Exploring the Impact of Centralised Translation Memory in OmegaT Translation Software

The Case of the Xitsonga Language

Hlamalani Dollence Baloyi
Department of African Languages
University of South Africa
Pretoria, South Africa
dollenzo@gmail.com

Abstract

The role of Information Communication Technology (ICT) is crucial in language development. As the technological era evolves, it is important for the language space to keep up with current trends. Computer Assisted Translation (CAT) tools are still inadequately designed for African indigenous languages, which hinders their precision, usage, and digital participation, ultimately perpetuating linguistic disparities. Using a corpus-based translation approach, the study aims to explore the impact of centralised translation memory in OmegaT translation software, specifically in the Xitsonga language. CAT applications are used to enhance, accelerate, enrich, and deepen language skills. A qualitative research approach was adopted which comprised desktop research of secondary materials. A comparative analysis was used for analysing data as it involves a direct comparison that methodically examines the OmegaT translation memories and translated documents to identify their similarities and differences. Inconsistencies were noted in the terminologies used by different translators due to the absence of centralisation of translation memories and glossaries. The lack of translation equivalents in the standardised terminology list indicates that the terms have not been included for development; it is essential to welcome new words into the corpus and offer definitions for these terms to support the evolution of language and terminology. Centralised translation memory could lead to reduced translation time and quicker turnaround for translation projects. It is recommended that OmegaT should include a sharing feature or a repository for storing translated terms, and this should be facilitated by Department of Sport, Arts and Culture, as they carry the mandate.

Keywords: OmegaT, Translation Equivalent, CAT Tool, Software, Terminology, Glossary

Introduction

Although the global use of Computer-Assisted Translation (CAT) tools is increasing, the incorporation and centralisation of translation memory (TM) systems for African indigenous languages, especially Xitsonga is still insufficiently advanced and researched. OmegaT CAT tool offers TM centralisation capabilities that allow translators to effectively share, and reuse translated segments. Even though that is the case, its implementation in

© 2025. Author(s)

Xitsonga translation workflows encounters various obstacles. The introduction of Machine Translation (MT), Computer-Assisted Translation (CAT) tool, and Artificial Intelligence (AI) in the language landscape is a major development since technology is evolving and gaining momentum daily. Efforts to preserve languages are often multifaceted, involving documentation, educational programs, community initiatives, and increasingly the use of technology (Sharon, 2024). Different software programs have emerged in the language field to improve, speed up, enrich, and intensify language abilities. These computer applications are called Computer-Assisted Translation (CAT) tools, some refer to them as Computer-Aided Translation or Computer-Aided Human Translation (CAHT). CAT falls under an interdisciplinary field in language studies called Computational linguistics, the study of computer systems for understanding and generating natural language (Grishman, 1986).

The application of Computer-Assisted Translation (CAT) tools has progressed notably in multilingual settings, yet their efficient incorporation of African indigenous languages, such as Xitsonga, is still insufficiently examined, especially concerning the centralisation of translation memory (TM) in tools like OmegaT application. Current research on CAT tools in Africa has mainly centred on user adoption, interface design, and overall translation effectiveness with minimal exploration of the linguistic and technical functioning of centralised TMs in languages with fewer resources. A significant gap is present in the access to and use of Xitsonga linguistic resources, including aligned bilingual corpora and standardised terminology databases, which are crucial for creating and sustaining dependable translation memories. In the absence of these, centralised TM systems cannot guarantee accuracy, consistency, or terminological uniformity. Additionally, prior studies have not empirically investigated how Omega's TM centralisation addresses Xitsonga-specific linguistic features, dialectal differences, or culturally rooted expressions. A notable gap exists in assessing the quality of translations generated by centralised TM systems in Xitsonga.

Throughout the years, CAT tools have developed to provide numerous features in addition to Translation Memory. They have begun to incorporate terminology management, glossaries, machine translation integration, quality assurance checks, and collaborative features, along with various other functionalities. CAT tools have continued to evolve, integrating artificial intelligence and machine learning technologies to improve translation precision, increase efficiency, and streamline workflows for translators and localisation specialists

Over the years, several CAT tools were developed and now available. Some are free and some are paid. The free applications include OmagaT, Autshumato, MateCat, Wordfast (online), CafeTran, Anywhere, Espresso, and SmartCat. The paid ones include Wordfast (PC installation), XTM Cloud, SDL Trados Studio, Crowdin, MemoQ, Memsource, Wordfast, Wordbee, and déjà vu, amongst others. Flórez & Alcina (2011) declare that there is another difference between free and payment software that in many cases is not considered. Unlike the proprietary model, which views the user as a customer purchasing a completed product, users in the free software community are an essential part of the development process. One of the easiest ways to contribute to a free software project is by using the program and notifying the developers about any bugs encountered or features that are lacking (Flórez and Alcina, 2011).

In the South African context, Autshumato Machine Translation Web Service has been developed and introduced in the translation field by the Centre for Text Technology at the North-West University. Autshumato is an initiative supported by the Department of Sports,

Arts and Culture to create, launch, and maintain open-source translation technologies with the purpose of facilitating translation and ultimately enhancing information accessibility for all South Africans.

The initiative began in 2007, with the goal of aiding the development of documents in all official South African languages within the public sector. This was done by creating user-friendly open-source technologies that facilitate the translation process, encourage terminology standardisation, and reduce translation time. Additionally, the project encompasses machine translation (MT) systems. MT involves the automated translation of a given text from one language (like English) to another by a computer. The current available MT system capable of translating from English into Afrikaans, isiZulu, Sepedi, Xitsonga, Sesotho, and Setswana. This is a limitation of some sort; South Africa is a diverse multilingual country with twelve (12) official languages, including South African sign language. However, one of the primary objectives of creating Machine Translation Web Service (MTWS) was to promote multilingualism and access to information to all South Africans (Nemutamvuni, 2018).

The Autshumato Integrated Translation Environment (ITE) consists of various plugins that tailor the well-known OmegaT software. It offers a unified space featuring translation memory, machine translation (MT), and a glossary to aid in the translation procedure. While Autshumato ITE is designed for the eleven (11) official languages of South Africa, OmegaT operates independently of language, allowing translation between any two languages. However, the Department of Sport, Arts and Culture provided training for both Autshumato and OmegaT which caters for any language.

Literature Review

A literature review surveys books, scholarly articles, and any other sources relevant to a particular issue, area of research, or theory, and by so doing, provides a description, summary, and critical evaluation of these works in relation to the research problem being investigated (Fink, 2004). Many scholars have written and shared their views regarding CAT tools, including their advantages, importance, and benefits. Many scholars and researchers posit that machine translation systems or CAT tools (computer-assisted translation, computer-aided translation) are software that help to translate faster (Garcia, 2014).

Digital learning tools such as machine translation (MT) and creating online dictionaries can also contribute to preserving these languages. Each of these strategies offers benefits on how technology could be employed effectively and facilitate the preservation of indigenous languages. The use of technology is used to adequately preserve minority indigenous languages of South Africa; namely, Xitsonga, Siswati, Tshivenda, and isiNdebele. These languages do not receive much attention in terms of preservation using technology in South Africa (Mlambo and Matfunjwa, 2024).

The view mentioned above is supported by Sundani (2023) when he mentions that there is a connection between indigenous languages of South Africa and digital technologies, emphasizing the difficulties and possibilities in access, promotion, and preservation. Digital technologies can greatly support the promotion and preservation of South African indigenous languages. This can be credited to their multimedia functions, storage abilities, and communication tools. Nevertheless, restricted access to digital tools that aid these languages adversely affects their promotion and preservation. Barriers like insufficient

expertise, collaboration, fair digital services, and effective advocacy obstruct access to digital technologies for indigenous languages in South Africa (Sundani, Op cit.).

The lack of ICTs in indigenous languages reduces the opportunities for producing and accessing content in local languages on the internet. This influences the culture and knowledge that might be accessible to researchers keen on examining them. ICTs ought to be localised by translating and culturally adapting applications and software graphical user interfaces into native languages (Osborn, 2006).

Terminology management is an essential element in CAT environments as it encompasses the development, maintenance, and access of terms specific to a domain to guarantee precision and consistency. Integrated terminology databases within CAT software offer translators established vocabularies that minimise confusion and promote consistent word application across translations, especially in specialised areas like financial, legal, medical, and technical translation. CAT tools have translation memories that record previously translated phrases and terminology for future use, reducing repetitive work. OmegaT also provides customisable glossaries and potential translation suggestions, enhancing its utility for translators.

CAT tools not only act as improvements to productivity but also facilitate quality control, particularly in fields that demand high localisation standards such as software localisation projects that illustrate the benefits of CAT environments in handling intricate projects with various stakeholders (Yixin, 2024).

The primary purpose of CAT tools is to retain translated text segments for use in future translations. This technology is referred to as Translation Memory (TM), which serves as a database that keeps track of past work for potential reuse, along with enabling rapid searches through the existing content. The capability to reuse translations allows CAT systems to conduct a search in memory and provide reused matches, not just at the sentence level but also at the phrase and individual word levels, which is beneficial in documents that have repeated phrases (Barrachina et al., 2009).

CAT tools can acquire knowledge by assembling a vast collection of source texts alongside their corresponding translations, which need to be processed with a highly capable multiprocessor. When translating new texts, fragments from this collection are identified and selected to create the target text (Olohan, 2011).

Computer-aided translation refers to a method of translating from one language to another where a human translator utilises computer technology to enhance the process and carry out the translation (Bowker and Fisher, 2010). This is supported by Han (2020) when he mentions that in contrast to machine translation systems, utilising a CAT program necessitates the involvement of a translator in the translation process. It does not translate for you, but it allows you as a translator to fully carry out your responsibility.

Methodology

The researcher adopted a qualitative research approach in this article, that comprised desktop research of secondary materials. Desk research or secondary research is a pivotal method for analysing existing data and literature to generate actionable insights (Gupta, 2024). The researcher has chosen this methodology because data is readily available to understand

the phenomenon in order to achieve the aim of the research. The Promotion of Access to Information Act (PAIA) documents translated by different organisations were compared whereby terms were selected and compared for consistency against the standardised terminology. It is challenging for organisations and government entities to have identical content due to their differing mandates. The translations will always reflect the nature of the content or mandates they uphold, which vary from one organisation to another. The researcher has selected the Promotion of Access to Information Act (PAIA) document because very few organisations have translated it. This motivated the researcher to choose it for comparative analysis. When several translations are evaluated, it becomes easier to identify the accuracy and consistency of the terminology used in the different translations against standardised terminology. Comparing translation versions also helps detect spelling errors and evaluate the quality of the translations by examining the consistency of terms and translations. Terms that were appearing in all, if not most of the available OmegaT translation memories or glossaries, were selected to make comparisons easier.

Utilising the available data from the current translation memories and PAIA translated documents is both economical and efficient, as it eliminates the need for costly fieldwork or data gathering. This study indicates that conducting desktop research allows research conclusions to be shaped by the pre-existing terminology glossaries in the OmegaT translation tool and already translated PAIA documents, particularly for under-resourced languages like Xitsonga where empirical data is scarce. Comparative analysis was employed for data analysis. A comparative analysis is a side-by-side comparison that systematically compares two or more things to pinpoint their similarities and differences (Kaluza, 2023). The researcher opted for comparative analysis as it provides a systematic approach to assess and juxtapose various terminologies used by different translators in their OmegaT glossaries with the chosen translated terms in the PAIA documents, all in relation to the standardised terminology to derive insightful conclusions. Data or evidence was gathered from already translated texts (from English to Xitsonga) and OmegaT saved translation memories from Xitsonga for translators to verify the uniformity and/or variations of translation equivalents of the same term. Validity and reliability have been ensured during data collection through triangulation. Triangulation refers to using different data sources, investigators, and methods of data collection (Lincoln and Guba, 1985). The researcher uses two methods of collecting data, that is by comparing translated PAIA documents by different translators, and OmegaT glossaries or translation memories against the authenticated terminology respectively. The aim of this study is to explore the impact of centralised translation memory in OmegaT translation software in the case of Xitsonga language. To achieve the aim, the following research questions are explored:

- In what ways does centralised translation memory aid in the preservation and dissemination of terminology?
- How does centralised translation memory influence the overall quality of translations?
- In what manner does centralised translation memory enhance collaboration among translators?

Theoretical framework

Swanson (2013: 122) avers that the theoretical framework is the structure that can hold or support a theory of a research study. The Skopos theory and functional translation theories are among the most recognised approaches impacting translation studies. However, these theories did not consider the advancements in technology. Balkul (2016) pointed out that

technological elements were not adequately incorporated into established translation theories, which have yet to recognise the cognitive impact of technology.

To achieve the purpose of the research, the Corpus-based Translation approach is adopted. Corpus-based translation approach, often referred to as 'reference translation', comprises text and its corresponding translations in the target language (TL). The source language (SL) and the TL are connected, and their respective translations are derived through an extraction based on statistical models. The corpus encompasses electronically stored texts in a single language or multiple languages. It can be categorised into two distinct approaches: the statistical method and the example-based method. These can be better understood in this manner:

- The statistical approach focuses on bilingual text corpora and statistical models. A sentence from the source language can be translated in several ways.
- The example-based approach relies on a bilingual corpus to operate. Examples are fetched and selected based on their proximity in pairs of sentences. If there are no sufficiently close matches in the bilingual corpus, the example-based method may not successfully find the appropriate pair, leading to a low-quality result.

The corpus-based translation approach was incorporated into the comparative evaluation of translation memory (TM) within the OmegaT CAT tool to establish an empirical foundation for assessing the quality, consistency, and efficacy of preserved terminology. Translation memories aim to leverage previously translated segments, yet their effectiveness and dependability differ among various CAT tools. By utilising both parallel and monolingual corpora, the research was able to compare TM outputs with authenticated and standardised terminology, ensuring that the retrieved segments were not only precise but also contextually suitable.

Discussion/Analysis of Findings and Recommendations

The analysis of findings will be divided into part A and B.

Part A (PAIA translated documents)

Data was obtained through evaluating the already translated texts from English to Xitsonga to check variations and uniformity of translation equivalents of the same terms. The four translations of Promotion of Access to Information Act (PAIA) document from four translators were evaluated and compared. They were analysed using a side-by-side comparison that systematically compares two or more things to pinpoint their similarities and differences, hence the comparative analysis. The two translations were from two government departments and two from state owned entities. Pseudonyms were used for privacy purposes, these documents were named document A, B, C, and D. The Promotion of Access to Information Act document was chosen because even though every department or state-owned entity has customised their own documents according to their mandates and duties, some information will remain formal or similar. The researcher wanted to check similarities and differences as far as terminology is concerned. These translated documents were readily available online for download. Only one document (PAIA) was chosen because almost every department or state-owned entity has translated it, the researcher did not want to evaluate or compare different documents. In these documents, some words and terms, or phrases were selected for the purpose of this study. Different published Xitsonga

terminology lists were also checked to verify if the selected words or terms are available on those lists. Those are the terms that have been authenticated by the Xitsonga National Language Body of the Pan South African Language Board, and they are considered to be standardised terms. Below are the selected words or terms and the authenticated ones:

Table 1: indicates the extracted terms from the PAIA documents translated by different translators to check consistency and spelling of the terminology used against the standardized or authenticated terminology (The researcher's own collection)

Term	Document A	Document B	Document C	Document D	Authenticated / Standardised Term
data	datara	switiviwa	switiviwa	data	switiviwa/deyitha
prescribed (fee)	phirisikirabiweke	vekiweke	vekiweke	vekiweke	–
contract (law)	kontiraka	kontiraka	kondiraka	kontiraka	kontiraka
strategy	xitirateji	kungu	–	qhinga	qhinga
processed	purosesa	phurosesa	phurosesa	phurosesa	–
printed	pirintiweke	kandziyisiweke	kandziyisiweke	kandziyisiweke	–
deposit	diphoziti	–	diphoziti	dipoziti	dipoziti/diphoziti
electronic	xielekitironiki	xilekitironiki	xielekitironiki	xitironiki/ xielekitironiki	xiilekitironiki/ xitironiki/ xielekitironiki (not a stand-alone term)
third party	thedi phathi	vandla ra vunharhu	vandla ra vunharhu	munhu wa vunharhu	ntlawa wa vunharhu
Information officer	muofisiri wa mahungu	muofisiri wa vuxokoxoko	muofisiri wa vuxokoxoko	muofisiri wa mahungu	not a stand-alone term, information - mahungu/officer- muofisiri

Data

The term data denotes information or details concerning a specific topic. Document A provided a translation equivalent of datara; however, despite adhering to the appropriate consonant-vowel pattern in localising the term, it was still deemed inaccurate according to the authenticated term. Above all, the translation equivalent datara was made to conform incorrectly to the spelling and orthography rule from Xitsonga Spelling Rules and Orthography (2019: 61), stating that:

> *Marito lama helaka hi ntwariso wa /-**ter**/, /-**ner**/, /-**dar**/, /-ure/ na man 'wana lamo tano eka tindzimi to fana na Xinghezi ya fanele ku hela hi /-**ra**/ loko ya hundzuluxeriwa eka Xitsonga.*
>
> (Words that end with the /-**ter**/, /-**ner**/, /-**dar**/, /-**ure**/ sounds and others similar to those ones in languages such as English must end with the sound, /-**ra**/ when translated into Xitsonga). The term data ends with a suffix or sound of /-**ta**/ not, /-**ter/,** /-**ner**/, /-**ure**/, so it is incorrect to translate it into *datara.*)

However, there are some exceptions involved when dealing with the above-mentioned rule, some words/terms are not able to conform and take the sound /-**ra**/. This is substantiated by Xitsonga Spelling Rules and Orthography (2019: 61) when it avers that:

> *XIYA: Swi nga endleka man'wana marito ya nga pfumeli ku tirhisa xilandzi xa /-**ra** / xa nawu* lowu, xk. collar > *kholoro*; spanner > *xipanere*; archar > *acha*.
>
> (NOTE: It is possible that some words may not conform to the rule of adopting the suffix /-ra/, e.g., collar > *kholoro*; spanner > *xipanere*; archar> *acha.*)

Translators require knowledge and understanding of the linguistic mechanisms of word-formation processes (Baloyi, 2023). Xitsonga language translators and Xitsonga speakers should familiarise themselves with the new spelling and orthography. The *banginkulu** saga would have been avoided if people knew the spelling rules in their language. Inconsistencies like these in the language hinder the ongoing terminology development efforts and drag down the quality of the translated documents. However, implementing a centralised translation memory might come with challenges with issues like:

- Management and version control: inquiries might emerge regarding who oversees the centralised TM, who modifies entries, and how to handle version control.
- Standardisation of terms: Consistent formats, spelling and orthographies, and metadata are essential for translation memories. Xitsonga, just like any African language, frequently experience variations in spelling and a lack of standardised terminology, complicating harmonisation efforts.
- Inter-institutional coordination: Language units, universities, government departments, and entities may use different CAT tools or workflows. PanSALB through Department of Sport, Arts and Culture (DSAC) must regulate this.
- Technical Infrastructure: Centralised translation systems need dependable servers, safe storage, and consistent internet. Numerous institutions continue to encounter bandwidth restrictions, obsolete systems, or a deficiency of cloud-based solutions customised for African environments.
- Privacy, ownership, and access rights: Who possesses the translations of the terms? Should every independent translator, government department or entity, educational institution, given open or complimentary access? Sharing is sometimes complicated by sensitive or copyrighted content. However, DSAC, as the recognised custodian, can manage this.

Contract

Documents A, B, D and the verified list (the standardised term) translated the contract as *kontiraka*, while Document C rendered it as *kondiraka*. The equivalent provided by Document C is inaccurate when considering the stipulation of the rule in the Xitsonga Spelling Rules and Orthography (2019: 61) which proposes that:

> *Malombiwa lama cinciweke mimpfumawulo ya twarisiwa hi sisiteme ya Xitsonga ya tsariwa hi laha ya twarisiwaka hi Xitsonga.*
>
> (Words taken from other languages that have been altered for pronunciation according to the Xitsonga system must be spelled as they are pronounced in Xitsonga.)

To prevent the creation of a new term, it was intended for the term to be pronounced as it is in English. The writing system needs to be considered and different terminology lists should be referenced to ensure the quality of our translations is maintained. Nonetheless, reviewing every single terminology list can be a lengthy process.

Strategy

The verified term for strategy appears in one of the terminology lists as *qhinga*, and Document D aligns with the verified term. Document A, on the other hand, has rendered it into *xitirateji*, while Document B has translated it into *kungu*. Why must we transliterate a word when we already have an equivalent in our language? *Xitirateji* adheres to the spelling, and the writing system is accurate (Consonant + Vowel), yet it is a wrong equivalent and negatively impacts the quality of our translations, inviting criticism.

Prescribed

In Document A, the term prescribed is translated as *phirisikirabiweke*, while Documents B, C, and D use the equivalent translation *vekiweke.* No standardised or verified term exists in the terminology lists; however, Document A chose transliteration, whereas Documents B, C, and D depended on the term's context. Although no standardised term exists, the translator should have explored each option and used the accurate equivalent by consulting the terminology list provided. This approach wastes time due to uncertainty about which list to reference, relying solely on the availability of equivalents in the translation memory and glossary of the CAT tool, OmegaT.

Processed

Documents B, C, and D possess the same translation equivalent *phurosesa,* whereas Document A contains *purosesa*. The verified term is not easily accessible, yet this is a straightforward word to translate. The term *purosesa* does not adhere to the Xitsonga spelling and orthography rules. The identical guideline that applies to the term "contract" should also be applied to this term. Borrowed words altered for pronunciation according to the Xitsonga system should be spelled as they sound in Xitsonga. In summary, translators need to have a strong understanding of word formation processes and translation strategies.

Printed

The strategy applied to this word is transliteration. Document A chose to transliterate the word printed as *pirintiweke,* whereas the other documents used the translation equivalent *kandziyisiweke*. The official terminology lists do not contain the standardised term. The translation used here is inaccurate, given that the equivalent is relatively easy to find. This could be due to the fact that sometimes translators do not put more effort in what they, as a result, poor quality translations are created.

Deposit

Documents A and C provided the same equivalent of *diphoziti*, while Document D offered an equivalent of *dipoziti*. Two verified terminology lists include *dipoziti* and *diphoziti*. This creates a disparity in translations because the two are validated in separate terminology lists and are regarded as distinct terms rather than synonyms. PanSALB should allow board members to gather occasionally to align the verified terms with similar meanings to prevent problems like these. Possessing various translation options for the identical term in the same context is problematic.

Electronic

Documents A and C have rendered the word electronic as *xielekitironiki*, while Document B has rendered it as *xilekitironiki,* and Document D has used *xitironiki* and xielekitironiki. In certain validated terminology lists, the term appears as an independent term, but it is noted as *xiilekitironiki, xitironiki*, and *xitironiki.* Because the terminology lists cannot maintain a consistent word or synonymous terms, this may lead to discrepancies in translations and issues with quality.

Third party

The translation equivalent for the term third party in Document A is *thedi phathi*.Iin Documents B and C, it is *vandla ra vunharhu*; in Document D, it is *munhu wa vunharhu*, and for the authenticated term, it is *ntlawa wa vunharhu*. From the comparison of the provided translation equivalents, it is clear that each translator offered an equivalent they believe is appropriate. Nonetheless, the differences between the equivalents *vandla ra vunharhu* and *munhu wa vunharhu* might be a matter of context. Thus, centralised translation memory must include terms with all translation equivalents based on various registers to accommodate contextual scenarios. Such factors lead to inconsistencies in translations and diminish their quality.

Information Officer

The standardised terminology list does not include the term 'information officer' as an independent term; however, the verified term for information is *mahungu* and for officer it is *muofisiri*, indicating that the correct term should be *muofisiri wa mahungu.* Documents B and C translated information to *vuxokoxoko*, therefore they possess *muofisiri wa vuxokoxoko. Vuxokoxoko* and *mahungu* can sometimes be used interchangeably in certain situations based on the context; translators must be cautious when addressing contextual considerations in translations.

Part B (OmegaT translation memory/glossaries)

Data were gathered by assessing OmegaT glossaries or translation memories from various Xitsonga translations to analyse discrepancies and consistency of translation counterparts for identical terms. Regarding glossaries, the terms that will be included are at the translator's discretion. Various translators will possess distinct terms in their glossaries,

as what holds significance for one translator might not be significant for another. In this context, it is essential for translators to access each other's glossaries and translation memories. In this manner, translators will gain access to terminologies from various fields and prevent the duplication of terms that have already been established. The researcher analysed four OmegaT glossaries created by various translators from both the private and public sectors. Pseudonyms were utilised here for confidentiality reasons, referred to as Glossaries A, B, C, and D. Below are the chosen terms along with the verified terms for the chosen expressions:

Table 2: indicates the extracted terms from the OmegaT translation memories or glossaries by different translators to check consistency and spelling of the terminology used against the standardized or authenticated terminology

Term	Glossary A	Glossary B	Glossary C	Glossary D	Authenticated Term
state capture	lawuriwa ka mfumo hi mavandla ma le tlhelo	nhlohlotelo wa le handle wa mfumo	–	–	–
vaccine	nsawutiso	xisawutisi	–	–	nsawutiso
non-compliance	nkavulandzelelanawu	ku nga landzeleriwi ka milawu	–	–	nkavulandzelela-wu
e-hailing services	–	vukorhokeri bya mathekisi lama vitaniwaka eka inthanete	–	–	–
disclaimed audit opinions	–	mihlahluvo ya oditi leyi oditara yi nga nyikiki mavonelo	–	–	–
disclaimed and adverse opinions	–	mihlahluvo leyi oditara yi nga nyikiki vonelo na leyi tirhekodo na switatimente swi andlariweke hindela leyi hambanaka na milawu ya tinkota	–	–	–
load-shedding	ntirhiso wa gezi hi ku siyerisana	–	–	–	–
terror financing	mali yo seketela vutherorisi	–	–	–	–
catfishing	kanganyisa vuwena	–	–	nkangayiso	–

State capture

Glossary A offered a translated equivalent for state capture as *ku Lawuriwa ka Mfumo hi Mavandla ma le Tlhelo*, while Glossary B rendered it as *Nhlohlotelo wa le Handle wa Mfumo*. Glossaries C and D, along with the authenticated term, lack translation equivalents or terminology. The lack of the translated term in Glossaries C and D may indicate that the

translators have never come across the term previously or it was not significant enough for them to retain or include it in their glossaries. The lack of presence in the standardised terminology list indicates that it has not been included for development; however, it is essential to welcome new words into the corpus and offer definitions for these terms to support the evolution of language and terminology. State capture is a relatively new and intricate concept; the two translators choose to employ paraphrasing as a translation approach, since finding a single term is quite difficult. Due to the lack of access each translator has to other translators' glossaries, the term will still end up with multiple translation equivalents or terms for a particular term. This undermines the quality of translations and leads to inconsistencies in terminology.

Vaccine

Glossary A offered a translation equivalent for vaccine as *nsawutiso*, whereas glossary B provided a translation as *swisawutisi*. The other glossaries did not offer any equivalent. The authenticated list offered a parallel as *nsawutiso*. The two translation equivalents differ in their structure; one is in uncountable form (*nsawutiso*) while the other is in countable form, yet both terms essentially convey the similar meaning. Nevertheless, because the standardised translation equivalent exists, translators must adhere to it. It can be exhausting at times to look for a term in terminology booklets or to search for it in a terminology list on a computer or online. As translators utilise OmegaT for their work, the term would be easily obtainable in a centralised translation memory or glossary, resulting in saved translation time and shorter turnaround times for translation projects.

Non-compliance

Glossary A provides the translation equivalent for non-compliance as *nkavulandzelelanawu*, while Glossary B offers an equivalent translation of *ku nga landzeleriwi ka milawu*. The other glossaries failed to offer translation equivalents. Nevertheless, the authorised terminology corresponds to *nkavulandzelelanawu*, which is the recognised term. The translation equivalent of Glossary B is identical to Glossary A and the verified term; nevertheless, Glossary B chose to paraphrase the term, providing an explanation of it which is acceptable. The joint work of the consultative meetings, verification, and authentication process by the PanSALB facilitated translators in creating a strong term instead of needing to devise an explanation.

E-hailing services

Glossary B rendered e-hailing services as *vukorhokeri bya mathekisi lama vitaniwaka eka inthanete*, while the other glossaries offered no translation equivalent. There is no standardised translation equivalent for the term; therefore, the official term is missing. This term is quite recent in the vocabulary, it was introduced when this mode of transport began. Its lack in other glossaries may be due to being new and having less content requiring translation; only those in the transport sector might have encountered it. The equivalent provided by Glossary B restricts the transportation to a specific kind of transport, namely taxis, whereas e-hailing services can encompass various types of transport, including vans, sedans, or even trucks, though the intended message is clear. Collaboration might

be beneficial in this case, translators can share their ideas and subsequently establish a standardised term, conserving time and preventing the need to create new terminology.

Disclaimed audit opinions

The translation equivalent provided for disclaimed audit opinions by Glossary B is *mihlahluvo ya oditi leyi oditara yi nga nyikiki mavonelo*. The other glossaries containing various authenticated terms or terminology lack the equivalents. The translator for Glossary B attempted to paraphrase but the equivalent appears too lengthy, and there was little the translator could accomplish on their own. Accounting terminology consistently presents challenges for translators. If the translators could be observed by fellow translators, it is thought that others could have enhanced it by utilising available translation strategies and word formation strategies. This approach would allow translators to adhere to the superior translation and simplify its validation, as no new term would surface. This would remove superfluous duplication and redundancy.

Disclaimed and adverse opinions

Glossary B is the only one that succeeded in finding a translation equivalent, meaning, *mihlahluvo leyi oditara yi nga nyikiki vonelo na leyi tirhekodo na switatimente swi andlariweke hindela leyi hambanaka na milawu ya tinkota.* A reader cannot fail to comprehend this in a translated text. Due to the absence of a suitable equivalent, the translator opted for paraphrasing, and there is always potential for enhancement since the term is excessively lengthy and transformed into a sentence. If the OmegaT translation memory of glossary was automatically accessible to other translators without the need for manual sharing, they could examine the contents of Glossary B and attempt to enhance the equivalent; translation necessitates creativity, collaboration, and teamwork. Translators do not operate in isolation for the same reason. Translators work together to create new terminology and enhance the existing content in the corpus. A term can have one meaning today and another tomorrow, as times are changing. Fresh ideas, innovation, and experience are consistently required in translations.

Load-shedding

Glossary A proposed a translation equivalent for load-shedding as *ntirhiso wa gezi hi ku siyerisana*. The term is missing in the other glossaries, including the authenticated terminology. Translators determine what to add and what to omit in their glossaries. If other translators could view what Glossary A generated, they could easily create a new term to replace this one or add a new synonym. Primarily, translators may retain the established term if they find it acceptable. In this manner, they conserve time instead of squandering it and overthinking without reason and maintaining consistency in translations, thus enhancing quality.

Terror financing

Glossary A translated terror financing as *mali yo seketela vutherorisi*. The other glossaries did not include the term, resulting in the lack of translation equivalents. Its appearance

in Glossary A may indicate that the translator primarily works with financial terminology or that the term holds significance based on the judgment. Developing terminology for financial terms is quite challenging. In an effort to close this gap, the former Financial Service Board Multilingual (FSB), now known as The Financial Sector Conduct Authority (FSCA), developed its Financial Terminology List released in 2017. The financial sector is extensive and so is its vocabulary; thus, additional terminology development projects are required to address its wide range of terms.

Catfishing

Glossary A is the sole source that provided a translation equivalent for *kanganyisa vuwena*, and a paraphrasing translating approach was employed in rendering the term. Glossary D rendered it as *nkangayiso*. The two counterparts differ in all respects; the aspect of inconsistency is highlighted here. If an individual encounters both translations in a document, they might be puzzled because the term was rendered differently, possessing two distinct meanings. *Nkanganyiso* in an acknowledged definition referring to fraud. The phrase is recent as the age of social media has developed and accelerated over time. The other glossaries omit the term, suggesting that translators have never encountered it or that it has not affected them in any manner. The introduction of new terms should be acknowledged and incorporated during the development and creation of terminologies. Acknowledgment must also be made to the Limpopo Department of Sport, Arts and Culture, which has initiated its consultation procedures concerning the creation of social media guidelines in Xitsonga, Sepedi, and Tshivenda.

Recommendations

For language to grow, as new words frequently enter language space, terminologies must be developed, maintained and preserved. The centralisation of the OmegaT glossary or translation memory should be embedded to the General Theory of Terminology (GTT) by Eugen Wuster, the main proponent. His core idea regarding the theory is based on removing uncertainty from technical languages through standardising terminology to transform them into effective communication tools to persuade all users of technical languages regarding the advantages of consistent terminology, and to define terminology as a field for all practical aims and to elevate it to the level of a science (Wüster, 1968). The glossary shown in the translation memories within OmegaT software is at the translators' discretion and it is improbable that the terms one translator possesses differ from those of other translators. In light of this context, it is suggested that OmegaT should include a sharing feature or a repository for storing translated terms. Post-translation, the translator will enable the sharing feature to store the document in a central location. Activating the sharing feature will make the document containing the terms accessible in the OmegaT glossary for any translator who opens a document with the same terms. In this manner, translation memories and glossaries are centralised for all translators to access, preventing inconsistencies in translation terms and equivalents. In dictionary auto-search, CAT tools emphasise terms within the text recognised by the terminology management system (Krüger, 2016). Should the document be confidential, the translator will refrain from enabling the sharing feature ensuring that only they have access to the document for privacy reasons. This approach will preserve translation quality, save time, and foster collaboration. They also need to ensure that the available authenticated terminology lists are loaded into a centralised glossary for

easy access to prevent the duplication of terminologies already created for various fields or subjects. The recommendation stated above is the responsibility of the Department of Sport, Arts and Culture, which is tasked with developing, implementing, and promoting the Computer Assisted Translation tool. They need to ensure that all translators adopt OmegaT for translation by organizing additional awareness workshops and mandating its use for every translator. Requests for OmegaT Enhancements (in English) can be submitted at the SourceForge site: https://sourceforge.net/p/omegat/feature-requests/. Translators must familiarise themselves with the linguistic mechanisms of word formation processes, spelling rules, and orthography, as language is dynamic rather than static. The recognition and inclusion of new terms is essential in the development and creation of terminologies, advancing the objectives of terminology management.

Conclusion

A CAT tool is a crucial element of ICT that translators need to use to manage and preserve languages through terminology development and management. The article examined the influence of centralised translation memory and glossary in OmegaT translation software, focusing on the Xitsonga language. Centralising the translation memories and glossaries will reduce inconsistencies, redundancies, streamline translation time, and result in high-quality translations. Not only will it improve the quality of the translations, but it will also promote collaboration among translators and facilitate the sharing of information regarding terminology learning. However, certain translators possess the software on their devices but are not utilising it, they should reassess and employ this tool to ensure preservation and maintain quality in their everyday translations. The researcher wanted to evaluate more OmegaT glossaries and documentation but was limited by the fact that some translators were not comfortable sharing their glossaries with the researcher. Some translators are not utilising any CAT tool in their translations, everything is done manually. To find the same document translated by a large number of translators is hard since the content that is translated is guided by the organisation's mandate or subject matter. This subject matter is under researched especially in indigenous languages or low resourced languages, a lot can still be done for future research. Future studies may focus on exploring how CAT tools can effectively share translation memories, standardise and validate translation equivalents across various organisations, and how to manage quality control in stored glossaries while removing outdated terminology from the CAT tool.

References

Autshumato. (n.d.). *The Autshumato MT Web Service*. Retrieved June 20,2025, from https://mt.nwu.ac.za/#.

Balkul, H. I. (2016). 'Translation technologies: A dilemma between translation industry and Academia', *Online Submission*, 4(4), 100-108.

Barrachina, S., Bender, O., Casacuberta, F., Civera Saiz, J., Cubel, E., Khadivi, S., Lagarda A. L., Ney, H., Tomas, J., Vidal, E., & Vilar, J. (2009). 'Statistical approaches to computer-assisted translation', *Computational Linguistics*, 35(11), 3-28. https://doi.org/10.1162/coli.2008.07-055-R2-06-29

Fink, A. (2014). *Conducting research literature reviews: From the internet to paper* (4th ed.). Thousand Oaks, CA: SAGE.

Flórez, S., & Alcina, A. (2011). 'Catálogo de software libre para la traducción', *Revista Tradumàtica*, 9, 57–73. https://doi.org/10.5565/rev/tradumatica.5

Folaron, D. A. (2010). Translation tools. In Y. Gambier & L. van Doorslaer (Eds.), *Handbook of translation studies* (Vol. 1, pp. 429–436). Amsterdam: John Benjamins Publishing Company. https://doi.org/10.1075/hts.1.tra9

Garcia, I. (2014). Computer-aided translation: Systems. In I. Garcia (Ed.), *Routledge encyclopaedia of translation technology* (pp. 106–125). London, New York, NY: Routledge.

Grishman, R. (1986). *Computational linguistics: An introduction*. New York: Cambridge University Press. https://doi.org/10.1017/CBO9780511611797

Gupta, R. (2024). 'Methodological and theoretical rigor in desk research', *ResearchGate*, https://www.researchgate.net/publication/386642850_Methodological_and_Theoretical_Rigor_in_Desk_Research.

Kaluza, J. (2023). *Comparative Analysis*. Retrieved July 10, 2025, from https://dovetail.com/research/comparative-analysis/

Han, B. (2020). 'Translation, from pen-and-paper to computer-assisted tools (CAT tools) and machine translation (MT)', *Proceedings*, 63(1), 56. https://doi.org/10.3390/proceedings2020063056

Krüger, R. (2016). 'Contextualising computer-assisted translation tools and modelling their Usability', *Journal of Translation and Technical Communication Research*, 9, 114–148.

Lincoln, YS. & Guba, EG. (1985). *Naturalistic Inquiry*. Sage Publications: Newbury, Park, CA. https://doi.org/10.1016/0147-1767(85)90062-8

Mahdi, H. (2018). 'A review of literature of computer-assisted translation', *ResearchGate*, https://www.researchgate.net/publication/328019119_A_Review_of_Literature_of_Computer-Assisted_Translation.

Mlambo, R. & Matfunjwa, M. (2024). 'The use of technology to preserve indigenous languages of South Africa', *Literator*, 45(1), 1–8. DOI: https://doi.org/10.4102/lit.v45i1.2007

Nemutamvuni, M. E. (2018). Investigating the effectiveness of available tools for translating into Tshivenda. Unpublished M. A. dissertation. University of South Africa, Pretoria.

Olaare, S. (2024). 'The role of technology in language preservation', *European Journal of Linguistics*, 3, 44–56. https://doi.org/10.47941/ejl.2046

Olohan, M. (2011). 'Translators and translation technology: The dance of agency', *Translation Studies*, 4(3), 342–357. https://doi.org/10.1080/14781700.2011.589656

OmegaT. OmegaT (n.d). *Translation Tool Overview*. Retrieved September 15, 2025, from (https://autshumato.nwu.ac.za/software/.

Osborn, D.Z. (2010). *African languages in a digital age: Challenges and opportunities for Indigenous languages computing*. Human Sciences Research Council: Cape Town.

Sundani, N.D. (2023). 'South African Indigenous languages and digital technologies: Access, promotion and preservation', *International Journal of Social Science Research and Review*, 6(8), 41–51. https://doi.org/10.47814/ijssrr.v6i8.1385

Swanson, R. A. (2013). *Theory building in applied disciplines.* San Francisco, CA: Berrett-Koehler.

Wang, Y. (2024). 'The impact of technology on human translators and translation quality: A study on machine translation and computer-assisted translation tools', *English Linguistics Research*, 13(1),19. https://doi.org/10.5430/elr.v13n1p19

Wüster, E. (1968). *The Machine Tool. An Interlingual dictionary of basic concepts*. London: Technical Press.