

Human language technology tools for indigenous South African languages and their potential use

**Authors:**Respect Mlambo¹ Muzi Matfunjwa¹ **Affiliations:**

¹South African Centre for Digital Language Resources, Faculty of Humanities, North-West University, Potchefstroom, South Africa

Corresponding author:

Respect Mlambo, respect.mlambo@nwu.ac.za

Dates:

Received: 11 Dec. 2023

Accepted: 09 Dec. 2024

Published: 28 Jan. 2025

How to cite this article:

Mlambo, R. & Matfunjwa, M., 2025, 'Human language technology tools for indigenous South African languages and their potential use', *Literator* 46(1), a2049. <https://doi.org/10.4102/lit.v46i1.2049>

Copyright:

© 2025. The Authors. Licensee: AOSIS. This work is licensed under the Creative Commons Attribution License.

Read online:

Scan this QR code with your smart phone or mobile device to read online.

Human language technology (HLT) contributes to the development of languages by providing various avenues through which languages can be interrogated. Through HLT, diverse questions can be raised and answered scientifically and objectively. In the context of South African indigenous languages (SAIL), several HLT tools support these languages. However, it seems that some language users are unaware of the availability and capabilities of these tools, which contributes to their underutilisation. This study aims to identify and describe briefly some of the HLT tools that support and analyse SAIL. It presents an overview of the open access HLT tools, namely part-of-speech (POS) taggers, morphological decomposers (MDs), morphological analysers (MAs), isiZulu.net, ZulMorph and Google Translate (GT). These tools are crucial in analysing and understanding SAIL, as well as for advancing these languages in the field of HLT. In this study, the researchers anticipate that by raising awareness of the existence of these tools, more users of indigenous languages will be eager to use them.

Contribution: This study fills the practical gap in the use of HLT to perform linguistic functions for SAIL. It seems that there is underutilisation of existing HLT tools for SAIL, which might be attributed to language users being unaware of these tools. Therefore, the study aims to identify and describe some HLT tools that support and analyse SAIL. It presents an overview of the open access HLT tools, namely POS taggers, MD, MA, isiZulu.net, ZulMorph and GT. The researchers intend to demonstrate the use of these tools and to raise awareness about their existence.

Keywords: human language technology; part-of-speech tagger; morphological decomposer; morphological analyser; ZulMorph; isiZulu.net; Google Translate.

Introduction

Human language technology (HLT) is necessary for South Africa, wherein about 25 languages are spoken (Grover, Van Huyssteen & Pretorius 2011), which include the 12 official languages: English, Afrikaans, isiZulu, isiXhosa, Setswana, Sesotho, Tshivenda, Sepedi, Xitsonga, Siswati, isiNdebele and South African Sign Language. Human language technology plays an important role in enhancing communication in multilingual communities and fulfils the needs of language practitioners as well as language speakers (Van Huyssteen et al. 2023). It includes software systems that recognise, grasp, interpret and interrogate human language in all forms, whether spoken, written or signed (Rivera Pastor et al. 2017). Rashel (2011) described HLT as a computational approach that uses computer programs and electronic devices that are capable of processing, analysing, producing and understanding human languages. They offer a wide range of applications through which users may engage with computer-assisted devices by using their languages. Therefore, HLT is the result of a collaboration between language and computer-assisted technology, and it brings together professionals from a variety of fields, including linguistics, engineering, computer science and programming. Tokenisers, part-of-speech (POS) taggers, machine translation systems, grammar checkers, spell checkers, named-entity recognisers, morphological analysers (MAs), syntactic parsers, information retrieval tools, speech recognition software and speech synthesis engines are among the core technologies and tools shared by HLT-based technologies for language analysis and production (Rivera-Pastor et al. 2017).

Human language technology is a field in South Africa that has been in existence for the past 20 years (Puttkammer et al. 2018). However, most South African indigenous languages (SAIL) are resource-scarce in terms of corpora, resources that are necessary to build HLT tools (Heeringa, De Wet & Van Huyssteen 2015). According to Tachbelie, Abate and Besacier (2011), HLT applications for these languages are built using a small amount of data acquired by researchers. As a result, the performance and accuracy of such tools are typically low when compared to those of developed languages (Grover et al. 2011; Koehn & Knowles 2017; Skosana & Mlambo 2021; Van Zaanen et al.

2020). This problem has caused many languages to be underserved by HLT, especially SAIL.

The South African government has recognised the role that HLT may play in bringing solutions to language-related difficulties in South Africa (Grover et al. 2011). The government took initiatives to fund HLT projects which were meant to promote all South African official languages and to close the gap that exists between these languages in the HLT sphere. This is in accordance with the South African constitutional mandate to take constructive steps to elevate the status and usage of languages in various fields (Mlambo, Skosana & Matfunjwa 2021). The government initiatives have been achieved with establishing research infrastructures such as universities, scientific councils, and a few private sectors that create several language technologies for South African languages. For instance, with government funding, the Centre for Text Technology (CTeX) at North-West University, University of Pretoria and language specialists were assigned as agencies responsible for the development of various text and language technologies such as corpora (one million words for each language), aligned corpora (50 000 words for each language, aligned on sentence level), wordlists for all eleven official languages, POS taggers, MAs and lemmatisers for the indigenous languages (Van Huyssteen & Griesel 2016:330). This has resulted in HLT tools and applications being created for SAIL.

However, researchers concede that many HLT tools that support indigenous languages are not widely used because of a lack of awareness and understanding by potential users. Abbott and Martinus (2019) concurred that as a result of a lack of research and community knowledge of the available tools built for these resource-limited languages, their use is minimal. Some of these tools are not user-friendly in terms of their graphical user interface and installation. Rogers (2003) also noted that it is very difficult to convince users to adopt a new technology regardless of the overt advantages that come with it. Moreover, factors contributing to the acceptance of a new technology include the amount to which it is perceived to be superior to what was available prior to its introduction, as well as how difficult it is to use and understand it (Rogers 2003). Also, factors like literacy level, prior experience and exposure to HLT tools, and attitudes towards the use of computers influence indigenous languages' adoption of HLT tools (Buabeng-Andoh 2012; Schiler 2003). This is perpetuated by the fact that some communities still do not have computers and access to stable Internet to download and use these technologies (Adefila, Oladokun & Adewojo 2024; Allam et al. 2022; Plomp et al. 2009). Despite these challenges, it is imperative to raise awareness about available HLT tools and their functionality to promote their use, especially in SAIL.

This article aims to present an overview of some useful HLT tools that support SAIL. The open-access HLT tools, namely POS taggers, morphological decomposers (MDs), morphological analysers (MAs), isiZulu.net, ZulMorph and Google Translate (GT) are described with their functionalities

in analysing SAIL, namely Sesotho sa Leboa, Siswati, isiNdebele, isiZulu and Xitsonga.

Literature review

Roux and Bosch (2006) conducted a web-based survey which was intended to provide an overview of initiatives aimed at advancing HLT tools and resources for South African languages. The survey found that the National Language Services (NLS) developed monolingual, multilingual and trilingual terminologies for various governmental sectors. The NLS also contributed by providing word lists for the development of spell checkers for all South African official languages. The Pan South African Language Board (PanSALB), through the Afrikaans National Lexicographic Units, developed an online Afrikaans dictionary. Roux and Bosch (2006) noted that various departments within South African universities also contributed to the advancement of HLT tools and language resources. The University of Pretoria (UP) compiled corpora that were used to develop spell checkers and tools for automatic corpus annotation. The University of South Africa (UNISA) initiated a project to construct MAs for several languages. At the University of Limpopo, synthesisers and automatic speech recognition systems were created for Northern Sotho, Setswana, Tshivenda and Xitsonga. The North-West University developed language resources collaboratively with UP and UNISA such as spell checkers for Afrikaans and five other South African languages. Moreover, Stellenbosch University developed telephone speech databases for Afrikaans, English, isiZulu, isiXhosa and Sesotho. Roux and Bosch (2006) also observed that the HLT Research Group created speech technology applications such as DictionaryMaker for electronic pronunciation dictionaries.

