

Comparing machine translation and human translation for South African languages

Respect Mlambo 

South African Centre for Digital Language Resources, North-West University, South Africa
E-mail: Respect.Mlambo@nwu.ac.za

Valencia Wagner 

Department of Languages and Communication, Sol Plaatje University, South Africa
E-mail: Valencia.Wagner@spu.ac.za

Nomsa Skosana 

South African Centre for Digital Language Resources, North-West University, South Africa
E-mail: Nomsa.Skosana@nwu.ac.za

Abstract

Machine translation (MT) has been available for several years but has only recently begun to be considered viable, particularly in the context of indigenous South African languages. Although the quality of an MT remains inferior to that of a human translation (HT), MT systems have gained popularity, making some significant contributions to translation studies. This study explored the similarities and differences between the two modes of translation. A case study approach based on a qualitative research method was used. The source data for this study are an MT and an HT of an article titled “Stellenbosch University to offer academic and psychosocial support to students”, written by Ntwaagae Seleka and published on News24. As a computer-aided translation tool, Autshumato Machine Translation Web Service (MTWS) produced the MT from English into Setswana as a target text. Meanwhile, a Setswana master’s student from the Tshwane University of Technology manually translated the source text. The findings show that the human translator outperformed the MTWS in providing high-quality translation. The MTWS was unable to offer a better translation in terms of case sensitivity and terminological inconsistencies. It also mistranslated, adding and omitting words that changed the intended meaning and leaving certain words untranslated. Human cognitive competency, intelligence, and flexibility enable human translators to deal with such translation problems to provide high-quality outputs. The initiatives discussed in this study show that even though the MTWS operates as a useful translation tool with the capacity to instantly translate a large number of documents, its output is not yet capable of replacing an HT when translating into indigenous South African languages. Post-editing of the MTWS outputs is always recommended.

Keywords: Machine translation; human translation; South African languages; Autshumato Machine Translation Web Service; translation; human translator; computer-aided translation

1. Introduction

South Africa is a multilingual country with twelve official languages. These official languages recognized by the constitution are English, Afrikaans, Xitsonga, isiNdebele, Setswana, isiXhosa, isiZulu, Sepedi, Sesotho, Siswati, Tshivenda, and South African Sign Language (Mlambo & Matfunjwa 2024:1). Apart from English and Afrikaans, these languages are referred to as resource-scarce languages because they currently do not benefit from modern digital technologies. Finlayson and Madiba (2002:40) confirmed that the nine indigenous languages have been partially developed, since they have written forms such as literary works, dictionaries and terminology lists, but are underdeveloped in modern technology compared to English and Afrikaans. The arena of modern digital technologies is being developed to provide the required knowledge and abilities to improve the efficiency with which machines mediate human-to-human communication and enable human-to-machine communication (Adegbola 2009:53).

Indigenous languages continue to lag behind in the advancement of natural language processing (NLP) applications due to a lack of representative data (Makgatho, Marivate, Sefara & Wagner 2022:1). The lack of data imposes challenges such as effective communication, standardization, consistency of the terminology, and conducting research using technologies in these languages. Having a strategy that can transfer information between languages could help overcome the lack of machine training data (Makgatho et al. 2022:1).

There have been major initiatives from several research infrastructures and government agencies across Africa and elsewhere to develop NLP applications that cater for indigenous languages. According to Adegbola (2009:53), the application of NLP to indigenous languages is relatively new, with the most consistent efforts inspired and led by state policy coming from nations such as South Africa; programmes in other countries are based mostly on private initiatives. To overcome the scarcity of digital technologies in the field of NLP, the South African government has developed many legal frameworks that support the usage and promotion of indigenous languages and has funded several initiatives during the previous two decades (Puttkammer & Du Toit 2022:1). Despite the South African government's attempts to build NLP applications, indigenous languages in South Africa continue to be resource-scarce. De Pauw, De Schryver, Pretorius and Levin (2011:266) list the following organizations as being the most notable regarding NLP applications for indigenous languages:

- Google Africa released a Swahili version of their translation system in 2009.
- Translate.org.za is a non-profit organization that focuses on translation and localization work of South Africa's official spoken languages.
- TshwaneDJe HLT, with offices in Africa and Europe, is mainly involved in computational lexicography (Joffe & de Schryver 2004) but has also developed African language spellcheckers, text messaging resources, digital corpora, and localization components.

A recent addition to this list is the Centre for Text Technology (CTeXT) at North-West University, Potchefstroom, which has developed an easily accessible machine translation (MT) tool known as Autshumato Machine Translation Web Service (MTWS). This tool was developed under the Autshumato Project, which was funded by the Department of Arts and Culture, to develop adequate NLP facilities to accommodate the official spoken languages of

South Africa. This facility provides sentence, document, and web page translation from one source language, English, into six target languages: Afrikaans, Xitsonga, Setswana, isiZulu, Sesotho and Sepedi; it is freely accessible on the website of the South African Centre for Digital Language Resources (SADiLaR) (<https://sadilar.org/>). According to Skosana and Mlambo (2021:1), the MTWS is a computer-aided translation (CAT) tool that works in combination with human involvement to translate both formal and informal writings. As a CAT tool, the MTWS includes a translation memory, which aids in the storage of translated words, terms, and segments with their associated source words, terms, and segments for reuse in future translations (Nemutamvuni 2018:51). The purpose of MTWS is to create, distribute and promote open-source translation technology that aids in the translation process and eventually provides further access to knowledge for all South Africans.

According to Abbott and Martinus (2019:98), the lack of research focus on and evaluation of existing technologies designed for resource-scarce languages makes it impossible for other researchers to establish a benchmark. Therefore, the purpose of this study was to compare MT and human translation (HT) for South African languages, with a focus on Setswana. The study contrasted the two types of translation to determine the effectiveness of the MT system and whether it is capable of replacing HT in the context of South African languages. As a comparative study, the MTWS was selected to produce MT output. The MTWS was used as an MT system in this study because it is a prevalent translation software used by translators and is freely accessible. A Setswana language practice master's student from the Tshwane University of Technology was recruited as the human translator. To achieve its aims, the study has been divided into five sections. In the next section (section 2), we discuss work related to this study. In section 3, we describe the research methodology, while the results and discussion of the study are detailed in section 4. Finally, in section 5, a conclusion is drawn.

2. Related work

Previous research that discusses the differences between the output of MT systems and HT has been conducted predominantly in European and Asian languages. Studies in these languages show that MT systems translate documents with a high degree of accuracy and at high speed; in some settings, however, human post-editing and proofreading are essential to maintain translation quality. Ahrenberg (2017), for example, compared MT with HT, focusing on the Swedish language. The British opinion article titled "Why I left my liberal London tribe", written by David Goodhart and published in March 2017, is used as source text (ST). The MT system output was obtained from Google Translate. The basic statistics results show that the number of words, sentences and characters of the HT has more tokens compared to that from the MT system. Concerning the output quality, Ahrenberg (2017) pointed out that the HT is of higher quality than the MT. The Ahrenberg study proposed profiling the number and types of post-editing processes judged to be necessary to increase the quality of the output.

In another study, Zong (2018) investigated the relations between an MT system and human translators. Both the MT system and human translators have their application scenarios, and the services used provide high-quality translations. For instance, the MT system can have a higher number of translations every time due to its ability to easily understand new language customizations, which takes more time for human translators to do. Zong also highlighted that the interpersonal understanding of semantics and the syntactic integration of target languages by human translators make it easier to produce quality translations, but it is not easy for the MT

system to do the same. The study concluded that if the MT system and human translators can complement each other, the efficiency and quality of translated texts will be markedly improved. In other words, the MT system as an artificial intelligence works together with human translators to produce translation works with high efficiency and better quality.

The current study also drew on the publication of Wu, Liang and He (2016) who make a comparative analysis of an MT and an HT. The study aims to compare the MT and HT to determine the limitations of the MT system at the lexical and syntactic levels and to suggest possible solutions. The Kingsoft Powerword tool is used to produce the MT. Meanwhile, the HT is completed by an editorial board member of the book, “A Survey of the Olympic Games from English Translation Practice II”, which is used as the ST in the study. The level of lexical accuracy is analysed primarily from a lexical ambiguity and named-entity perspective; word order, word segmentation, noun phrase, and verb phrase are analysed on a syntactic basis. The study reveals that human translators outperform MT in various aspects of translation, including context recognition, flexible word order arrangement, and linguistic errors in the target text (TT). Wu et al. (2016) highlighted that the machine translator, unlike a human translator, lacks human cognitive competence. To overcome these limitations, Wu et al. (2016) supported the view that machine translation should be complemented with the support of a human translator in the pre-editing stage of the ST and in post-editing the TT in order to achieve a high-quality translation.

Arvianti (2018) analysed differences between the output of an MT system and HT using formal and non-formal language texts. The data used in Arvianti’s study are captions that cover news items and entertainment on Instagram. The captions selected are translated by four university semester students studying Translation modules, and a translation tool is provided by the Instagram platform. The findings showed that the human translators are better at translating special terms and demonstrate more diverse vocabulary than the MT system. The human translators can also translate difficult words and produce readable texts that follow the rules of grammar. However, when the source language uses common words and formal language, the MT system produces a better translation. These findings show that the MT system cannot translate non-formal language as it lacks vocabulary in its storage. The human translators produced translation with a high level of accuracy, acceptability, and readability in both formal and non-formal language, while the MT system produced translation with a low level of accuracy, acceptability and readability in the non-formal language.

Despite all the research on the importance of MT systems in translation studies, there is a lack of work investigating the similarities and differences between the outputs of MT systems and HT in South African languages. Our study attempts to fill this research gap by investigating the similarities and differences between the translation outputs provided by the MTWS as our MT system and a Setswana master’s student from the Tshwane University of Technology offering HT.

3. Methodology: A case study

The research design used in this study is that of a case study based on a qualitative research approach. Yin (2003) indicates that a research design aimed at a contemporary phenomenon, focusing on the dynamics of the case within its real-life context and generating hypotheses, is known as a case study. The data for this study comprised of two forms of translation; one was

rendered by the MT system, and the other by a human translator. The English ST is an article titled “Stellenbosch University to offer academic and psychosocial support to students”, written by Ntwaagae Seleka and published by News24 on 17 August 2020. Table 1 displays the ST and the two types of translation: MT and HT.

Table 1. Source text with translations by the MT system and a human translator.

English source text	Setswana MT	Setswana HT
<p>Stellenbosch University to offer academic and psychosocial support to students News24 Author: Ntwaagae Seleka Published: 17/08/2020</p> <p>1. Stellenbosch University has promised to strengthen academic and psychosocial support for its students. The university will utilise collaborative efforts to help undergraduate students during virtual learning.</p> <p>2. The institution together with the Michael & Susan Dell Foundation will support students who are vulnerable and at risk of not graduating or moving on to their next year of study.</p> <p>3. Rector and vice-chancellor Professor Wim de Villiers said the Covid-19 pandemic had brought with it many new challenges and risks the university had been working hard to mitigate.</p> <p>4. “We remain acutely aware of the challenges that our students face in adapting to this new and largely unfamiliar mode of learning during the lockdown period. So, we are doing everything we can to ensure that all of our</p>	<p>Yunibesiti ya Stellenbosch go neelana ka tshegetso ya tsa Akatemi le jwa baithuti Dikgang 24 Mokwadi: Ntwaagae Seleka Gatisitsweng.....: <17/08/2020></p> <p>1. Yunibesiti ya Stellenbosch e solofeditse go nonotsha boemo jwa tlhaloganyo le mo go tsa Akatemi le Kemonokeng ya baithuti. Yunibesiti e tla dirisa nne maiteko a go thusa baithuti ba pele ga kalogo ba ka nako e go ithuta.</p> <p>2. Setheo mmogo le Michael & Susan dell o tla tshegetsa baithuti ba ba nang le mo kotsing ya go gola ga kgotsa ka go bona ya ngwaga e e latelang ya thuto.</p> <p>3. Moreketoro le Motlatsamokanseliri Moporofesara wim de Villiers o rile covid-19 (e) tlišitse leroborobo e ne e le dikgwetlho tse diša le mathata a a yunibesiti e ne e setse e dira ka natla go fokotsa.</p> <p>4. “re sa ntse re ka bogale tlhokego ya gore baithuti ba rona ba lebagane le dikgwetlho mo go twe o mošwa le dintsi tse a mokgwa wa go ithuta ka nako ya lockdown. Ka jalo, re dira sengwe le sengwe se re ka se kgonang go netefatsa</p>	<p>Yunibesiti ya Stellenbosch e tšile go neela baithuti tshegetso ya dithuto le kamano magareng ga maitshwaro le dikakanyo tsa baithuti mo loagong News24 Mokwadi: Ntwaagae Seleka E phasaladitswe ka: 17/08/2020</p> <p>1. Yunibesithi ya Stellenbosch e tshepeditse go tiisa tshegetso ya dithuto mmogo le ya tsa maikutlo le ditlhaloganyo tsa baithuti ba sekolo. Yunibesithi e tšile dirisa ditsela tsa go dirisana mmogo go leka go thusa baithuti ba basenadikirii ka nako eno ya go rutwa ka tiriso ya mafatlharatlhara.</p> <p>2. Ntlo ya thutelo mmogo le ba ga Michael & Susan Dell foundation ba tsile go tshegetsa barutwana bao ba leng mo kotsing ya go se falole kgotsa go tšwelela pele go ya kwa ngwageng o o latelang wa dithuto tsa bona.</p> <p>3. Moreketoro le Motlatsamokanseliri wa Yunibesithi Wim de Villiers o buile gore bolwetsi jwa Covid-19 bo tlišitse seemo se se bokete le kotsi tseo Yunibesithi e dirang ka thata go di fokotsa.</p> <p>4. “Re nna re itse ka seemo se se thata se se šwa seo baithuti ba rona ba lomoganyeng le sona se e seng tlwaelo mo go bona ka mokgwa o mo ntšhwa wa go</p>

<p>students – and especially those who are vulnerable – can successfully complete the academic year,” added De Villiers.</p> <p>5. The university has made available laptops on a loan basis to financially disadvantaged students and data bundles have been procured for all students to access the internet.</p> <p>6. De Villiers said the university’s study management system had been boosted to handle the increased demand now that most activities were taking place online.</p> <p>7. “In addition, the university’s Centre for Student Counselling and Development is now providing online and telephonic services to those who need academic, emotional or emergency assistance during these times.</p> <p>8. “A grant provided by the Michael & Susan Dell Foundation, a US-based philanthropic non-profit organisation, will go towards tailored academic and psychosocial support for students who are vulnerable and at risk of not graduating or moving on to their next year of study. The grant will bolster academic support to ensure that every student stays on course, despite the challenges that they may experience.</p> <p>9. “We are grateful to the Michael & