumert and Masinyana [57]	2008	The analysis of the language choices and communication patterns of bilingual South African users in text messages (SMS).	The study contributes to isiXhosa lexicon development and digitization by examining how bilingual South African users incorporate English and isiXhosa in their text messages, shedding light on the linguistic choices, norms, and cultural content present in digital communication, thus providing insights into the evolving lexicon and digitization of isiXhosa in a mobile communication context.
Parker et al. [63]	2015	It involves translating and validating the Brief Pain Inventory into BPI-Xhosa.	By translating a widely used pain assessment tool into isiXhosa and validating it for a specific population, the study enhances the lexicon of medical terminology in isiXhosa and contributes to the digitization of healthcare assessment tools for the Xhosa-speaking community.
Bosch et al. [66]	2018	The preparation and utilisation of Xhosa lexicographical data in a multilingual, federated environment using a dedicated OWL ontology.	The study contributes to isiXhosa lexicon development and digitization by presenting a new source of lexicographical data for Xhosa, modelling it using an OWL ontology for Bantu languages, and emphasising the importance of open source data for enhancing language processing capabilities and promoting resource reuse in a multilingual, federated environment.
Klerk [67]	2005	It involves analysing the usage of the pragmatic marker "well" in a large corpus of Xhosa English.	The study contributes to isiXhosa lexicon development and digitization by examining the use of the discourse marker "well" in a corpus of Xhosa English, shedding light on how non-native speakers incorporate and use this marker in their language use.
Mostert [68]	2020	It involves using interlinear morphemic glossing to present mathematical examples in a language different from the language of publication, facilitating cross-linguistic understanding in mathematics education research.	The study contributes to isiXhosa lexicon development and digitization by exploring the challenges and strategies for presenting mathematical concepts in isiXhosa, particularly through the use of interlinear morphemic glossing to bridge language barriers in mathematics education research.
Grabar et al. [71]	2014	It involves utilising natural language processing methods and tools to automatically detect and categorise medical words.	The study contributes to isiXhosa lexicon development and digitization by analysing medical terminology in Xhosa, building a bilingual Xhosa-English lexicon, and exploring the comprehension levels of native speakers towards medical words in the language.
Sam [73]	2010	The development and implementation of IsiXhosa computer terminology to promote digital literacy and access to technology in educational settings.	The study contributes to IsiXhosa lexicon development and digitization by creating a specialised set of computer terminology in IsiXhosa, facilitating the integration of the language into the digital realm, and promoting its use in the context of technology and education.
Gunnink [74]	2022	It focuses on reconstructing phonemic clicks from Proto-Nguni and examining the historical development of clicks in Nguni languages.	The study contributes to isiXhosa lexicon development and digitization by providing insights into the historical presence of phonemic clicks in Nguni languages, including isiXhosa, through the reconstruction of Proto-Nguni click phonemes and their subsequent evolution in individual Nguni languages.
Møller et al. [76]	2016	Involved translating the National Wellbeing Index into isiXhosa and exploring the understanding of well-being concepts among native isiXhosa speakers.	The study's contribution to isiXhosa lexicon development and digitization lies in its exploration of how well-being concepts are translated and understood in isiXhosa, shedding light on the nuances of language and culture in the context of measuring national well-being.
Eckart et al. [87]	2019	The use of computational linked data-based techniques for aligning multilingual fragmented and heterogeneous data for under-resourced Bantu languages.	The contribution to isiXhosa lexicon development and digitization in the paper lies in the prototypical implementation for aligning Xhosa language dictionaries based on their English translations, which helps in enhancing the quality and interconnectedness of digital dictionaries for isiXhosa.

TABLE 4. (Continued.) Efforts on IsiXhosa lexicon development and digitization.

Makhenyane [90]	2024	Evaluate and propose standardised research terminology in isiXhosa for master’s and doctoral research reports.	The study contributes to isiXhosa lexicon development and digitization by proposing standardised research terminology for master’s and doctoral research reports in isiXhosa, addressing the lack of standardisation and enhancing the language’s capacity for academic research.
Eckart et al. [91]	2018	Utilising linked data techniques, specifically RDF format, to create a highly interoperable and machine-processable lexical resource for isiXhosa and other Bantu languages.	The study contributes to isiXhosa lexicon development and digitization by creating a dedicated ontology, the Bantu Language Model (BLM), to represent isiXhosa lexical and morphological language data, facilitating future language tools and applications for interconnection and enrichment with other language data.
Allwood et al. [94]	2010	The collection, transcription, and analysis of spoken language samples for all the South African official languages.	The study contributes to isiXhosa lexicon development and digitization by analysing the influence of English and Afrikaans on spoken Xhosa, studying word frequencies and loans, and highlighting the importance of corpus data for understanding language contact and development in multilingual environments.

Xhosa lexicographical data in a multilingual setting, highlighting the significance of open-source data for language processing.

Gunnink [74] examines the historical development of the isiXhosa vocabulary by reconstructing phonemic clicks from Proto-Nguni. This analysis offers valuable insights into the evolution of Nguni languages. Allwood et al. [94] examine the effects of English and Afrikaans on spoken Xhosa, emphasising the importance of language contact on the vocabulary of isiXhosa. Grabar et al. [71] also employ natural language processing techniques to identify and classify medical terms, construct a bilingual Xhosa-English dictionary, and improve local speakers’ understanding of medical terminology. These contributions demonstrate a strong and diverse approach to developing and digitising the isiXhosa vocabulary. The use of annotated corpora, specialised dictionaries, computational tools, and instructional resources is essential to the preservation, improvement, and digital integration of isiXhosa.

2) GRAMMAR AND MORPHOLOGY ANALYSIS

Efforts in grammar and morphology analysis make valuable contributions to the fields of language documentation, theoretical linguistics, and language technology creation. These efforts conserve significant linguistic knowledge for future generations and gain insights into the typological aspects of IsiXhosa by documenting grammatical structures and morphological instances. In addition, computational models that rely on grammatical and morphological analyses are fundamental components of natural language processing applications, including machine translation, information retrieval, and text production.

Significant and diverse efforts have been made to analyse the grammar and morphology of isiXhosa (Table 5), indicating a wide-ranging interest in both theoretical linguistics and practical uses. A considerable amount of research has been dedicated to the computational elements of isiXhosa

grammar. Mahlaza and Keet [21], [55] created computational grammar rule sets and context-free grammars for isiXhosa and isiZulu verbs. This has improved the understanding of morphosyntactic similarity and accuracy in two closely related languages. They play a crucial role in the development of computational linguistics and the creation of language technology for isiXhosa, specifically in the area of weather forecast verbs. In a similar vein, Bosch et al. [43] and Pretorius and Bosch [44], [46] made significant contributions to the advancement of finite-state morphological analyzers and computational morphological analyzers for the Nguni language group. They have used cross-linguistic similarities to effectively initiate the construction of analyzers specifically for isiXhosa.

In addition, a number of studies have explored the syntactic and morphological features of isiXhosa, as well as computational approaches. Carstens and Zeller [18] investigated the syntax and structure of isiXhosa, with a particular emphasis on the use of focus particles. Taraldsen et al. [30] explored the nominal class prefixes in Southern Bantu languages, shedding light on their function as specifiers in lexicalization. Carstens and Mletshe [60] conducted a study on nominal structures, specifically focusing on the syntactic and distributional characteristics of nominals in negative contexts. This study shed light on the negative concord aspects of nominal licensing. The additional area of interest has been the intersection between phonology and morphology. Braver and Bennett [65] investigated the way in which isiXhosa speakers use root phonotactics to assign noun classes to nonce words, emphasising the relationship between phonological structure and morphological information. In their study, Gxowa-Dlavedwa [51] examined the orthographic patterns of initial click consonants in isiXhosa syllables. The study focused on improving the understanding of syllable structure and the pronunciation of click clusters. Furthermore, other researchers use unsupervised and machine learning methods to analyse and

TABLE 5. Efforts on IsiXhosa grammar and morphology analysis.

Source	Publication Year	Digitization Effort	Summary/Key Findings
Dowling Krause [17]	2019	It explores the use of isiXhosa grammar and morphology in English teaching.	The study contributes to the analysis of isiXhosa grammar and morphology within the context of English teaching.
Carstens and Zeller [18]	2020	It delves into the syntax and structure of isiXhosa, particularly in relation to the use of focus particles.	By examining the syntax and structure of isiXhosa and Zulu languages, particularly focusing on the use of the phrase-final focus particles 'only', the paper provides insights into how these particles interact with various syntactic elements in the sentence.
Persohn [19]	2020	It focuses on the analysis of grammar and morphology in Bantu languages, specifically the Xhosa language.	The study makes a significant contribution to the analysis of isiXhosa grammar and morphology by providing a detailed examination of the aspectual periphrasis involving the construction ya 'go' plus subordinate imperfective.
Mahlaza and Keet [21]	2019	It focuses on developing computational grammar rule sets for verbs in closely related languages (Zulu and Xhosa) and analyzing the syntactic and semantic correctness of these rules within the context of weather forecast verbs.	The research enhances the understanding of isiXhosa grammar and morphology, particularly in the context of computational linguistics and language technology development.
Taraldsen et al. [30]	2018	Analyzing the syntactic and morphological properties of nominal class prefixes in Southern Bantu languages.	The study contributes to isiXhosa Grammar and Morphology Analysis by providing insights into the structure and function of nominal class prefixes in isiXhosa, specifically highlighting their role as specifiers in lexicalization within the Bantu language family.
Mzamo et al. [41]	2019	The evaluation of combined bi-directional branching entropy language models for morphological segmentation of isiXhosa	The study contributes to isiXhosa grammar and morphology analysis in the evaluation and implementation of the IsiXhosa Branching Entropy Segmenter (XBES), which introduces a novel approach using branching entropy techniques for unsupervised morphological segmentation of isiXhosa text.
Bosch et al. [43]	2008	The development of finite-state morphological analysers for Xhosa, Swati, and Ndebele.	The study contributes to isiXhosa grammar and morphology analysis by demonstrating the feasibility of using fast-tracking techniques to develop accurate and usable morphological analyzers for isiXhosa.
Pretorius and Bosch [44]	2009	It leverage cross-linguistic similarities between Zulu and isiXhosa to bootstrap a morphological analyser for the language.	The study contributes to isiXhosa grammar and morphology analysis by demonstrating how cross-linguistic similarities with Zulu can be utilised to bootstrap a morphological analyzer for isiXhosa.
Carstens and Mletshe [60]	2015	Analysis of the syntactic and distributional properties of nominals in Xhosa and Zulu in negative contexts, particularly focusing on the role of negative concord features and nominal licensing.	The study makes a significant contribution to isiXhosa grammar and morphology analysis by providing insights into the syntactic constraints and distributional patterns of nominals in negative contexts, shedding light on the role of negative concord features and nominal licensing in the language.
Mostert [61]	2019	Analysing the linguistic features of isiXhosa number names to understand their impact on the learning and teaching of mathematics in early-grade education.	The study contributes to isiXhosa grammar and morphology analysis by examining the syntactic category, transparency, regularity, length, and spoken vs. written differences of number names in isiXhosa, comparing them with English, and discussing the implications of these linguistic features for counting and mathematical thinking in early-grade education.
Taraldsen [62]	2019	Analysing the properties and similarities between augment-less NPs in Xhosa and de NPs in French to understand their syntactic structures and implications.	The study contributes to isiXhosa grammar and morphology analysis by examining the properties of augment-less NPs in Xhosa, highlighting similarities with de NPs in French, and providing insights into the syntactic structures and implications of these linguistic phenomena.
Braver and Bennett [65]	2016	Explore how speakers of isiXhosa utilise root phonotactics to assign noun classes to nonce words.	The study contributes to isiXhosa grammar and morphology analysis by demonstrating how speakers use root phonotactics to assign noun classes to words, highlighting the interaction between phonological structure and abstract morphological information in the grammar of isiXhosa.

TABLE 5. (Continued.) Efforts on IsiXhosa grammar and morphology analysis.

Andrason [80]	2019	Analyse the categorial status and polyfunctionality of the conjunctive coordinator NA in Xhosa.	The study makes a significant contribution to isiXhosa grammar and morphology analysis by providing a detailed examination of the morpheme NA as a conjunctive coordinator in Xhosa, discussing its categorial status, polysemy, and polyfunctionality within the language.
Savic [77]	2017	Investigate the semantics of Xhosa's perfective and imperfective aspects.	The study makes a significant contribution to isiXhosa grammar and morphology analysis by providing insights into how aspects function in the language, clarifying the distinctions between perfective and imperfective verbs, and highlighting the role of boundedness in understanding eventuality conceptualization in Xhosa.
Bloom Ström [78]	2022	Involves analysing and documenting the prosodic features of Xhosa sentence types.	The study contributes to isiXhosa grammar and morphology analysis by providing insights into the role of vowel length and pitch in sentence type intonation and shedding light on the prosodic features that play a crucial role in distinguishing different linguistic functions within the language.
Ma and Simango [79]	2023	Analyse the syntactic structure and ellipsis phenomenon in isiXhosa to determine the presence or absence of V-stranding VP ellipsis.	The study contributes to isiXhosa grammar and morphology analysis by providing empirical evidence that V-stranding VP ellipsis does not exist in isiXhosa, highlighting the syntactic positions of VP-internal constituents and the reasons behind the impossibility of V-stranding VP ellipsis in the language.
Andrason and Matutu [80]	2019	A corpus-driven examination of the syntactic properties of interjections in isiXhosa.	The study makes a significant contribution to isiXhosa grammar and morphology analysis by providing a detailed examination of the syntactic properties of interjections in the language, shedding light on their compliance with linguistic typology and the interjectional prototype.
Mzamo et al. [81]	2019	The development of unsupervised morphological segmenters for isiXhosa.	The study contributes to isiXhosa grammar and morphology analysis by proposing and evaluating unsupervised morphological segmenters that utilise branching entropy techniques and language heuristics, providing insights into automated segmentation methods for the complex morphological structure of isiXhosa.
Myler and Mali [88]	2021	The syntactic and semantic analysis of isiXhosa causative constructions.	The study contributes to isiXhosa grammar and morphology analysis by proposing a novel explanation for the case alternation with respect to causees in productive morphological causatives.
Oosthuysen [45]	2015	It extricate the description of the grammar of isiXhosa from a Euro-centric approach.	The study contributes to isiXhosa grammar and morphology analysis by highlighting the importance of approaching the language from its own genius, setting aside grammatical presuppositions from other languages, and emphasising a phenomenological method to uncover the true structures and features of isiXhosa.
Pretorius and Bosch [46]	2010	The development of computational morphological analyzers for the Nguni cluster of languages, namely Zulu, Xhosa, Swati, and Ndebele.	The study contributes to isiXhosa grammar and morphology analysis by outlining the bootstrapping process of the Zulu morphological analyzer to create analyzer prototypes for Xhosa.
Gxowa-Dlayedwa [51]	2018	It involves analyzing the orthographic patterns of onset click consonants in isiXhosa syllables to facilitate the pronunciation and understanding of click clusters	The study contributes to isiXhosa grammar and morphology analysis by providing insights into the structure of syllables, the formation of click consonant clusters, and the impact of orthographic patterns on pronunciation, thereby enhancing the understanding of the linguistic features of isiXhosa for both native and non-native speakers.
Mahlaza and Keet [55]	2018	The development of Context Free Grammars for isiXhosa and isiZulu verb fragments to measure morphosyntactic similarity.	The paper contributes to isiXhosa grammar and morphology analysis by providing insights into the morphosyntactic similarity between isiXhosa and isiZulu verbs through the development of context-free grammars and the application of binary similarity measures.

TABLE 5. (Continued.) Efforts on IsiXhosa grammar and morphology analysis.

Mzamo [59]	2021	It focuses on using unsupervised machine learning techniques for isiXhosa's morphological segmentation.	The study's contribution to isiXhosa grammar and morphology analysis is to advance the use of unsupervised machine learning techniques for surface segmentation, thereby improving the understanding and processing of the intricate morphological structures of isiXhosa, a Nguni language with significant linguistic complexity.
Ron Simango [93]	2011	Analyze the code-switching grammar between English and isiXhosa.	The study contributes to isiXhosa grammar and morphology analysis by examining how the grammar of code-switching with English influences the structural patterns and features of isiXhosa spoken by bilingual speakers, shedding light on the impact of language contact on isiXhosa linguistic systems.

segment the morphology of isiXhosa. Mzamo et al. [41], [81] and Mzamo [59] conducted an assessment of bi-directional branching entropy language models and developed unsupervised morphological segmenters for isiXhosa. The studies introduced novel methods to automate the segmentation of intricate morphological features. Andrason [80] examined the conjunctive coordinator NA, exploring its categorial position and polyfunctionality. Savic [77] examined the meaning of the perfective and imperfective aspects in isiXhosa, providing a clear understanding of the differences and the significance of boundedness in the conceptualization of occurrences. Myler and Mali [88] conducted a study that specifically examined the syntactic and semantic analysis of isiXhosa causative constructs. The study put forth theories for the occurrence of case alternation in productive morphological causatives.

All of these efforts strengthen the understanding of isiXhosa grammar and morphology, offering essential knowledge on its computational, syntactic, phonological, and morphological characteristics.

3) SPEECH SYNTHESIS TECHNOLOGIES

Speech technologies are essential for enabling communication, improving accessibility, and preserving languages. Researchers are using advancements in signal processing, machine learning, and linguistics to develop speech technologies that are robust and efficient. These technologies are designed to meet the demands of a wide range of users, including individuals with low literacy or disabilities.

The spectrum of approaches and applications being explored in the development and improvement of isiXhosa speech synthesis technologies (Table 6) is broad as well as diverse. Significant effort has been devoted to developing text-to-speech (TTS) systems specifically designed for isiXhosa. An example of this is the work done by Kogeda [6] in creating a TTS system that transforms isiXhosa text into a voice that sounds natural. This is an important step in making isiXhosa available on digital platforms. Roux and Visagie [34] employed HTS (Hidden Markov Model-based Speech Synthesis) to construct a speech synthesiser for isiXhosa. The study utilised data-driven techniques to handle syntactic and other text characteristics,

resulting in enhanced naturalness and comprehensibility of the synthesised speech. Louw and Moodley [26] conducted research on improving TTS systems by developing automated methods for transforming, coding, and modelling intonation patterns particular to the isiXhosa language. The idea was to enhance the prosodic characteristics of synthesised speech, a vital aspect of achieving a more natural and comprehensible sound in speech, particularly for languages with limited resources such as isiXhosa. This study builds upon the research conducted by Schlünz et al. [64], which utilised natural language processing techniques including part-of-speech tagging and chunking to improve the authenticity of synthesised speech in isiXhosa.

The advances in automatic speech recognition (ASR) systems have played a major role in the development of isiXhosa speech synthesis technology. Naidoo and Tsoeu [40] conducted an assessment of open-source toolkits for automatic speech recognition (ASR) in isiXhosa. The study identified the most effective settings that can improve the accuracy of speech recognition. Precise recognition of input speech is a necessary requirement for effective speech synthesis, as it is crucial for enhancing the quality of synthesised speech. Biswas et al. [31] expanded upon previous work by enhancing automatic speech recognition (ASR) systems to effectively handle code-switched speech, a prevalent occurrence in multilingual settings that involve the languages isiXhosa and English. This study on language models and acoustic models played an invaluable role in improving the accuracy of recognising isiXhosa code-switched utterances.

Nthite and Tsoeu's [38] study, which implemented and assessed an end-to-end TTS system for under-resourced South African languages, notably isiXhosa and Sesotho, demonstrates another innovative technique. The approach demonstrated promising results in terms of clarity and authenticity, emphasising the practicality and possibilities of end-to-end systems for improving speech synthesis in isiXhosa.

The development of isiXhosa voice synthesis technologies is aided by these initiatives, which explore different approaches, including the building of a TTS system, modelling intonation, improving ASR, and integrating multi-lingual applications.

TABLE 6. Efforts on IsiXhosa speech synthesis technologies.

Source	Publication Year	Digitization Effort	Summary/Key Findings
Kogeda [6]	2014	Designed, developed, and implemented a text-to-speech system to convert isiXhosa text into natural-sounding isiXhosa speech.	The study text-to-speech system contributes to isiXhosa speech synthesis technology by producing audio utterances for the isiXhosa text-to-speech system.
Louw and Moodley [26]	2017	It focuses on the automatic stylization, coding, and modeling of intonation in text-to-speech for under-resourced languages.	The study contributes to isiXhosa Speech Synthesis Technologies by presenting an automatic method for modeling intonation specific to isiXhosa speakers, enhancing the naturalness and intelligibility of synthesized speech in under-resourced languages like isiXhosa.
Marais et al. [27]	2020	It develop a multilingual, multi-modal speech-to-speech translation application for maternal health care.	The paper contributes to isiXhosa speech synthesis technologies by incorporating isiXhosa as one of the target languages in the speech-to-speech translation application for maternal health care, enabling health care providers to communicate with patients in isiXhosa.
Yılmaz et al. [28]	2018	It focuses on developing a Unified Code-Switching ASR System for South African Languages	The paper contributes to isiXhosa Speech Synthesis Technologies by incorporating isiXhosa as one of the languages in the Unified Code-Switching ASR System for South African Languages, thereby advancing research in processing and synthesizing isiXhosa speech in a multilingual context.
Biswas et al. [31]	2022	Improving an ASR system to process code-switched South African speech containing English and four indigenous languages.	The study contributes to isiXhosa Speech Synthesis Technologies by exploring methods to improve automatic speech recognition (ASR) for code-switched speech involving isiXhosa and the development of language models and acoustic models to enhance the recognition accuracy of isiXhosa code-switched utterances.
Roux and Visagie [34]	2007	It uses HTS to build an isiXhosa speech synthesizer.	It contributes to isiXhosa speech synthesis technologies by developing a data-driven text-to-speech system for syntactic and other text features in isiXhosa.
Nthite and Tsoeu [38]	2020	the implementation and evaluation of an end-to-end Text-To-Speech system for under-resourced South African languages, specifically isiXhosa and Sesotho.	The study contributes to isiXhosa speech synthesis technologies by introducing and evaluating an end-to-end text-to-speech system for isiXhosa, showcasing promising results in terms of intelligibility and naturalness.
Naidoo and Tsoeu [40]	2019	The evaluation of open-source toolkits for automatic speech recognition of South African languages, specifically English and isiXhosa.	The study contributes to isiXhosa speech synthesis technologies by evaluating and identifying optimal toolkit configurations for automatic speech recognition in isiXhosa, which can enhance the development of speech synthesis technologies for the language.
Jansen van Vuren and Niesler [48]	2022	It involves enhancing N-Best rescoring in under-resourced code-switched speech recognition.	The study contributes to isiXhosa speech synthesis technologies by demonstrating improvements in N-Best rescoring for code-switched speech involving isiXhosa through the use of pretraining and data augmentation methods, ultimately enhancing speech recognition accuracy for the language.
Schlünz et al. [64]	2016	Applying the natural language processing tasks of part-of-speech tagging and chunking to enhance text-to-speech synthesis for South African languages.	The paper contributes to isiXhosa speech synthesis technologies by applying part-of-speech tagging and chunking to enhance the naturalness of synthesised speech in isiXhosa language within the text-to-speech synthesis process.
van der Westhuizen and Niesler [70]	2019	It involves using word embeddings to synthesise artificial bigrams for improving automatic speech recognition of code-switched speech between English and four South African Bantu languages.	The study contributes to isiXhosa speech synthesis technologies by addressing the challenges of code-switching between English and isiXhosa, providing insights into improving language models for bilingual language switches, and synthesising artificial bigrams using word embeddings to enhance automatic speech recognition.

4) MACHINE TRANSLATION

Machine translation has great potential for overcoming language barriers and enabling communication between

different languages. Researchers frequently prioritise the development and enhancement of machine translation systems that facilitate smooth translation. Researchers are

TABLE 7. Efforts on IsiXhosa machine translation.

Source	Publication Year	Digitization Effort	Summary/Key Findings
Stahlberg et al. [16]	2018	It involves combining pre-trained language models with translation models in neural machine translation (NMT) to improve translation quality and performance.	The study contributes to isiXhosa machine translation by demonstrating the effectiveness of integrating pre-trained language models with translation models in Neural Machine Translation (NMT), leading to improved translation quality and performance on the isiXhosa-English language pair.
Campbell and Young [22]	2016	Involves the translating the CORE-OM questionnaire into Xhosa.	The paper contributes to isiXhosa Machine Translation by providing insights into the translation process of a psychometric tool, the CORE-OM, from English to Xhosa.
McKellar and Puttkammer [35]	2020	It involves creating a parallel dataset with four reference translations for each of the 11 official South African languages	The paper contributes to isiXhosa machine translation by providing a valuable dataset that includes four reference translations for isiXhosa, enabling researchers to evaluate and improve machine translation systems for the language.
Steele and Edwards [37]	2008	It involves translating the Beck Depression Inventory-II, Beck Anxiety Inventory, and Beck Hopelessness Scale into Xhosa	The study contributes to isiXhosa machine translation by demonstrating the successful translation of psychological assessment scales (Beck inventories) into Xhosa, showcasing the potential for accurate and culturally relevant machine translation in the field of mental health assessment and research within the Xhosa-speaking population.
Hendy et al. [39]	2021	It involves training multilingual and bilingual models using a multitask objective with parallel and monolingual data, augmented with back translation and knowledge distillation techniques.	The study contributes to isiXhosa machine translation by focusing on the language pair Xhosa and Zulu as part of the constrained track in the WMT21 shared news translation task.
Chiguvare and Cleghorn [49]	2021	Focused on improving transformer model translation for low resource South African languages using BERT.	The paper contributes to isiXhosa machine translation by evaluating the performance of pre-trained transformer models on low-resource South African languages, including isiXhosa, and introducing SALR-mBERT, a model that outperforms mBERT and transformer models for certain language corpora sizes.
Emezue and Dossou [56]	2021	The development of a many-to-many multilingual translation system, MMTAfrica, for African languages.	The paper contributes to isiXhosa Machine Translation by introducing MMTAfrica, a many-to-many multilingual translation system that includes isiXhosa (Xhosa) as one of the African languages.
Martínez [75]	2021	It involves utilising low-resource language pairs and implementing a simple system with techniques such as BPE dropout, sub-subword features, and back-translation to achieve good results in the WMT 2021 News Translation and Biomedical Translation tasks.	The paper contributes to isiXhosa machine translation by focusing on the Xhosa to Zulu language pair in the News Translation Task of WMT 2021, showcasing the effectiveness of techniques like BPE dropout, sub-subword features, and back-translation in improving translation quality for low-resource language pairs.
Hedderich et al. [82]	2020	Leverage multilingual transformer models, transfer learning, and distant supervision techniques to improve NLP tasks for low-resource African languages like Hausa, isiXhosa, and Yoruba.	The study contributes to isiXhosa machine translation specifically in the exploration of transfer learning and distant supervision techniques to enhance NLP tasks, showcasing the potential for improved machine translation capabilities in isiXhosa through innovative approaches.
Sefara et al. [83]	2021	Training and evaluating neural machine translation (NMT) models for low-resourced languages.	The study contributes to isiXhosa machine translation by training and evaluating Neural Machine Translation (NMT) models, specifically the Transformer NMT with self-attention, which achieved state-of-the-art performance for isiXhosa language translation.

using advancements in natural language processing, neural networks, and computational linguistics to narrow the divide in translating natural language between different widely spoken languages.

A wide range of areas, including neural machine translation (NMT), dataset development, and specific translation problems, have been covered in isiXhosa machine translation

efforts (Table 7). One major emphasis has been on using pre-trained language models to improve translation quality. Stahlberg et al. [16] show this by combining pre-trained language models with translation models in NMT. This led to big improvements in the quality and speed of translation between isiXhosa and English. Hendy et al. [39] investigated the training of multilingual and bilingual models using

TABLE 8. Efforts on IsiXhosa digitization and language preservation tools.

Source	Publication Year	Digitization Effort	Summary/Key Findings
Rabie and Naidoo [15]	2019	The adapting and translating the South African Career Interest Inventory into isiXhosa, thereby creating a digital version of the inventory in an indigenous South African language.	By creating a digital version of a career assessment tool in isiXhosa, the study enhances the availability of culturally relevant resources for vocational assessment in South Africa.
Faaß et al. [20]	2014	It discusses effort involved in transforming printed dictionaries into electronic formats using Optical Character Recognition (OCR) technology.	The study contributes to isiXhosa digitization and language preservation tools by proposing a model for storing lexicographic data that includes isiXhosa language entries.
du Toit and Kidd [24]	2016	The translation and evaluation of the EORTC QLQ-CX24 questionnaire into Xhosa and Afrikaans languages.	The study's translation and validation of the EORTC QLQ-CX24 questionnaire into Xhosa contributes to isiXhosa digitization and language preservation tools by providing a standardized and culturally relevant quality of life assessment tool for women with cervical cancer in the Xhosa-speaking population, enhancing healthcare research and practice in South Africa.
Diko [25]	2023	The archiving and preserving isiXhosa language materials.	The paper contributes to isiXhosa digitization and language preservation tools by highlighting the significance of information science in archiving isiXhosa materials.
Oluwagbemi et al. [54]	2020	The development of a fuzzy logic-based HIV diagnostic tool with indigenous multi-lingual interfaces.	The study contributes to isiXhosa digitization and language preservation tools by integrating isiXhosa language interfaces into the fuzzy logic-based HIV diagnostic system, which improves communication and healthcare services in isiXhosa-speaking rural African communities.
Modise and Suleman [58]	2017	The development of mobile search interfaces (text and voice) to support isiXhosa speakers in accessing information on the web.	The study's contribution to isiXhosa digitization and language preservation tools lies in the development of mobile search interfaces tailored for isiXhosa speakers, providing them with accessible means to search for information on the web in their native language, thus enhancing digital access and utilisation of the isiXhosa language.
Botha et al. [69]	2020	It involves reconceptualizing poetry education in South African classrooms by infusing indigenous poetry texts and practices to revitalise high school poetry.	The study contributes to isiXhosa digitization and language preservation tools by highlighting the rich oral tradition and poetic practices of isiXhosa poetry and emphasising the importance of incorporating indigenous knowledge and cultural practices in poetry education.
Banda [84]	2017	The utilisation of learners' multilingual repertoire, including African language mother tongues, can enhance teaching and learning practices.	The study contributes to isiXhosa digitization and language preservation tools by showcasing how translanguaging practices in a black township school in Cape Town enable learners to utilise their multilingual repertoire, including isiXhosa, to enhance their educational experience and bridge the gap between home language and English content learning.
Smit et al. [85]	2006	Translating and culturally adapting a mental health battery for use in a Xhosa-speaking community.	The study contributes to isiXhosa digitization and language preservation tools by detailing the process of translating and culturally adapting mental health screening instruments for a Xhosa-speaking community and demonstrating an approach to digitising and preserving isiXhosa language in the field of mental health research.
Gqada et al. [86]	2021	Translating and linguistically validating the EORTC QLQ-PAN26 questionnaire into isiXhosa and Afrikaans.	The study contributes to isiXhosa digitization and language preservation tools by providing a validated translation of a health-related quality of life questionnaire, the EORTC QLQ-PAN26, into isiXhosa, thereby enhancing the availability of medical assessment tools in the indigenous language and promoting the understanding of patients' experiences in their mother tongue.

TABLE 8. (Continued.) Efforts on IsiXhosa digitization and language preservation tools.

Diko [89]	2022	Utilise qualitative research methodology to analyse Welile Shasha's poetry collection "Umsila Wembulu" and unlock the multiplicity of meanings in one selected poem addressing health-related issues.	The study contributes to isiXhosa digitization and language preservation tools by highlighting the significance of indigenous knowledge systems in Welile Shasha's poetry, emphasising the need to bridge the gap in isiXhosa literature, especially in addressing health-related issues, and promoting the use of African languages in academia and contemporary societal challenges.
Juul et al. [92]	2012	Adapting and validating the King's Health Questionnaire into English, Afrikaans, and isiXhosa languages.	The study's contribution to isiXhosa digitization and language preservation tools lies in the adaptation and validation of the King's Health Questionnaire into isiXhosa, ensuring linguistic and cultural relevance for assessing the quality of life of the isiXhosa-speaking community.

multifunctional objectives, back translation, and knowledge distillation approaches. The study emphasizes the possibility of improving translation accuracy in under-resourced languages such as isiXhosa by utilizing advanced training methods.

Another crucial area of study focuses on the development and application of parallel datasets to enhance machine translation systems. McKellar and Puttkammer [35] made an important step by creating a parallel dataset that includes four reference translations for each of the 11 official languages of South Africa, including isiXhosa. This dataset is an ideal tool for researchers looking to assess and improve machine translation systems. Similarly, Steele and Edwards [37] showed how psychological assessment instruments like the Beck Depression Inventory-II and Beck Anxiety Inventory can be translated into isiXhosa. The study highlights the significance of precise and culturally appropriate translations in specialised domains such as mental health. The incorporation of isiXhosa into multilingual translation systems is another vital effort. Emezue and Dossou [56] created MMTAfrica, a multilingual translation system that handles multiple African languages, including isiXhosa. The primary objective of this system is to cater to the requirements of African languages, thereby fostering linguistic diversity and enhancing the accessibility of digital translation tools. Martínez [75] also worked on language pairs with limited resources, employing methods such as BPE dropout, sub-subword features, and back-translation to attain remarkable outcomes in the WMT 2021 News Translation and Biomedical Translation tasks.

The efforts to improve transformer models for isiXhosa machine translation are equally remarkable. Chiguvare and Cleghorn [49] enhanced the translation of low-resource South African languages in transformer models by integrating BERT. The SALR-mBERT model surpasses conventional transformer models in specific language corpus sizes, showcasing the advantages of employing sophisticated pre-trained models in environments with limited resources. Hedderich et al. [82] conducted additional research on multilingual transformer models, transfer learning, and distant supervision techniques to improve natural language

processing (NLP) tasks for African languages with limited resources, such as isiXhosa. This study demonstrates novel methods for enhancing the capabilities of machine translation through transfer learning.

Researchers have also investigated the practical applicability of machine translation in specific circumstances. Campbell and Young [22] conducted a translation of the CORE-OM questionnaire into isiXhosa, providing insight into the process of translating psychometric tools. Together, these efforts advance the study of isiXhosa machine translation by leveraging pre-trained models, generating significant datasets, including multilingual systems, and employing creative ways to improve translation accuracy.

5) DIGITIZATION AND LANGUAGE PRESERVATION TOOLS

Digitizing and preserving language resources is critical for documenting and preserving linguistic legacies, supporting language revitalization initiatives, and encouraging cultural diversity. The development of digital archives, language corpora, and computer tools to preserve, analyse, and disseminate linguistic data is significantly transforming our cultural heritage. Researchers strive to digitize oral traditions, written texts, and linguistic artefacts, ensuring their easy accessibility and preservation for future generations through collaborative efforts and community involvement.

Initiatives in isiXhosa digitalization and language preservation (Table 8) involve a variety of efforts to improve the accessibility and significance of isiXhosa in several domains. A major area of emphasis has been the modification and adjustment of evaluation instruments to guarantee their cultural and linguistic suitability for IsiXhosa-speaking communities. Rabie and Naidoo [15] played a crucial role in this matter by modifying and converting the South African Career Interest Inventory into isiXhosa. This project ensures the accessibility and relevance of career evaluation tools, thereby assisting in vocational coaching, particularly within the isiXhosa-speaking community. Similarly, du Toit and Kidd [24] and Gqada et al. [86] conducted research on the translation and validation of health-related quality of life

questionnaires into isiXhosa. Du du Toit and Kidd concentrated on the EORTC QLQ-CX24 questionnaire, augmenting its applicability for isiXhosa-speaking groups of women with cervical cancer. Gqada et al. translated the EORTC QLQ-PAN26, which is a valuable tool for assessing the quality of life of patients who speak isiXhosa. These efforts guarantee that medical evaluations are not only precise but also culturally responsive, consequently enhancing patient care. Another notable area of contribution is the creation of digital and mobile technologies to improve isiXhosa utilization. Modise and Suleman [58] created mobile search interfaces in isiXhosa that allow people to easily search the internet for information using their own language. This effort aims to improve digital literacy and access for IsiXhosa speakers, hence encouraging increased use of the language on the internet. Botha et al. [69] concentrated on reimagining poetry education in high schools by incorporating indigenous isiXhosa poetry texts and practices into the curriculum, with the aim of maintaining and developing isiXhosa literature and oral traditions. It not only safeguards the abundant oral customs of isiXhosa poetry, but also rejuvenates poetry instruction by incorporating indigenous knowledge systems.

Diko [25], [89] emphasised the significance of archiving and safeguarding IsiXhosa materials. The study emphasizes the significance of information science in preserving isiXhosa materials and examines the various interpretations found in Welile Shasha's poetry. This indicates the need to close the divide in isiXhosa literature and advocate for the inclusion of African languages in academic and modern societal issues. To summarise, the combined efforts in digitising isiXhosa and preserving the language encompass a wide range of areas, including healthcare, education, digital tools, and literature. These initiatives are essential for enhancing the accessibility and relevance of isiXhosa in today's settings, thereby guaranteeing the language's continued survival and cultural value.

IV. DISCUSSION

The purpose of this review was to understand, from the literature, the digitalization efforts implemented to preserve and promote the IsiXhosa language and culture in this digital era. The results indicate isiXhosa language efforts across various domains—lexicon development and digitization, speech recognition, speech synthesis, machine translation, and digitization tools—show both significant progress and notable gaps. Among other areas, there has been significant progress in isiXhosa speech recognition and synthesis. Biswas et al. [31] and Yilmaz et al. [28] have made efforts to include isiXhosa into multilingual ASR systems, improving the accuracy of recognising code-switched speech. In a similar vein, the studies conducted by Kogeda [6] and Roux and Visagie [34] have resulted in the development of text-to-speech systems that transform written IsiXhosa language into a voice that sounds natural.

Significant progress has been achieved in enhancing the quality and efficiency of translation systems for isiXhosa in the field of machine translation. Research conducted by Stahlberg et al. [16] and Chiguvare and Cleghorn [49] has shown that including pre-trained language models and transformer models in neural machine translation leads to improved translation accuracy. In addition, the creation of parallel datasets, as emphasised by McKellar and Puttkammer [35], has offered useful resources for the training and assessment of translation systems. The digitalization and language preservation efforts have been remarkable. Projects carried out by Rabie and Naidoo [15] and Turner [8] have concentrated on the translation and adaptation of diverse evaluation instruments and educational resources into isiXhosa. This effort seeks to protect and promote the use of the language in many contexts. Modise and Suleman's [58] research on creating customized mobile search interfaces for isiXhosa speakers highlights the importance of ensuring that digital information is easily accessible to native speakers of the language. Researchers in the field of grammar and morphology analysis are pursuing various efforts to create computational models that can effectively analyze isiXhosa's grammatical structure and morphological characteristics. Studies by Bosch et al. [43] and Pretorius and Bosch [44], [46] contributed to the advancement of finite-state morphological analyzers and computational morphological analyzers for the isiXhosa. These efforts can improve isiXhosa grammar comprehension and facilitate the development of more advanced language technologies.

Despite these advancements and digitization efforts, there are still several gaps in the development of isiXhosa language technologies. The review revealed several gaps, including some of the following:

- **Large and high-quality dataset:** A significant gap still exists in the isiXhosa dataset due to the limited availability of large, high-quality datasets for training language models. Many of the studies reviewed rely on relatively small datasets, which can limit the performance and generalizability of the resulting models.
- **Cultural and contextual aspects of language technology:** Many efforts fall short of fully capturing the cultural context and idiomatic expressions that are exclusive to isiXhosa. This weakness can have negative impacts on the accuracy and precision of speech synthesis and translation.
- **Lexicon development:** There is a need to continuously and consistently update and broaden the isiXhosa lexicons to incorporate current and specialised vocabulary, as well as the need to effectively combine isiXhosa lexicon resources with other language technologies in order to provide seamless functionality across various applications.
- **Speech evaluation metrics:** To fully understand the work that has gone into isiXhosa speech synthesis and

technologies, it is important to have evaluation metrics that take into account the language's unique features, like tone and intonation.

- Publication demographics and trends: About 78% (Fig. 3(b)) of the research and digitization activities are led by South African authors, indicating a significant national commitment. However, fostering greater international collaboration will enhance research and provide further support. The year of publication (Fig. 3(a)) also provides insights into notable patterns in both the growth and decline of research outputs. The overall trend indicates that although there are periods of increased attention, consistent and persistent effort is required to ensure continual development in isiXhosa digitalization efforts.

In order to close these gaps, more broad data gathering efforts must first be made in order to produce sizable, varied datasets that accurately reflect the diversity of the IsiXhosa language and cultural variety. Collecting such data could be facilitated through collaborative endeavours among academic institutions, government organisations, and local communities. Also, it is imperative for future research to prioritise the development of models that exhibit both language accuracy and cultural sensitivity. This entails integrating cultural context and idiomatic expressions into language models and developing evaluation metrics that consider isiXhosa's distinctive characteristics. Researchers should further investigate the possibilities of transfer learning and data augmentation techniques to enhance the performance of isiXhosa language technology when there is a scarcity of data. Furthermore, it is imperative to make a collaborative endeavour to ensure that these technologies are easily accessible and user-friendly for individuals who speak isiXhosa. This entails not only the creation of easy-to-use interfaces but also incorporating these technologies into typical applications like education, healthcare, and digital communication platforms. It is crucial to prioritize community interaction and feedback when developing and deploying these technologies to ensure that they effectively cater to the requirements and preferences of isiXhosa speakers. It is essential to consistently update and broaden digital lexicons and incorporate them with other language technologies. This may entail establishing forums for community contributions to ensure the lexicons remain up-to-date and relevant.

A. LIMITATIONS

While this scoping review offers vital insights into the present state of isiXhosa language digitization efforts, it is crucial to recognise certain limitations that come with the study's focus. Only the studies screened and classified into various themes form the basis of the review. Therefore, this may not include all other research and advancements in the field. The approach might overlook valuable contributions from non-peer-reviewed sources or research conducted outside

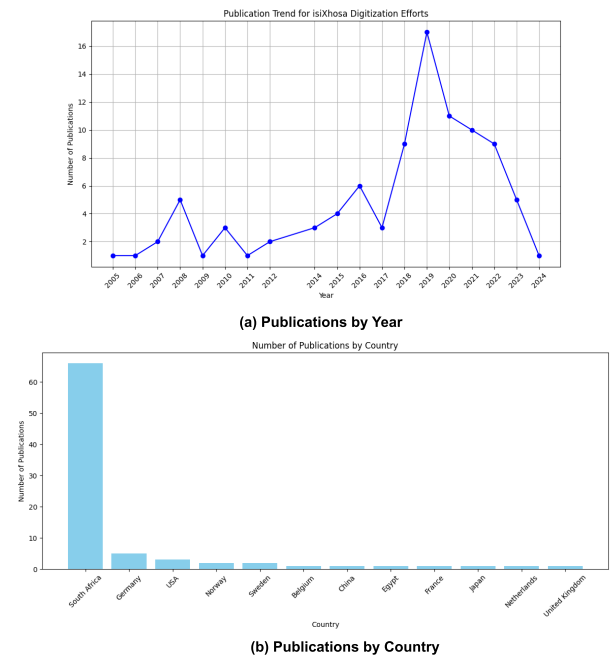


FIGURE 3. Publication Demographics and Trends.

of academic publications such as SADiLaR,⁴ established in 2019, to produce and maintains freely available digital language annotated corpora in all South African languages including resources in and for isiXhosa; wikipedia's isiXhosa encyclopaedia⁵ that focus on increasing the presence of African languages on the platform including isiXhosa; dictionary creation units in South Africa that were state funded, namely the Bureau of the Woordboek van die Afrikaanse Taal (WAT)⁶ and the Dictionary for South African English (DSAE),⁷ Pan-South African Language Board (PanSALB),⁸ and national lexicography units (NLUs) for the African languages [95] situated at tertiary institutions in South Africa among many other current and ongoing interventions.

Similarly, the review's emphasis on academic literature did not capture practical implementations or initiatives led by non-academic groups, such as community organisations, corporate partners, or government agencies. Academic research may not fully account for the critical role these actors play in influencing language technology development and implementing practical solutions.

Notwithstanding these limitations, the review provides a thorough overview of current digitization initiatives and highlights important gaps that require more research and

⁴<https://sadilar.org/en/>

⁵<https://af.wikipedia.org/wiki/Xhosa>

⁶<https://www.wat.co.za/>

⁷<https://dsae.co.za/>

⁸<https://www.pansalb.org/>

collaboration in the context of digitizing the isiXhosa language for the digital age.

V. CONCLUSION

This review focuses on the progress and ongoing work in digitizing the IsiXhosa language. It covers several areas, such as developing the lexicon, assessing grammar and morphology, developing speech recognition and synthesis, machine translation, and the development of language preservation tools. These efforts demonstrate an increasing acceptance of the significance of protecting and improving indigenous African languages, such as isiXhosa, in this era of digital technology.

The review highlights some significant developments in specific areas, such as speech recognition, grammar/morphology and machine translation. However, it also highlights the existence of unresolved gaps and challenges. It emphasises the importance of additional research and development in isiXhosa language technologies that could promote the language preservation, cultural legacy, and societal inclusivity.

Further research should focus on addressing the observed gaps and implementing the recommended strategies to develop more resilient and easily accessible language technologies that empower isiXhosa language and contribute to the wider scope of digitising and preserving African languages.

REFERENCES

- [1] K. A. B. Chew, V. Anthony-Stevens, A. LeClair-Diaz, S. E. Nicholas, A. Sobotta, and P. Stevens, "Enacting hope through narratives of Indigenous language and culture reclamation," *Transmotion*, vol. 5, no. 1, pp. 132–151, 2019, doi: [10.20204/UniKent/03/tm.570](https://doi.org/10.20204/UniKent/03/tm.570).
- [2] J. Ferguson and M. Weaselboy, "Indigenous sustainable relations: Considering land in language and language in land," *Current Opinion Environ. Sustainability*, vol. 43, pp. 1–7, Apr. 2020.
- [3] J. Gumperz, "The speech community," *Int. Encyclopedia Social Sci.*, vol. 9, no. 3, pp. 381–386, 1968.
- [4] R. Darvin and B. Norton, "Identity and a model of investment in applied linguistics," *Annu. Rev. Appl. Linguistics*, vol. 35, pp. 36–56, Mar. 2015, doi: [10.1017/s0267190514000191](https://doi.org/10.1017/s0267190514000191).
- [5] P. Maseko, "Intellectualisation of African languages with particular reference to isiXhosa," Ph.D. thesis, Fac. Humanities, Rhodes Univ., Grahamstown, South Africa, 2011. [Online]. Available: <https://tinyurl.com/phz5hvc9>
- [6] O. P. Kogeda, S. Mhlana, T. Mamello, and T. Olwal, "An implementation of isiXhosa text-to-speech modules to support e-services in marginalized rural areas," in *User-Centric Technology Design for Nonprofit and Civic Engagements* (Public Administration and Information Technology), vol. 9, S. Saeed, Ed., Cham, Switzerland: Springer, 2014, doi: [10.1007/978-3-319-05963-1_8](https://doi.org/10.1007/978-3-319-05963-1_8).
- [7] Q. Mgogo and S. Nokutya, "Reinterpreting technological language: The use of Facebook emojis to construct new meanings in isiXhosa language," in *African Language Media*. Evanston, IL, USA: Routledge, 2023, pp. 79–97.
- [8] I. Turner, "Decolonisation through digitalisation? African languages at South African universities," *Curriculum Perspect.*, vol. 43, no. S1, pp. 73–82, May 2023, doi: [10.1007/s41297-023-00196-w](https://doi.org/10.1007/s41297-023-00196-w).
- [9] N. wa Thiong'o, "The language of African literature," in *Colonial Discourse and Post-Colonial Theory*. Evanston, IL, USA: Routledge, 2015, pp. 435–455.
- [10] World Econ. Forum. *These Countries Rank Highest for Digital Competitiveness*. Accessed: Mar. 17, 2024. [Online]. Available: <https://www.weforum.org/agenda/2021/09/countries-rank-highest-digital-competitiveness/>
- [11] A. Warner. (Apr. 15, 2022). *How Digitization Helps Preserve Languages*. MultiLingual Media. [Online]. Available: <https://multilingual.com/digitization-language-preservation/>
- [12] M. Eisenhart, "On the subject of interpretive reviews," *Rev. Educ. Res.*, vol. 68, no. 4, p. 391, 1998, doi: [10.3102/00346543068004391](https://doi.org/10.3102/00346543068004391).
- [13] J. S. du Toit and M. J. Puttkammer, "Developing core technologies for resource-scarce Nguni languages," *Information*, vol. 12, no. 12, p. 520, Dec. 2021, doi: [10.3390/info12120520](https://doi.org/10.3390/info12120520).
- [14] E. van der Westhuizen and T. Niesler, "A first South African corpus of multilingual code-switched soap opera speech," in *Proc. 11th Int. Conf. Language Resources Eval. (LREC)*, May 2018. [Online]. Available: <https://aclanthology.org/L18-1451.pdf>
- [15] S. Rabie and A. V. Naidoo, "Validating the adaptation of the first career measure in isiXhosa: The South African career interest inventory–isiXhosa version," *South Afr. J. Psychol.*, vol. 49, no. 1, pp. 109–121, Apr. 2018, doi: [10.1177/0081246318772419](https://doi.org/10.1177/0081246318772419).
- [16] F. Stahlberg, J. Cross, and V. Stoyanov, "Simple fusion: Return of the language model," in *Proc. 3rd Conf. Mach. Transl. Res. Papers*, 2018, pp. 204–211, doi: [10.18653/v1/w18-6321](https://doi.org/10.18653/v1/w18-6321).
- [17] T. Dowling and L. Krause, "'Ndifuna imeaningyakhe': Translingual morphology in English teaching in a South African township classroom," *Int. J. Multilingualism*, vol. 16, no. 3, pp. 205–225, Jul. 2019.
- [18] V. Carstens and J. Zeller, "'Only' in Nguni: A phrase-final particle meets antisymmetry theory," *Linguistic Inquiry*, vol. 51, no. 2, pp. 199–235, Mar. 2020, doi: [10.1162/ling_a_00337](https://doi.org/10.1162/ling_a_00337).
- [19] B. Persohn, "A description of the Xhosa constructionya'go' plus subordinate imperfective," *J. Afr. Lang. Linguistics*, vol. 41, no. 1, pp. 57–98, Aug. 2020, doi: [10.1515/jall-2020-2004](https://doi.org/10.1515/jall-2020-2004).
- [20] G. Faab, S. E. Bosch, and R. H. Gouws, "A general lexicographic model for a typological variety of dictionaries in African languages," *Lexikos*, vol. 24, no. 1, pp. 94–115, Oct. 2014, doi: [10.5788/24-1-1254](https://doi.org/10.5788/24-1-1254).
- [21] Z. Mahlaza and C. M. Keet, "A method for measuring verb similarity for two closely related languages with application to Zulu and Xhosa," *South Afr. Comput. J.*, vol. 31, no. 2, pp. 34–56, Dec. 2019, doi: [10.18489/sacj.v31i2.698](https://doi.org/10.18489/sacj.v31i2.698).
- [22] M. M. Campbell and C. Young, "A xhosa language translation of the CORE-OM using South African university student samples," *Transcultural Psychiatry*, vol. 53, no. 5, pp. 654–673, Oct. 2016, doi: [10.1177/1363461516661643](https://doi.org/10.1177/1363461516661643).
- [23] V. D. Mgijima, "Advancing text prediction skills through translanguaging," *Reading Writing*, vol. 12, no. 1, pp. 1–10, Jun. 2021, doi: [10.4102/rw.v12i1.284](https://doi.org/10.4102/rw.v12i1.284).
- [24] G. C. du Toit and M. Kidd, "An analysis of the psychometric properties of the translated versions of the European Organisation for the research and treatment of cancer QLQ CX24 questionnaire in the two South African indigenous languages of Xhosa and Afrikaans," *Eur. J. Cancer Care*, vol. 25, no. 5, pp. 832–838, Sep. 2016, doi: [10.1111/ecc.12333](https://doi.org/10.1111/ecc.12333).
- [25] M. Diko, "Archiving of indigenous languages with specific reference to isiXhosa," *Mousaion South Afr. J. Inf. Stud.*, p. 19, Oct. 2023, doi: [10.25159/2663-659x/12684](https://doi.org/10.25159/2663-659x/12684).
- [26] J. A. Louw and A. Moodley, "Automatic stylization, coding and modelling of intonation in text-to-speech for under-resourced languages," in *Proc. Pattern Recognit. Assoc. South Afr. Robot. Mechatronics (PRASA-RobMech)*, Nov. 2017, pp. 197–202, doi: [10.1109/RoboMech.2017.8261147](https://doi.org/10.1109/RoboMech.2017.8261147).
- [27] L. Marais, J. A. Louw, J. Badenhorst, K. Calteaux, I. Wilken, N. van Niekerk, and G. Stein, "AwezaMed: A multilingual, multimodal speech-to-speech translation application for maternal health care," in *Proc. IEEE 23rd Int. Conf. Inf. Fusion (FUSION)*, Jul. 2020, pp. 1–8, doi: [10.23919/FUSION45008.2020.9190240](https://doi.org/10.23919/FUSION45008.2020.9190240).
- [28] E. Yilmaz, A. Biswas, E. van der Westhuizen, F. de Wet, and T. Niesler, "Building a unified code-switching ASR system for South African languages," in *Proc. Interspeech*, Sep. 2018, pp. 1923–1927, doi: [10.21437/interspeech.2018-1966](https://doi.org/10.21437/interspeech.2018-1966).

- [29] M. Alberts, "Centre for political and related terminology in Southern African languages (CEPTSA)—Translating and explanatory dictionaries," *Lexikos*, vol. 25, pp. 461–494, Nov. 2015. [Online]. Available: <https://www.ajol.info/index.php/lex/article/view/126263>
- [30] K. T. Taraldsen, L. T. Medová, and D. Langa, "Class prefixes as specifiers in Southern Bantu," *Natural Lang. Linguistic Theory*, vol. 36, no. 4, pp. 1339–1394, Jan. 2018, doi: [10.1007/s11049-017-9394-8](https://doi.org/10.1007/s11049-017-9394-8).
- [31] A. Biswas, E. Yılmaz, E. van der Westhuizen, F. de Wet, and T. Niesler, "Code-switched automatic speech recognition in five South African languages," *Comput. Speech Lang.*, vol. 71, Jan. 2022, Art. no. 101262, doi: [10.1016/j.csl.2021.101262](https://doi.org/10.1016/j.csl.2021.101262).
- [32] F. Gillis-Webber, "Conversion of the English–Xhosa dictionary for nurses to a linguistic linked data framework," *Information*, vol. 9, no. 11, p. 274, Nov. 2018, doi: [10.3390/info9110274](https://doi.org/10.3390/info9110274).
- [33] S. Packham and H. Suleman, "Crowdsourcing a text corpus is not a game," in *Digital Libraries: Providing Quality Information*, 2015, pp. 225–234, doi: [10.1007/978-3-319-27974-9_23](https://doi.org/10.1007/978-3-319-27974-9_23).
- [34] J. C. Roux and A. S. Visagie, "Data-driven approach to rapid prototyping Xhosa speech synthesis," in *Proc. 6th ISCA Workshop Speech Synthesis*, vol. 95, Bonn, Germany, Aug. 2007. [Online]. Available: <https://core.ac.uk/download/pdf/15953597.pdf#page=167>
- [35] C. A. McKellar and M. J. Puttkammer, "Dataset for comparable evaluation of machine translation between 11 South African languages," *Data Brief*, vol. 29, Apr. 2020, Art. no. 105146, doi: [10.1016/j.dib.2020.105146](https://doi.org/10.1016/j.dib.2020.105146).
- [36] G. I. Steele and D. J. A. Edwards, "Development and validation of the Xhosa translations of the beck inventories: 1. Challenges of the translation process," *J. Psychol. Afr.*, vol. 18, no. 2, pp. 207–215, Jan. 2008, doi: [10.1080/14330237.2008.10820188](https://doi.org/10.1080/14330237.2008.10820188).
- [37] G. I. Steele and D. J. A. Edwards, "Development and validation of the Xhosa translations of the beck inventories: 2. Item analysis, internal consistency and factor analysis," *J. Psychol. Afr.*, vol. 18, no. 2, pp. 217–226, Jan. 2008, doi: [10.1080/14330237.2008.10820189](https://doi.org/10.1080/14330237.2008.10820189).
- [38] T. Nthite and M. Tsoeu, "End-to-end text-to-speech synthesis for under resourced South African languages," in *Proc. Int. SAUPEC/RobMech/PRASA Conf.*, Jan. 2020, pp. 1–6, doi: [10.1109/SAUPEC/RobMech/PRASA48453.2020.9041030](https://doi.org/10.1109/SAUPEC/RobMech/PRASA48453.2020.9041030).
- [39] A. Hendy, E. A. Gad, M. Abdelghaffar, J. S. ElMosalami, M. Afify, A. Y. Tawfik, and H. H. Awadalla, "Ensembling of distilled models from multi-task teachers for constrained resource language pairs," 2021, *arXiv:2111.13284*.
- [40] A. Naidoo and M. Tsoeu, "Evaluating open-source toolkits for automatic speech recognition of South African languages," in *Proc. Southern Afr. Universities Power Eng. Conf./Robot. Mechatronics/Pattern Recognit. Assoc. South Afr. (SAUPEC/RobMech/PRASA)*, Jan. 2019, pp. 160–165, doi: [10.1109/RoboMech.2019.8704774](https://doi.org/10.1109/RoboMech.2019.8704774).
- [41] L. Mzamo, A. Helberg, and S. Bosch, "Evaluation of combined bi-directional branching entropy language models for morphological segmentation of isiXhosa," in *Proc. FAIR*, 2019, pp. 77–89. [Online]. Available: <https://tinyurl.com/3d8n6u62>
- [42] L. Mzamo, A. Helberg, and S. Bosch, "Evaluation of heuristic guided character level word models for morphological segmentation of isiXhosa," in *Proc. SACAIR Appl. AI*, 2020, pp. 162–178. [Online]. Available: <https://tinyurl.com/2xybvmdda>
- [43] S. Bosch, L. Pretorius, K. Podile, and A. Fleisch, "Experimental fast-tracking of morphological analysers for Nguni languages," in *Proc. 6th Int. Conf. Lang. Resour. Eval. (LREC)*, Marrakech, Morocco: European Language Resources Association (ELRA), 2008, pp. 2588–2595. [Online]. Available: http://www.lrec-conf.org/proceedings/lrec2008/pdf/643_paper.pdf
- [44] L. Pretorius and S. Bosch, "Exploiting cross-linguistic similarities in Zulu and Xhosa computational morphology," in *Proc. 1st Workshop Lang. Technol. African Lang.*, 2009, pp. 96–103. [Online]. Available: <http://hdl.handle.net/10204/4040>
- [45] J. Oosthuysen, "Extricating the description of the grammar of isiXhosa from a eurocentric approach," *South Afr. J. Afr. Lang.*, vol. 35, no. 1, pp. 83–92, Jan. 2015, doi: [10.1080/02572117.2015.1056467](https://doi.org/10.1080/02572117.2015.1056467).
- [46] L. Pretorius and S. Bosch, "Finite state morphology of the Nguni language cluster: Modelling and implementation issues," in *Finite-State Methods and Natural Language Processing*, 2010, pp. 123–130, doi: [10.1007/978-3-642-14684-8_13](https://doi.org/10.1007/978-3-642-14684-8_13).
- [47] C. Marquard and H. Suleman, "Focused crawling for automated IsiXhosa corpus building," in *South African Institute of Computer Scientists and Information Technologists*, 2023, pp. 19–31, doi: [10.1007/978-3-031-39652-6_2](https://doi.org/10.1007/978-3-031-39652-6_2).
- [48] J. Jansen van Vuren and T. Niesler, "Improving N-best rescoring in under-resourced code-switched speech recognition using pretraining and data augmentation," *Languages*, vol. 7, no. 3, p. 236, Sep. 2022, doi: [10.3390/languages7030236](https://doi.org/10.3390/languages7030236).
- [49] P. Chiguvare and C. W. Cleghorn, "Improving transformer model translation for low resource South African languages using BERT," in *Proc. IEEE Symp. Ser. Comput. Intell. (SSCI)*, Dec. 2021, pp. 1–8, doi: [10.1109/SSCI50451.2021.9659923](https://doi.org/10.1109/SSCI50451.2021.9659923).
- [50] L. Mzamo, A. Helberg, and S. Bosch, "Introducing XGL—A lexicalised probabilistic graphical lemmatiser for isiXhosa," in *Proc. Pattern Recognit. Assoc. South Afr. Robot. Mechatronics Int. Conf. (PRASA-RobMech)*, Nov. 2015, pp. 142–147, doi: [10.1109/RoboMech.2015.7359513](https://doi.org/10.1109/RoboMech.2015.7359513).
- [51] N. Gxowa-Dlavedwa, "Investigating click clusters in isiXhosa syllables," *South Afr. J. Afr. Lang.*, vol. 38, no. 3, pp. 317–325, Sep. 2018, doi: [10.1080/02572117.2018.1518043](https://doi.org/10.1080/02572117.2018.1518043).
- [52] R. Eiselien and A. Bukula, "IsiXhosa named entity recognition resources," *ACM Trans. Asian Low-Resource Lang. Inf. Process.*, vol. 22, no. 2, pp. 1–19, Dec. 2022, doi: [10.1145/3531478](https://doi.org/10.1145/3531478).
- [53] T. Gaustad and M. J. Puttkammer, "Linguistically annotated dataset for four official South African languages with a conjunctive orthography: IsiNdebele, isiXhosa, isiZulu, and Siswati," *Data Brief*, vol. 41, Apr. 2022, Art. no. 107994, doi: [10.1016/j.dib.2022.107994](https://doi.org/10.1016/j.dib.2022.107994).
- [54] O. O. Oluwagbemi, F. E. Oluwagbemi, A. Jatto, and C. Hui, "MAVSCOT: A fuzzy logic-based HIV diagnostic system with indigenous multi-lingual interfaces for rural Africa," *PLoS ONE*, vol. 15, no. 11, Nov. 2020, Art. no. e0241864, doi: [10.1371/journal.pone.0241864](https://doi.org/10.1371/journal.pone.0241864).
- [55] Z. Mahlaza and C. M. Keet, "Measuring verb similarity using binary coefficients with application to isiXhosa and isiZulu," in *Proc. Annu. Conf. South Afr. Inst. Comput. Scientists Inf. Technologists*, Sep. 2018, doi: [10.1145/3278681.3278690](https://doi.org/10.1145/3278681.3278690).
- [56] C. C. Emezue and B. F. P. Dossou, "MMTAfrica: Multilingual machine translation for African languages," 2022, *arXiv:2204.04306*.
- [57] A. Deumert and S. O. Masinyana, "Mobile language choices—The use of English and isiXhosa in text messages (SMS)," *English World-Wide. A J. Varieties English*, vol. 29, no. 2, pp. 117–147, Apr. 2008, doi: [10.1075/eww.29.2.02deu](https://doi.org/10.1075/eww.29.2.02deu).
- [58] M. Modise and H. Suleman, "Mobile search interfaces for isiXhosa speakers: A comparison between voice and text," in *Proc. Conf. Inf. Commun. Technol. Soc. (ICTAS)*, Mar. 2017, pp. 1–6, doi: [10.1109/ICTAS.2017.7920649](https://doi.org/10.1109/ICTAS.2017.7920649).
- [59] L. Mzamo, "Morphological segmentation of isiXhosa using unsupervised machine learning," Ph.D. dissertation, North-West Univ., Potchefstroom, South Africa, 2021. [Online]. Available: <https://repository.nwu.ac.za/handle/10394/38844>
- [60] V. Carstens and L. Mletshe, "Negative concord and nominal licensing in Xhosa and Zulu," *Natural Lang. Linguistic Theory*, vol. 34, no. 3, pp. 761–804, Nov. 2015, doi: [10.1007/s11049-015-9320-x](https://doi.org/10.1007/s11049-015-9320-x).
- [61] I. Mostert, "Number names: Do they count?" *Afr. J. Res. Math., Sci. Technol. Educ.*, vol. 23, no. 1, pp. 64–74, Jan. 2019, doi: [10.1080/18117295.2019.1589038](https://doi.org/10.1080/18117295.2019.1589038).
- [62] K. T. Taraldsen, "On augment-less NPs in Xhosa and de NPs in French," *Stellenbosch Papers Linguistics Plus*, vol. 58, Dec. 2019, doi: [10.5842/58-0-836](https://doi.org/10.5842/58-0-836).
- [63] R. Parker, J. Jelsma, and D. J. Stein, "Pain in amaXhosa women living with HIV/AIDS: Translation and validation of the brief pain inventory-Xhosa," *J. Pain Symptom Manage.*, vol. 51, no. 1, pp. 126–132.e2, Jan. 2016, doi: [10.1016/j.jpainsymman.2015.08.004](https://doi.org/10.1016/j.jpainsymman.2015.08.004).
- [64] G. I. Schlünz, N. Dlamini, and R. P. Kruger, "Part-of-speech tagging and chunking in text-to-speech synthesis for South African languages," in *Proc. Interspeech*, Sep. 2016, doi: [10.21437/interspeech.2016-1040](https://doi.org/10.21437/interspeech.2016-1040).
- [65] A. Braver and Wm. G. Bennett, "Phonotactic c(l)ues to Bantu noun class disambiguation," *Linguistics Vanguard*, vol. 2, no. 1, Dec. 2016, doi: [10.1515/lingvan-2016-0062](https://doi.org/10.1515/lingvan-2016-0062).
- [66] S. Bosch, T. Eckart, and B. Klimek, "Preparation and usage of Xhosa lexicographical data for a multilingual, federated environment," in *Proc. 11th Int. Conf. Language Resources Eval. (LREC)*, 2018, pp. 4372–4378. [Online]. Available: <https://aclanthology.org/L18-1692.pdf>

- [67] V. de Klerk, "Procedural meanings of well in a corpus of Xhosa English," *J. Pragmatics*, vol. 37, no. 8, pp. 1183–1205, Aug. 2005, doi: [10.1016/j.pragma.2004.11.001](https://doi.org/10.1016/j.pragma.2004.11.001).
- [68] I. E. Mostert, "Relative difficulty of early grade compare type word problems: Learning from the case of isiXhosa," *Pythagoras*, vol. 41, no. 1, Dec. 2020, doi: [10.4102/pythagoras.v41i1.538](https://doi.org/10.4102/pythagoras.v41i1.538).
- [69] L. Botha, P. Y. De Villiers, and R. Maungedzo, "Research that is real and Utopian: Indigenous knowledge as a resource to revitalise high school poetry," *Educ. Change*, vol. 24, Dec. 2020, doi: [10.25159/1947-9417/7897](https://doi.org/10.25159/1947-9417/7897).
- [70] E. van der Westhuizen and T. R. Niesler, "Synthesised bigrams using word embeddings for code-switched ASR of four South African language pairs," *Comput. Speech Lang.*, vol. 54, pp. 151–175, Mar. 2019, doi: [10.1016/j.csl.2018.10.002](https://doi.org/10.1016/j.csl.2018.10.002).
- [71] N. Grabar, I. van Zyl, R. de La Harpe, and T. Hamon, "The comprehension of medical words—Cross-lingual experiments in French and Xhosa," in *Proc. 7th Int. Conf. Health Inform. (HEALTHINF)*, vol. 2, Mar. 2014, pp. 334–342, doi: [10.5220/0004803803340342](https://doi.org/10.5220/0004803803340342).
- [72] A. Andrason, "The conjunctive coordinator NA in Xhosa: Its categorial status and map of polyfunctionality," *Italian J. Linguistics*, vol. 31, no. 1, pp. 3–60, 1120. [Online]. Available: <https://tinyurl.com/bdzzfvp4>
- [73] M. S. Sam, "The development and implementation of computer literacy terminology in isiXhosa," Ph.D. thesis, Fac. Humanities, Rhodes Univ., Grahamstown, South Africa, 2010. [Online]. Available: <http://hdl.handle.net/10962/d1002155>
- [74] H. Gunnink, "The early history of clicks in Nguni," *Diachronica*, vol. 39, no. 3, pp. 311–368, Mar. 2022, doi: [10.1075/dia.19061.gun](https://doi.org/10.1075/dia.19061.gun).
- [75] A. Martínez, "The Fujitsu DMATH submissions for WMT21 news translation and biomedical translation tasks," in *Proc. 6th Conf. Mach. Transl.*, 2021, pp. 162–166. [Online]. Available: <https://aclanthology.org/2021.wmt-1.13>
- [76] V. Møller, B. J. Roberts, and D. Zani, "The national wellbeing index in the IsiXhosa translation: Focus group discussions on how South Africans view the quality of their society," *Social Indicators Res.*, vol. 135, no. 1, pp. 167–193, Nov. 2016, doi: [10.1007/s11205-016-1481-4](https://doi.org/10.1007/s11205-016-1481-4).
- [77] S. Savic, "The perfective and the imperfective aspects in Xhosa," *Stellenbosch Papers Linguistics Plus*, vol. 52, pp. 45–72, Apr. 2017, doi: [10.5842/52-0-713](https://doi.org/10.5842/52-0-713).
- [78] E.-M. Bloom Ström, "The role of vowel length and pitch in Xhosa sentence type intonation," *Stellenbosch Papers Linguistics Plus*, vol. 62, no. 2, Dec. 2022, doi: [10.5842/62-2-894](https://doi.org/10.5842/62-2-894).
- [79] X. Ma and S. R. Simango, "The search for V-stranding VP ellipsis in isiXhosa," *Southern Afr. Linguistics Appl. Lang. Stud.*, vol. 41, no. 3, pp. 315–335, Apr. 2023, doi: [10.2989/16073614.2022.2136222](https://doi.org/10.2989/16073614.2022.2136222).
- [80] A. Andrason and H. Matutu, "The syntax of interjections in isiXhosa: A corpus-driven study," *Stellenbosch Papers Linguistics Plus*, vol. 58, pp. 1–16, Dec. 2019, doi: [10.5842/58-0-833](https://doi.org/10.5842/58-0-833).
- [81] L. Mzamo, A. Helberg, and S. Bosch, "Towards an unsupervised morphological segmenter for isiXhosa," in *Proc. Southern Afr. Universities Power Eng. Confe./Robot. Mechatronics/Pattern Recognit. Assoc. South Afr. (SAUPEC/RobMech/PRASA)*, Bloemfontein, South Africa, Jan. 2019, pp. 166–170, doi: [10.1109/RoboMech.2019.8704816](https://doi.org/10.1109/RoboMech.2019.8704816).
- [82] M. A. Hedderich, D. Adelani, D. Zhu, J. Alabi, U. Markus, and D. Klakow, "Transfer learning and distant supervision for multilingual transformer models: A study on African languages," in *Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP)*, Oct. 2020, pp. 2580–2591, doi: [10.18653/v1/2020.emnlp-main.204](https://doi.org/10.18653/v1/2020.emnlp-main.204).
- [83] T. J. Sefara, S. G. Zwane, N. Gama, H. Sibisi, P. N. Senoamadi, and V. Marivate, "Transformer-based machine translation for low-resourced languages embedded with language identification," in *Proc. Conf. Inf. Commun. Technol. Soc. (ICTAS)*, Durban, South Africa, Mar. 2021, pp. 127–132, doi: [10.1109/ICTAS50802.2021.9394996](https://doi.org/10.1109/ICTAS50802.2021.9394996).
- [84] F. Banda, "Translanguaging and English–African language mother tongues as linguistic dispensation in teaching and learning in a black township school in cape town," *Current Issues Lang. Planning*, vol. 19, no. 2, pp. 198–217, Jul. 2017, doi: [10.1080/14664208.2017.1353333](https://doi.org/10.1080/14664208.2017.1353333).
- [85] J. Smit, C. E. van den Berg, L. G. Bekker, S. Seedat, and D. J. Stein, "Translation and cross-cultural adaptation of a mental health battery in an African setting," *Afr. Health Sci.*, vol. 6, no. 4, pp. 215–222, 2006. [Online]. Available: <https://www.ajol.info/index.php/ahs/article/view/6956>
- [86] K. Gqada, U. Kotze, V. Soldati, C. Kloppers, J. Krige, and E. Jonas, "Translation and linguistic validation of the EORTC QLQ-PAN26 questionnaire for assessment of health-related quality of life in patients with pancreatic cancer and chronic pancreatitis into isiXhosa and afrikaans," *South Afr. J. Surg.*, vol. 59, no. 4, pp. 153–156, 2021. [Online]. Available: <https://www.ajol.info/index.php/sajs/article/view/245535>
- [87] T. Eckart, S. Bosch, D. Goldhahn, U. Quasthoff, and B. Klimek, "Translation-based dictionary alignment for under-resourced Bantu languages," *OpenAccess Ser., Inform.*, vol. 70, May 2019, doi: [10.4230/OASICS.LDK.2019.17](https://doi.org/10.4230/OASICS.LDK.2019.17).
- [88] N. Myler and Z. O. Mali, "Two places for causees in productive IsiXhosa morphological causatives," *Syntax*, vol. 24, no. 1, pp. 1–43, Mar. 2021, doi: [10.1111/synt.12208](https://doi.org/10.1111/synt.12208).
- [89] M. Diko, "Uhlalutyzo bunzululwazi ngokomxholo ochongiweyo kwingqokelela yesihobe sikaWelile shasha ethi umsila wembulu," *South Afr. J. Afr. Lang.*, vol. 42, no. 3, pp. 265–271, Sep. 2022, doi: [10.1080/02572117.2022.2132689](https://doi.org/10.1080/02572117.2022.2132689).
- [90] L. E. Makhenyane, "Ukuphuhliselwa ezobuNgcali kweeLwimi zaseAfrika: IsiXhosa njengolwimi lokuphanda," *South Afr. J. Afr. Lang.*, vol. 43, no. suppl, pp. 351–360, Mar. 2024, doi: [10.1080/02572117.2024.2322279](https://doi.org/10.1080/02572117.2024.2322279).
- [91] T. Eckart, B. Klimek, D. Goldhahn, and S. Bosch, "Using linked data techniques for creating an isiXhosa lexical resource—A collaborative approach," in *Proc. CLARIN Annu. Conf.*, 2018, pp. 26–29. [Online]. Available: <https://core.ac.uk/download/pdf/572206780.pdf#page=33>
- [92] L. Juul, J. A. Van Rensburg, and P. S. Steyn, "Validation of the king's health questionnaire for South Africa in English, Afrikaans and isiXhosa," *South Afr. J. Obstetrics Gynaecol.*, vol. 18, no. 3, pp. 82–84, Oct. 2012. [Online]. Available: <https://www.ajol.info/index.php/sajog/article/view/82941>
- [93] S. Ron Simango, "When English meets isiXhosa in the clause: An exploration into the grammar of code-switching," *Southern Afr. Linguistics Appl. Lang. Stud.*, vol. 29, no. 2, pp. 127–134, Jun. 2011, doi: [10.2989/16073614.2011.633361](https://doi.org/10.2989/16073614.2011.633361).
- [94] J. Allwood, H. Hammarström, A. Hendrikse, M. N. Ngcobo, N. Nomdebevana, L. Pretorius, and M. van der Merwe, "Work on spoken (multimodal) language corpora in South Africa," in *Proc. 7th Int. Conf. Lang. Resour. Eval. (LREC)*, 2010, pp. 880–884. [Online]. Available: <https://urn.kb.se/resolve?urn=urn:nbn:se:hb:diva-6500>
- [95] M. Alberts, "The South African national lexicography units—Two decades later," *Lexikos*, vol. 32, no. 3, 2022, doi: [10.5788/32-3-1729](https://doi.org/10.5788/32-3-1729).



ABAYOMI AGBEYANGI received the B.Sc. degree in computer engineering, the M.Sc. degree in computer science, and the Ph.D. degree in computer engineering. His area of expertise is in natural language processing, machine learning, embedded systems, and ICT4D research. He has published his research work in several esteemed journals and conferences. He is also a member of many learned professional bodies.



ROBERT JERE received the Ph.D. degree in computer science. He is currently an Associate Professor. He is also a Pan Africanist. He has a passion for sustainable digital solutions that empower disadvantaged citizens. He is in the ICT for sustainable development research niche area. He has published over 50 articles in accredited journals and conference proceedings. He has supervised the completion of over 15 postgraduate students at the Ph.D. and master's levels. His research interests include sustainable emerging technologies, i.e., AI, big data, and machine learning. He is also a South African National Research Foundation (NRF)-Rated Researcher.

...