Grover and colleagues (2011) conducted an audit of HLT in South Africa. This audit was meant to explore speech and text technologies that are available for all South African official languages. The authors noted that HLT was a priority for the government as text, and speech applications were needed for the South African languages. The need for HLT language resources and tools was classified into three levels of priority. The first level consisted of HLT tools that required immediate development, such as machine translation, proofreading and information extraction for text, computer-assisted language learning, audio search and accessibility applications for speech. The second level included text and speech applications that require attention based on requirements. Speech recognisers, computer-assisted training, speech devices and access control were examples of speech applications. For text, optical character recognition, computer-assisted language learning, multilingual comprehension assistance and authorship recognition were stated. The third level included language-enhancing applications. For text, these applications were text generation, document categorisation, question answering and automated summarisation. For speech, transcription and detection, multimodal information access, speech-to-speech translation and audio books were needed. It was also found

that in terms of HLT tools and language resources, Afrikaans was leading, followed by English, isiZulu, isiXhosa, Setswana, Sepedi, Sesotho, Tshivenda, Siswati, isiNdebele and Xitsonga. Grover and colleagues (2011) also noted that Tshivenda, Siswati, isiNdebele and Xitsonga did not have any translation tools. The significance of HLT in improving the educational environment for students has been highlighted in studies (Alhawiti 2014; Cloete 2017; Popenici & Kerr 2017). Alhawiti emphasised that HLT is a key factor in enhancing the learning process, showcasing its effectiveness in language acquisition and academic performance improvement. Human language technology is integrated into various educational contexts such as research, science, languages, e-learning and assessment systems, thus contributing to positive outcomes in both schools and higher education settings. Alhawiti (2014) also noted that the use of language technologies, including grammar checkers and spell checkers, is identified as a time-saving measure for both teachers and learners. Cloete's (2017) research aligned with these findings, demonstrating how the integration of HLT into education can significantly enhance the quality of teaching and learning. This author stated that merging HLT and education creates opportunities for improvements in evaluations, data collection, learning progress and the development of new teaching tactics. Popenici and Kerr (2017) also supported the importance of HLT, such as artificial intelligence in higher education, because it opens new opportunities and challenges for teaching and learning at universities. The authors also realised that HLT plays a crucial role in transforming and advancing educational practices.

Mlambo and Matfunjwa (2022) ascertained that using HLT such as the MonoConc Pro program (concordancer) provides a practical approach to the teaching and learning of Xitsonga lexical collocations. This analysis was facilitated using the corpus and MonoConc Pro program. It was determined that the use of this tool was efficient, as opposed to relying on traditional methods of teaching and learning linguistic traits such as collocations. The study also demonstrated the importance of concordancers and corpora in enhancing language proficiency and collocational competence awareness among language speakers.

From the consulted related work, scholars have conducted studies on the use of HLT tools for various purposes such as teaching and learning, as well as exploring language patterns. Some scholars have shown that there are different HLT tools and language resources for South African official languages. According to the researchers' knowledge, however, no study has addressed the HLT tools for SAIL that are presented in this paper to describe and demonstrate their use.

Methodology

In this study, a descriptive qualitative approach was employed to describe some HLT tools for SAIL that can be openly accessed online. Purposive sampling was utilised to select six HLT tools. The researchers first searched online for

HLT tools that support SAIL and found a variety of them. Thereafter they purposefully chose tools that support specific SAIL, focusing on those they believed warranted attention for potential users. The selected tools included a POS tagger, MD, MA, isiZulu.net, ZulMorph and GT. This intentional selection was conducted so that there is representation of a tool that does tagging of word categories, morphological analysis, segmentation of morphemes and automatic translation in the SAIL. Therefore, descriptions of the tools were given in that order. To demonstrate the use of these HLT tools, we provided examples from isiZulu, a majority language, as well as Xitsonga, Siswati and isiNdebele, which are minority languages, and Sesotho Sa Leboa, a language with a relatively high number of speakers but not classified as a majority language.

Theoretical framework

The theoretical underpinning of this study is the technology adoption model (TAM). Straub (2009) explained that the TAM theory examines how people accept and use innovations (technologies) to fully incorporate them into suitable contexts for their benefit. The TAM theory posits that two key factors influence an individual's decision to adopt new technology, namely perceived usefulness and ease of use (Holden & Karsh 2010; Mois & Beer 2020). Perceived usefulness refers to the belief that technology enhances job performance or productivity, while ease of use assumes that if a tool is user-friendly, more users will utilise it (Davis 1989; Durodolu 2016). Therefore, the perceived usefulness of HLT tools for SAIL can be influenced by factors like improved communication, increased access to resources and promotion of language preservation, while ease of use depends on the user interface, the availability of training materials and compatibility (Venkatesh et al. 2003). For this study, the TAM theory offers insights into understanding that language users must be aware of existing HLT tools and their benefits before accepting them.

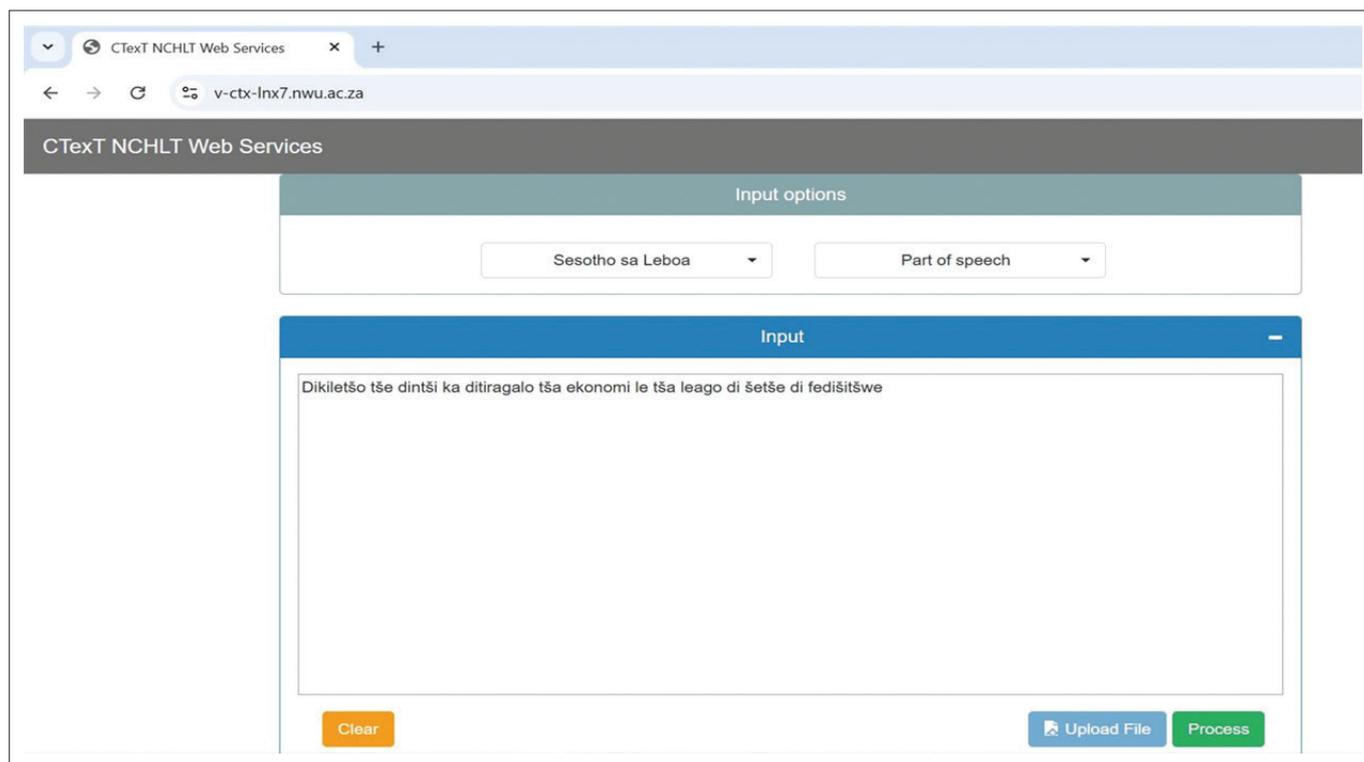
Human language technology tools and their usefulness

This section delves into the intricacies of some HLT tools. It offers a comprehensive exploration of the tools' functionalities and highlights diverse ways in which they contribute to our daily lives in the South African linguistic context. The researchers unravel the profound usefulness of these tools, showcasing their impact across language domains.

Part of speech tagger

A POS tagger is an automated computer application which allocates word categories to all words in a text that it analyses in a specific language (Kumawat & Jain 2015). Antony and Soman (2011) explained that POS tagging involves a:

[P]rocess of assigning to each word in a sentence, a label which indicates the status of that word within some system of categorizing the words of that language according to their morphological and/or syntactic properties. (p. 22)



Source: CText NCHLT Web Services <https://v-ctx-lnx7.nwu.ac.za/>

FIGURE 1: CText NCHLT Web Services POS tagger for Sesotho sa Leboa.

Part-of-speech tagging techniques are classified into three types: statistical tagging, rule-based tagging and hybrid tagging approach. The statistical technique of tagging is based on the most used tag for a certain term in the annotated training data, and this information is utilised to tag that word in the unannotated text (Kumawat & Jain 2015). The rule-based approach necessitates the creation of rules, based on the language structure to assign POS tags to words (Malema et al. 2020).

The hybrid techniques make use of both supervised and unsupervised datasets (Dandapat, Sarkar & Basu 2004). Therefore, POS tagging is essential in the preprocessing task for linguistic text analysis and grammar checking, as this tool predicts word category classes such as nouns, pronouns, adjectives, verbs, adverbs, prepositions and conjunctions based on contextual information (Ekbal, Haque & Bandyopadhyay 2007). Voutilainen (2003:221–222) stated that POS taggers could be advantageous in a variety of tasks:

- More abstract levels of analysis benefit from reliable low-level information (e.g. POS), so a good tagger can serve as a preprocessor.
- Large, tagged text corpora (e.g. British National Corpus; Bank of English Corpus) are used as data for linguistic studies.
- Information technology applications (e.g. text indexing and retrieval) can benefit from POS information (e.g. nouns and adjectives are better candidates for index terms than adverbs, verbs, or pronouns).
- Speech processing benefits from tagging. For instance, the pronoun is pronounced differently from the conjunction.

For the South African context there are two open-access POS taggers for all official languages, excluding South African Sign Language (SASL), which can be used for various purposes. These include automated identification and labelling of word categories to create tagged corpora for various linguistic analysis. These are the National Centre for Human Language Technology (NCHLT) and the CText NCHLT Web Services POS taggers which were developed by CText at the North-West University. Part-of-speech taggers are useful in analysing human language, as users can tag POS in their native and other languages (Antony, Mohan & Soman 2010). For this study, the researchers used the CText NCHLT Web Services POS tagger as shown in Figure 1.

From Figure 2, it is observed that the words in the sentence *Dikiletšo tše dintši ka ditiragalo tša ekonomi le tša leago di šetše di fedišitšwe* [Nearly all restrictions on economic and social activity have already been lifted] have been automatically tagged accurately by the Sesotho sa Leboa CText NCHLT Web Services POS tagger, except the word */tše/*, which is tagged as a demonstrative concord of Class 10 instead of an adjectival concord. This tagger can be utilised as a computer-assisted educational tool to help Sesotho sa Leboa learners to understand the grammatical structure of sentences. The POS tagging can also contribute to the development of other linguistic resources for SAIL. It can provide annotated corpora for developing various NLP applications, including text analysis, machine translation, information retrieval, word sense disambiguation, question answering parsing and information extraction (Albared et al. 2010; Antony et al.

1.The NCHLT Tagger is available at <https://repo.sadilar.org/handle/20.500.12185/351>

Token	Part of speech
Dikiletšo	N10
tše	CD10
dintši	ADJ10
ka	PART
ditiragalo	N10
tša	CPOSS10
ekonomi	N09
le	CONJ
tša	CPOSS10
leago	N05
di	CS10
šetše	VAUX
di	CS10
fedišitšwe	V
.	ŀE

Source: CText NCHLT Web Services <https://v-ctx-lnx7.nwu.ac.za/>

FIGURE 2: CText NCHLT Web Services POS tagger results for Sesotho sa Leboa as an example of a disjunctively written language.



Source: Text Annotation Tools, viewed 15 August 2023, from <https://repo.sadilar.org/handle/20.500.12185/548>

FIGURE 3: Nguni languages morphological decomposer.

2010; Chiche & Yitagesu 2022). Therefore, the use of CText NCHLT Web Services POS tagger for creating annotated corpora can be significant in training and enhancing the accuracy of other HLT tools for some SAIL which are perceived as under-resourced in terms of language resources. Furthermore, linguists and researchers can also use POS tagging to analyse an author's writing style in literature. Bacciu and colleagues (2019) concurred that POS output is effective in identifying syntactic patterns in an author's style. Antony and Soman (2011) also supported that the outputs from a POS tagger can be employed to identify patterns of language which are characteristic of an individual author's style.

Morphological decomposer

An MD is a computational word analysis tool that performs segmentation by identifying a word's components or morphemes (Du Toit & Puttkammer 2021). The availability of MDs for SAIL, such as the one for Nguni languages, could facilitate the use of technology for morphological analysis of these languages. Nguni languages include isiZulu, isiXhosa, isiNdebele and Siswati, which are written conjunctively and are mutually intelligible (Zulu 2013). The MDs were created through the CText core technologies, which include a lemmatiser, POS tagger and an MA for each of the four Nguni languages (Du Toit & Puttkammer 2021). The outputs

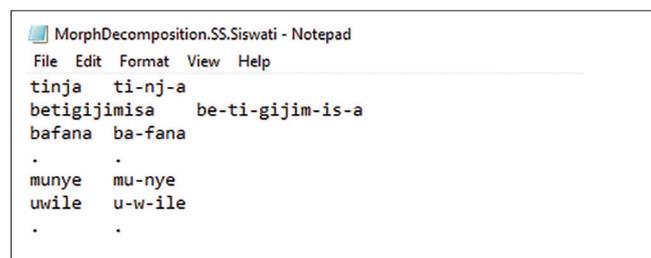
of MDs which show morphemes that form words could then be used by language users who want to know and understand various kinds of morphemes such as prefixes, roots, stems and suffixes. Figure 3 shows an MD for four Nguni languages in which Siswati has been selected as a language to be analysed, and Figure 4 shows the results of two sentences which consist of five words that have been analysed by the MD.

In Figure 4, the Siswati words in the sentence *tinja betigijimisa bafana* [Dogs were chasing boys] are automatically and correctly split by the MD into *ti-nj-a be-ti-gijim-is-a ba-fana*, while in the sentence *munye uwile* [one fell] the words *munye* and *uwile* have been segmented into *mu-nye* and *u-w-ile*. After obtaining these results from the MD, learners can then be given a task to manually label the words' morphemes, that is, for *ti-nj-a* (dogs), /ti-/ is a Class 10 noun prefix, /-nj-/ is a root and /-a/ is a terminative vowel; in the word *be-ti-gijim-is-a*, /be-/ is a past tense marker, /ti-/ is a Class 10 subject concord, /-gijim-/ is a verb root, /-is-/ is a causative verbal extension and /-a/ is a terminative vowel; for *ba-fana* (boys), /ba-/ is a Class 2 noun prefix and /-fana/ is a stem. In the word *mu-nye* (one), /mu-/ is a Class 1 enumerative concord, and /-nye/ is an enumerative stem, while in the verb *u-w-ile* (he fell), /u-/ is a Class 1 subject concord, /-w-/ is a verb radical and /-ile/ is a past tense marker. From the MD results,

we noted how language users can use the MD to automatically obtain morphological analysis of words. The MD is advantageous over manual segmentation and analysis because it can swiftly process many words. Linguists can then clean obtained data in case of minor errors. This process can facilitate the creation of a large morphologically annotated corpus for SAIL that can be used for research purposes.

Morphological analyser

Suriyah and colleagues (2020) described an MA as an HLT application that takes inflected, derived or compound forms of words as inputs to retrieve numerous grammatical aspects of language such as word(s) root, prefixes and suffixes. It also recognises, segments and assigns grammatical information to the input words while producing their grammatical structures as outputs (Akilan & Naganathan 2014). In contrast to the MD tool, the MA tool recognises morphemes and assigns tags depending on their grammatical functions (Eiselen & Puttkammer 2014). The MA is useful for languages with high morpheme density such as Nguni languages. Such a tool can be beneficial in giving morphological descriptions of these languages.

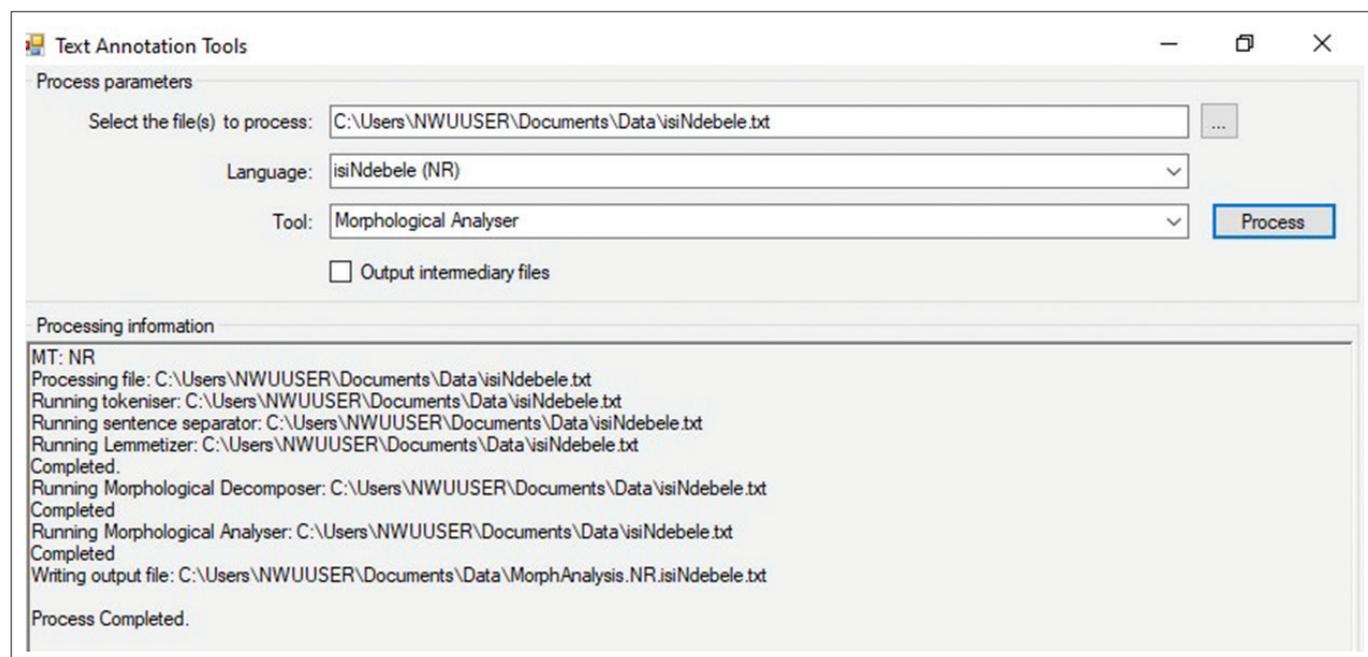


Source: Text Annotation Tools, viewed 15 August 2023, from <https://repo.sadilar.org/handle/20.500.12185/548>

FIGURE 4: Morphological decomposer output for Siswati as an example of a conjunctively written language.

Morphological analysers are significant not only for identifying word morphemes or for conducting accurate corpus searches, but also as basic tools that can be used to facilitate the development of other HLT tools such as tokenisers, POS taggers, parsers and machine translation systems (Abudouwaili et al. 2021; Bosch 2020; Zinkevičius, Daudaravičius & Rimkutė 2005). The MA can also be used to semi-automatically create a morphologically annotated corpus for SAIL. Figure 5 shows the MA tool for Nguni languages in which isiNdebele has been selected as a language to be analysed, and Figure 6 shows the results.

In Figure 6, most of the words in the sentence *izulu lentando yenengi lilethe ukuvuseleleka nokubelethwa kavesitjhaba esitjha* [the rain of democracy has brought the renewal and birth of the new nation] have been morphologically analysed correctly. The noun *izulu* [rain] consists of the noun Class 5 preprefix /i-/ and the noun stem /-zulu/. Such results can be used to get an insight into various morphemes that make up nouns in isiNdebele. From the word *lentando* [of democracy], language users would be aware that the possessive is made up of the possessive concord /le-/, which was formed when the vowel /a-/ of the possessive concord /la-/ for Class 5 nouns assimilated with the preprefix /i-/ to form the possessive concord /le-/ (viz. la- + i- > le-), and the basic prefix /-n-/ as well as the noun stem /t(h)ando/. The MA can also be used to demonstrate that these morphemes have marked meanings. For instance, the morpheme /ye-/ means 'of' in the word *yenengi*. However, such tools do not always produce accurate results; for example, the adjective *esitjha* [the new] has been inaccurately analysed because the /esi-/ is tagged by MA as a relative concord instead of an adjectival concord, while /-tjha/ has been tagged as a noun stem, yet it is an adjectival stem. Regardless of this shortcoming, language users can benefit by being given all



Source: Text Annotation Tools, viewed 15 August 2023, from <https://repo.sadilar.org/handle/20.500.12185/548>

FIGURE 5: Morphological analyser for Nguni languages.

```

MorphAnalysis.NR.ndebele test - Notepad
File Edit Format View Help
izulu i[NPrePre]-zulu[NStem]
lentando le[PossConc]-i[NPrePre]-n[BPre]-thando[NStem]
yenengi ya[PossConc]-e[AdjPref]-nengi[AdjStem]
lilethe li[SC]-leth[VRoot]-ile[Perf]
ukuvuseleleka u[NPrePre]-ku[BPre]-vuseleleka[NStem]
nokubelethwa na[AdvPre]-u[NPrePre]-ku[BPre]-beleth[VRoot]-w[PassExt]-a[VerbTerm]
kwesitjhaba kwe[PossConc]-si[BPre]-tjhaba[NStem]
esitjha esi[RelConc]-tjha[NStem]
. .[Punc]

```

Source: Text Annotation Tools, viewed 15 August 2023, from <https://repo.sadilar.org/handle/20.500.12185/548>

FIGURE 6: Output for isiNdebele morphological analyser.

Finite state morphology

About ZulMorph

ZulMorph demo

Zulu language

Zulu morphology

ZulMorph tagset

References

Citing output

Contact us

ZulMorph demo

Version: 20190625

Type one or more Zulu words as input to the analyzer, for example "abasiki abadala bazozithunga izingubo".

All input is processed as lowercase text. For more information, please consult the [documentation](#).

Input:

Submit

thina sonke lapha siyahamba

- thi[PronStem][1pp]na[PronSuf]
- so[QC][1pp]nke[QuantStem]
- so[QC][7]nke[QuantStem]
- lapha[Adv]
- lapha[Conj]
- lapha[Dem][16][Pos1]
- la[SCPT][5]ph[VRoot]a[VT]
- si[SC][1pp]ya[LongPres]hamb[VRoot]a[VT]
- si[SC][7]ya[LongPres]hamb[VRoot]a[VT]

Source: ZulMorph Demo, computer software, viewed 13 August 2023, from <https://portal.sadilar.org/FiniteState/demo/zulmorph/doc.html#about>

FIGURE 7: Output for Zulu ZulMorph.

the results so that they can analyse and deduce whether the words were tagged correctly or incorrectly. This will allow them to provide an accurate analysis manually. Such tools will also promote and strengthen their understanding of morphological perspectives of indigenous languages.

ZulMorph

ZulMorph is a morphological finite state analyser for isiZulu in which base forms of words are determined by analysing their surface form. ZulMorph² is a computational application; the Xerox Finite-State Tool and Finite-State Lexicon Compiler were used to create this analyser, and it conforms with Foma. The advantage of ZulMorph over other MAs developed for SAIL is that it does not only automatically segment words

²See <https://portal.sadilar.org/FiniteState/demo/zulmorph/doc.html#about>

into their base form, but it can also show some homonymy, as shown in Figure 7.

From Figure 7, it is observed that, unlike the MDs and MAs, ZulMorph has done more than a morphological analysis. It has also done POS tagging for the word *lapha*. The POS tagging shows that the word *lapha* is homonymous. This word belongs to three distinct POS in isiZulu, depending on the context in which it has been employed: it can be used as an adverb of place, a conjunction and a demonstrative pronoun. Through the automated results obtained from ZulMorph, one can ascertain that the word *lapha* belongs to these three types of word categories. However, a fourth possible analysis of *lapha* has been given in which /la-/ is a past tense subject concord of Class 5 nouns, /-ph-/ is a verb root and /-a/ is a verb terminative. Therefore, language

users can deduce the basic forms of the words in the sentence and realise that some words are homonymous in the language.

IsiZulu.net

IsiZulu.net is a bilingual online dictionary that supports isiZulu and English. This dictionary can be accessed freely at <https://isizulu.net/>. According to Jin and Deifell (2013), online dictionaries facilitate learning outside and inside the classroom environment. The role of online dictionaries is to assist language users in broadening their vocabularies and communicating effectively (Lan 2005). IsiZulu.net gives a basic overview of the most important isiZulu grammatical features, such as morphological analysis, conjugation, word equivalents, spell checking, pronunciation, noun classes, concords and transcribing isiZulu words phonetically (Klein 2012). The main linguistic advantage of isiZulu.net is that it is comprehensive because no stem search is required, and it automatically provides morphological analysis of a word, its stem, prefix and suffix (Prinsloo 2011). IsiZulu.net also has a forum where users can participate in general discussions, submit new lemmas, suggest additional translations and report errata such as incorrect translations and typos (Prinsloo 2012). Bosch (2020) concurred that isiZulu.net can be used for learning purposes such as learning isiZulu grammar, checking spelling and translation. Based on the usefulness of this dictionary, language users can utilise it to find isiZulu word equivalents in English and to learn isiZulu grammar, as presented in Figure 8.

In Figure 8, the isiZulu word *ubuhlakani* has been correctly translated into English as 'cleverness,' 'craftiness,' 'smartness' and 'subtlety.' This example demonstrates that *ubuhlakani* has more than one equivalent in English, which would assist language learners in developing English vocabulary. In this Figure, it is also observed that the dictionary indicates that the word *ubuhlakani* is a noun in Class 14 and its stem is *-hlakani*/. This grammatical information is useful to individuals learning isiZulu as it assists them in realising word categories and their stems.

Moreover, isiZulu.net plays an important role in learning pronunciations of International Phonetic Alphabets (IPA). Klein (2012) affirmed that isiZulu.net is an ideal online dictionary because of its ability to offer IPA and its pronunciations. Therefore, language users can learn how

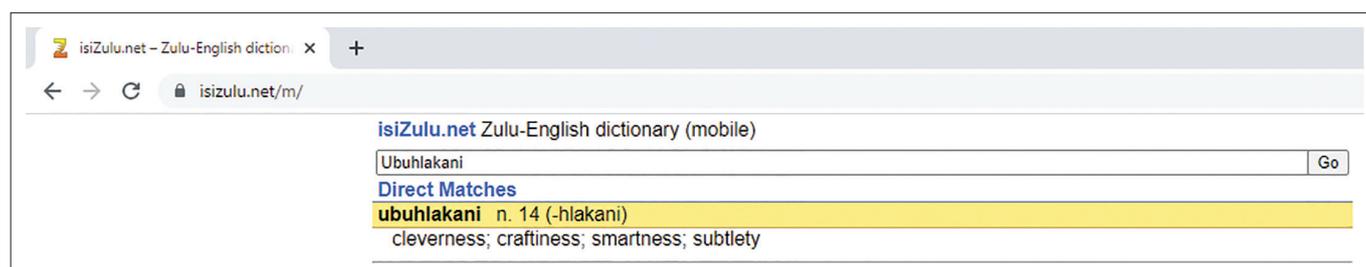
the IPA is pronounced using both isiZulu and English words by listening to audio recordings, as shown in Figure 9.

Figure 9 shows the IPA symbols with isiZulu and English examples, in which audio pronunciations can be played and phonetic transcriptions of each word are provided. Using IPA audio pronunciations, learners can pronounce words correctly. Setter and Jenkins (2005) supported that the IPA system is useful because it indicates the way in which sounds should be produced accurately. Trinh, Nguyen and Le (2022) asserted that language learners benefit from phonetic transcriptions as they learn the language easily, efficiently and precisely. Therefore, isiZulu.net is an important dictionary which language users can use to improve their pronunciation skills.

Google Translate

Google Translate is a free and open online neural translation machine service offered by Google LLC. This program is available in over 100 languages globally and can translate written words, websites, documents and speech from a source language to a target language (Larassati et al. 2019). The texts or corpora used to develop GT come from diverse books, organisations such as the United Nations, and the Internet (Lotz & Van Rensburg 2014). Among the 133 languages in GT, seven are official languages of South Africa, namely Afrikaans, English, Southern Sotho [Sesotho], isiZulu, isiXhosa, Sepedi and Xitsonga (Illidge 2022). Users use GT as a learning tool, particularly for language acquisition, because of its useful features such as translation of multiple languages, time-saving, effortless ease of use and enhancement of the basic pronunciation of words based on the languages involved (Pham et al. 2022). In the context of SAIL, GT is beneficial to users who want to acquire linguistic aspects such as the pronunciation of words, synonyms, translation equivalents, translation variations and sentence reformulations. Figure 10 shows features that can be used to learn linguistic components.

From Figure 10, it is observed that the sentence 'The fuel price will be reduced on Wednesday' was translated from English into Xitsonga by this HLT tool. The GT speech recognition feature may help with language acquisition as it allows users to translate spoken words and phrases. This can assist users in accurately pronouncing words and phrases in the source language so that accurate translation is achieved



Source: Online Dictionary, 2025, *isiZulu.net Zulu-English dictionary*, viewed 23 October 2023, from <https://isizulu.net/>

FIGURE 8: IsiZulu.net translation.

isiZulu.net
zulu-english dictionary

Zulu pronunciation guide

The phonetic transcriptions in this dictionary use the phonetic alphabet of the IPA. They are generated on-the-fly by an algorithm I developed based on the pronunciation guide of the *Speak Zulu With Us* beginners course from African Voices.

The following table shows examples of how to pronounce each of the IPA symbols in the context of this dictionary. For each symbol there are Zulu and English example words with sound recordings and a link to the Wikipedia article on the IPA symbol. I chose the English example recordings with respect to matching the Zulu pronunciation as close as possible. They do *not* correspond as to the meanings nor do they necessarily conform to your personal way of pronouncing them. So don't blame me for that. :-)

List of sources

- IPA recordings.
- By permission. From the Merriam-Webster online dictionary (www.merriam-webster.com).
- By permission. From the African Voices *Speak Zulu With Us* beginners course (www.africanvoices.co.za).

IPA symbol	Zulu example	English example (approx.)
[]	The following syllable is stressed.	
[]	The preceding vowel is lengthened.	
[a]	amakhaza [ama'kʰa:za]	buck [bak]
[e]	umgibeli [umgi'be:li]	bayed [bejd]
[e]	ihembe [i'he:mbe]	bed [bed]
[i]	isibili [isi'bi:li]	bead [bi:d]
[o]	iGoli [i'go:li]	soil [soil]

Source: Online Dictionary, 2025, isiZulu.net Zulu-English dictionary, viewed 23 October 2023, from <https://isizulu.net/p11n/>

FIGURE 9: International Phonetic Alphabets symbol pronunciation.

Google Translate

Text Documents Websites

DETECT LANGUAGE ENGLISH ZULU TSONGA

term equivalence

TSONGA ZULU ENGLISH

The fuel price will be reduced on Wednesday

speech recogniser

43 / 5,000

Ntsengo wa mafurha wu ta hungutiwa hi Ravunharhu

Ravunharhu

synonyms

- ✓ Ntsengo wa mafurha wu ta hungutiwa hi Ravunharhu
The price of fuel will be reduced on Wednesday
- Ntsengo wa petirolo wu ta hungutiwa hi Ravunharhu
The price of gasoline will be reduced on Wednesday

Send feedback

Source: Google Translate, English-Tsonga, computer software, viewed 15 November 2023, from <https://translate.google.com/?sl=en&tl=ts&op=translate>

FIGURE 10: Google Translate results for Xitsonga.

in the target language. If words are pronounced incorrectly, the system will not be able to recognise them. Howard and Messum (2014) stated that being able to pronounce words correctly is an essential first step in learning a language. Therefore, GT speech recognition can be used to support the learning of proper pronunciation. There is also a text-to-speech feature that may be used for language acquisition in English and Afrikaans. However, other SAIL are not yet supported by this feature.

Google Translate provides translations from a source to a target language for self-learning of synonyms and translation equivalents in the target language. As illustrated in Figure 10, GT has offered synonyms of 'fuel' as 'gasoline' in English and *mafurha* as *petirolo* in Xitsonga. From these results, language users can learn synonyms and translation equivalents, which are beneficial for language learning. Such information can also assist in lexicography, the process of compiling dictionaries or glossaries. Moreover, GT facilitates multilingualism

through its user-friendly interface and language support, encouraging cultural appreciation and exchange among users as it allows any language to be used as a source language. Therefore, language users can use GT to translate text into their desired language. Furthermore, GT can also play a crucial role in breaking down language barriers and fostering communication between individuals who speak different languages. Ahmad (2019:64) stated that 'language barriers are the root causes of many problems or obstacles in health care, aviation, maritime, business, and education'. This can have detrimental impacts on individuals' well-being, social development and economic prosperity. The use of GT can assist in overcoming these challenges by translating the source text into the native user language. It is also imperative to note that GT may not always provide perfect translations, especially for complex or context-dependent content, but it can provide a gist for understanding the context.

Conclusion

The study has shown that open-access HLT tools, namely CTeX NCHLT Web Services POS tagger, Nguni Languages MD, Nguni Languages MA, isiZulu.net, ZulMorph and GT can play an important role in understanding SAIL. These tools can be used to learn and gain insight into various linguistic aspects of SAIL such as word categories, their formation and pronunciation, synonyms, vocabulary expansion and translation equivalents. The study has ascertained that some of the HLT tools can be used effectively to perform automated linguistic analysis which can create annotated corpora. These corpora are significant in developing and improving the accuracy of already existing HLT tools for SAIL. The use of some tools such as isiZulu.net and GT to acquire linguistic aspects can also contribute to the promotion of multilingualism in South Africa. Therefore, the HLT tools discussed in this study are essential in the development of SAIL in the HLT field, and language users are encouraged to utilise them.

Acknowledgements

We would like to thank Prof Koch for his help and advice in finalising this article.

Competing interests

The author reported that they received funding from the South African Centre for Digital Language Resources (SADiLaR) which could have affected the research reported in the enclosed publication. The author has disclosed those interests fully and has implemented an approved plan for managing any potential conflicts arising from their involvement. The terms of these funding arrangements have been reviewed and approved by the affiliated University in accordance with its policy on objectivity in research.

Authors' contributions

All authors contributed equally to the writing of this research article.

Ethical considerations

This article does not contain any studies involving human participants performed by any of the authors.

Funding information

This publication was made possible with the support from the South African Centre for Digital Language Resources (SADiLaR), a research infrastructure established by the Department of Science and Innovation of the South African government as part of the South African Research Infrastructure Roadmap (SARIR).

Data availability

Data sharing is not applicable to this article as no new data were created or analysed in this study.

Disclaimer

The views and opinions expressed in this article are those of the authors and are the product of professional research. It does not necessarily reflect the official policy or position of any affiliated institution, funder, agency or that of the publisher. The authors are responsible for this article's results, findings and content.

References

- Abbott, J. & Martinus L., 2019, 'Benchmarking neural machine translation for Southern African languages', in A. Axelrod, D. Yang, R. Cunha, S. Shaikh & Z. Waseem (eds.), *The association for computational linguistics 2019 workshop on widening natural language processing*, Florence, Italy, July 28, 2019, pp. 98–101.
- Abudouwaili, G., Abiderexiti, K., Wushouer, J., Shen, Y., Maimaitimin, T. & Yibulayin, T., 2021, 'Morphological analysis corpus construction of Uyghur', in S. Li, M. Sun, Y. Liu, H. Wu, L. Kang, W. Che, S. He & G. Rao (eds.), *Chinese computational linguistics*, pp. 1076–1086, Springer, Cham.
- Adefila, E., Oladokun, B.D. & Adewojo, A.A., 2024, *Beginning of a new era: Leveraging digital twin technology in the preservation of indigenous knowledge systems in the disruptive era*, viewed 08 October 2024, from <https://www.emerald.com/insight/content/doi/10.1108/lhnt-05-2024-0088/full/pdf>.
- Ahmad, A.I., 2019, 'Language barriers to effective communication', *Utopia y Praxis Latinoamericana* 24, 64–76.
- Akilan, R. & Naganathan, E.R., 2014, 'Morphological analyzer for classical Tamil texts: A rule-based approach', *International Journal of Innovative Science, Engineering & Technology* 1(5), 563–568.
- Albared, M., Omar, N., Aziz, M.J.A. & Ahmad Nazri, M.Z., 2010, 'Automatic part of speech tagging for Arabic: An experiment using Bigram hidden Markov model', in J. Yu, S. Greco, P. Lingras, G. Wang & A. Skowron (eds.), *Rough set and knowledge technology*, pp. 361–370, Springer, Berlin.
- Alhawiti, K.M., 2014, 'Natural language processing and its use in education', *International Journal of Advanced Computer Science and Applications* 5(12), 72–76. <https://doi.org/10.14569/IJACSA.2014.051210>
- Allam, Z., Sharifi, A., Bibri, S.E., Jones, D.S. & Krogstie, J., 2022, 'The metaverse as a virtual form of smart cities: Opportunities and challenges for environmental, economic, and social sustainability in urban futures', *Smart Cities* 5(3), 771–801.
- Antony, P.J. & Soman, K.P., 2011, 'Parts of speech tagging for Indian languages: A literature survey', *International Journal of Computer Applications* 34(8), 22–29.
- Antony, P.J., Mohan, S.P. & Soman, K.P., 2010, 'SVM based part of speech tagger for Malayalam', paper presented at the International Conference on Recent Trends in Information, Telecommunication and Computing, Information, Telecommunication and Computing (ITC), Kerala, 12–13th March.
- Bacciu, A., La Morgia, M., Mei, A., Nemmi, E.N., Neri, V. & Stefa, J., 2019, 'Cross-domain authorship attribution combining instance based and profile-based features', in L. Cappellato, N. Ferro, D.E. Losada & H. Müller (eds.), *CLEF 2019 labs and workshops, notebook papers*, Lugano, Switzerland, September 9–12, 2019, pp. 1–14.
- Bosch, S., 2020, 'Computational morphology systems for Zulu – A comparison', *Nordic Journal of African Studies* 29(3), 1–28.
- Buabeng-Andoh, C., 2012, 'Factors influencing teachers' adoption and integration of information and communication technology into teaching: A review of the literature', *International Journal of Education and Development using Information and Communication Technology* 8(1), 136–155.

- Chiche, A. & Yitagesu, B., 2022, 'Part of speech tagging: A systematic review of deep learning and machine learning approaches', *Journal of Big Data* 9(10), 1–25. <https://doi.org/10.1186/s40537-022-00561-y>
- Cloete, A.L., 2017, 'Technology and education: Challenges and opportunities', *HTS: Theological Studies* 73(3), 1–7. <https://doi.org/10.4102/hts.v73i3.4589>
- CTeX NCHLT Web Services, *computer software*, viewed 23 October 2023, from <https://hlt.nwu.ac.za/>.
- Dandapat, S., Sarkar, S. & Basu, A., 2004, 'A hybrid model for part-of-speech tagging and its application to Bengali', paper presented at the International Conference on Computational Intelligence, International Conference on Computational Intelligence Society (ICCS), Istanbul, 17–19th December.
- Davis, F.D., 1989, 'Perceived usefulness, perceived ease of use, and user acceptance of information technology', *MIS Quarterly* 13(3), 319–340. <https://doi.org/10.2307/249008>
- Du Toit, J.S. & Puttkammer, M.J., 2021, 'Developing core technologies for resource-scarce Nguni languages', *Information* 12(12), 1–12. <https://doi.org/10.3390/info12120520>
- Durodolu, O., 2016, 'Technology acceptance model as a predictor of using information system to acquire information literacy skills', *Library Philosophy & Practice* 1450, 1–27, viewed 23 October 2023, from <http://digitalcommons.unl.edu/libphilprac/1450>.
- Eiselen, R. & Puttkammer, M.J., 2014, 'Developing text resources for ten South African Languages', in N. Calzolari, K. Choukri, T. Declerck, H. Loftsson, B. Maegaard, J. Mariani, A. Moreno, J. Odijk & S. Piperidis (eds.), *The 9th International Conference on Language Resources and Evaluation*, Reykjavik, Iceland, May 26–31, 2014, pp. 3698–3703.
- Ekbal, A., Haque, R. & Bandyopadhyay, S., 2007, 'Bengali part of speech tagging using conditional random field', paper presented at the 7th international symposium on natural language processing, a network of excellence forging the multilingual Europe technology alliance, Pattaya, 13–15th December.
- Google Translate, English-Tsonga, *computer software*, viewed 15 November 2023, from <https://translate.google.com/?sl=en&tl=ts&op=translate>.
- Grover, A.S., Van Huyssteen, G.B. & Pretorius, M.W., 2011, 'The South African human language technology audit', *Language Resources and Evaluation* 45(3), 271–288. <https://doi.org/10.1007/s10579-011-9151-2>
- Heeringa, W., De Wet, F. & Van Huyssteen, G.B., 2015, 'Afrikaans and Dutch as closely related languages: A comparison to west Germanic languages and Dutch dialects', *Stellenbosch Papers in Linguistics Plus* 47, 1–18. <https://doi.org/10.5842/47-0-649>
- Holden, R.J. & Karsh, B.T., 2010, 'The technology acceptance model: Its past and its future in health care', *Journal of Biomedical Informatics* 43(1), 159–172. <https://doi.org/10.1016/j.jbi.2009.07.002>
- Howard, I.S. & Messum, P., 2014, 'Learning to pronounce first words in three languages: An investigation of caregiver and infant behavior using a computational model of an infant', *PLoS One* 9(10), 1–21. <https://doi.org/10.1371/journal.pone.0110334>
- Illidge, M., 2022, 'Google translate adds two new South African languages', *MyBroadband*, viewed 18 September 2023, from <https://mybroadband.co.za/news/internet/444186-google-translate-adds-two-new-south-african-languages>.
- isiZulu.net, *computer software*, viewed 23 October 2023, from <https://isizulu.net/>.
- Jin, L. & Deifell, E., 2013, 'Foreign language learners' use and perception of online dictionaries: A survey study', *Journal of Online Learning and Teaching* 9(4), 515–533.
- Klein, J., 2012, 'Luxdico: A Lëtzebuergesch – German online dictionary: What could it learn from South African online dictionaries?', paper presented at the 17th African association for lexicography international conference, University of Pretoria, Pretoria, 2–5th July.
- Koehn, P. & Knowles, R., 2017, 'Six challenges for neural machine translation', paper presented in first workshop on neural machine translation, Vancouver, Canada, 30 July–04 August 2017.
- Kumawat, D. & Jain, V., 2015, 'POS tagging approaches: A comparison', *International Journal of Computer Applications* 118(6), 32–38. <https://doi.org/10.5120/20752-3148>
- Lan, L., 2005, 'The growing prosperity of online dictionaries', *English Today* 21(3), 16–21. <https://doi.org/10.1017/S0266078405003044>
- Larassati, A., Setyaningsih, N., Nugroho, R.A., Suryaningtyas, V.W., Cahyono, S.P. & Pamelasari, S.D., 2019, 'Google vs. Instagram machine translation: Multilingual application program interface errors in translating procedure text genre', paper presented at the international seminar on application for technology of information and communication, Universitas Dian Nusawantoro, Semarang, 21–22nd September.
- Lotz, S. & Van Rensburg, A., 2014, 'Translation technology explored: Has a three-year maturation period done Google Translate any good?', *Stellenbosch Papers in Linguistics Plus* 43, 235–259. <https://doi.org/10.5842/43-0-205>
- Malema, G., Okgetheng, B., Tebalo, B., Motlanka, M. & Rammidi, G., 2020, 'Complex Setswana parts of speech tagging', paper presented at the first workshop on resources for African indigenous languages, South African Centre for Digital Language Resources, Marseille, 11–16th May.
- Mlambo, R. & Matfunjwa, M., 2022, 'Using MonoConc pro to teach and learn Lexical collocations in Xitsonga', *Journal of the Digital Humanities Association of Southern Africa* 3(3), 1–7. <https://doi.org/10.55492/dhasa.v3i03.3821>
- Mlambo, R., Skosana, N. & Matfunjwa, M., 2021, 'The extraction of terminology list using ParaConc for creating a quadrilingual dictionary', *Southern African Linguistics and Applied Language Studies* 39(1), 82–91. <https://doi.org/10.2989/16073614.2021.1896971>
- Mois, G. & Beer, J.M., 2020, 'Robotics to support aging in place', in R. Pak, E.J. De Visser & E. Rovira (eds.), *Living with robots: Emerging issues on the psychological and social implications of robotics*, pp. 49–74, Academic Press, London.
- NCHLT Tagger, *computer software*, viewed 15 November 2023, from <https://repo.sadilar.org/handle/20.500.12185/351>.
- Online Dictionary, 2025, *isiZulu.net Zulu-English dictionary*, viewed 23 October 2023, from <https://isizulu.net/>
- Pham, A.T., Nguyen, Y.N., Tran, L.T., Huynh, K.D., Le, N.T.K. & Huynh, P.T., 2022, 'University students' perceptions on the use of Google translate: Problems and solutions', *International Journal of Emerging Technologies in Learning* 17(4), 74–94. <https://doi.org/10.3991/ijet.v17i04.28179>
- Plomp, T., Anderson, R.E., Law, N. & Quale, A. (eds), 2009, *Cross-national information and communication technology: Policies and practices in education*, Information Age Publishing, Charlotte, NC.
- Popenici, S.A. & Kerr, S., 2017, 'Exploring the impact of artificial intelligence on teaching and learning in higher education', *Research and Practice in Technology Enhanced Learning* 12(1), 1–13. <https://doi.org/10.1186/s41039-017-0062-8>
- Prinsloo, D.J., 2011, 'A critical analysis of the lemmatisation of nouns and verbs in isiZulu', *Lexikos* 21(1), 169–193. <https://doi.org/10.5788/21-1-42>
- Prinsloo, D.J., 2012, 'Electronic lexicography for lesser-resourced languages: The South African context', in S. Granger & M. Paquot (eds), *Electronic lexicography*, pp. 119–144, Oxford University Press, Oxford.
- Puttkammer, M., Eiselen, R., Hocking, J. & Koen, F., 2018, 'NLP web services for resource-scarce languages', in F. Liu & T. Solorio (eds.), *Proceedings of association for computational linguistics*, Melbourne, Australia, July 15–20, 2018, pp. 43–49.
- Rashel, M.M., 2011, 'Introducing language technology and computational linguistics in Bangladesh', *International Journal of English Linguistics* 1(1), 179–186. <https://doi.org/10.5539/ijel.v1n1p179>
- Rivera Pastor, R., Tarín Quirós, C., Villar García, J.P., Badía Cardús, T. & Meleru Nogués, M., 2017, *Language equality in the digital age-towards a human language project*, viewed 12 October 2023, from <https://op.europa.eu/en/publication-detail/-/publication/fa0a50e7-cda4-11e7-a5d5-01aa75ed71a1/language-en>.
- Rogers, E.M., 2003, *Diffusion of innovations*, 5th edn., The Free Press, New York, NY.
- Roux, J. & Bosch, S., 2006, 'Language resources and tools in Southern Africa', paper presented at the Fifth international conference on language resources and evaluation, European Language Resources Association, Genoa, 22–28th May.
- Schiler, J., 2003, 'Working with ICT: Perceptions of Australian principals', *Journal of Educational Administration* 41(3), 171–185. <https://doi.org/10.1108/09578230310464675>
- Setter, J. & Jenkins, J., 2005, 'State of the art review article: Pronunciation', *Language Teaching* 38(1), 1–17. <https://doi.org/10.1017/S026144480500251X>
- Skosana, N.J. & Mlambo, R., 2021, 'A brief study of the Autshumato machine translation web service for South African languages', *Literator* 42(1), a1766. <https://doi.org/10.4102/lit.v42i1.1766>
- Straub, E.T., 2009, 'Understanding technology adoption: Theory and future directions for informal learning', *Review of educational research* 79(2), 625–649. <https://doi.org/10.3102/0034654308325896>
- Suriyah, M., Anandan, A., Narasimhan, A. & Karky, M., 2020, 'Piripori: Morphological analyser for Tamil', in L. Kumar, L. Jayashree & R. Manimegalai (eds.), *Proceedings of international conference on artificial intelligence, smart grid and smart city applications*, Coimbatore, India, January 03–05, 2019, pp. 801–809.
- Tachbelie, M.Y., Abate, S.T. & Besacier, L., 2011, 'Part-of-speech tagging for under-resourced and morphologically rich languages – The case of Amharic', paper presented at the international conference on human language technology for development, Bibliotheca Alexandrina, Alexandria, 2nd – 5th May.
- Text Annotation Tools, viewed 15 August 2023, from <https://repo.sadilar.org/handle/20.500.12185/548>
- Trinh, Q.L., Nguyen, T.D.L. & Le, T.T., 2022, 'Using explicit instruction of the international phonetic alphabet system in English as a foreign language adult classes', *European Journal of Educational Research* 11(2), 749–761. <https://doi.org/10.12973/eurj.11.2.749>
- Van Huyssteen, G.B., & Griesel, M., 2016, 'Translation technology in South Africa', in C. Sin-wai (ed.), *The Routledge encyclopedia of translation technology*, pp. 327–336, Routledge, New York, NY.
- Van Huyssteen, G.B., Puttkammer, M., McKellar, C. & Griesel, M., 2023, 'Translation technology in South Africa', in C. Sin-wai (ed.), *Routledge encyclopedia of translation technology*, 2nd edn., pp. 373–383, Routledge, London.
- Van Zaenen, M., Trollip, B., Ramukhadi, P.M. & Mlambo, R., 2020, 'Identifying relations between characters in Afrikaans, Tshivenda, and Xitsonga books', in *annual conference of the Alliance of Digital Humanities Organizations (ADHO)*, July 20–25, 2020, Ottawa, Canada, viewed 24 September 2020, from <https://hcommons.org/deposits/item/hc:32053/>.
- Venkatesh, V., Morris, M.G., Davis, G.B. & Davis, F.D., 2003, 'User acceptance of information technology: Toward a unified view', *MIS Quarterly* 27(3), 425–478. <https://doi.org/10.2307/30036540>
- Voutilainen, A., 2003, 'Part-of-speech tagging', in R. Mitkov (ed.), *The Oxford handbook of computational linguistics*, pp. 219–232, Oxford University Press, Oxford.
- Zinkevičius, V., Daudaravičius, V. & Rimkutė, E., 2005, 'The morphologically annotated Lithuanian corpus', in M. Langemets & P. Penjam (eds.), *Second Baltic conference on human language technologies: Proceedings*, Tallinn, Estonia, April 4–5, 2005, pp. 365–370.
- ZulMorph Demo, *computer software*, viewed 13 August 2023, from <https://portal.sadilar.org/FiniteState/demo/zulmorph/doc.html#about>.
- Zulu, P.N., 2013, 'Classification of South African languages using text and acoustic based methods: A case of six selected languages', in L. Vanderwende, H. Daumé III & K. Kirchoff (eds.), *Proceedings of the 2013 conference of the North American chapter of the association for computational linguistics: Human language technologies*, Atlanta, Georgia, June 9–14, 2013, pp. 280–287.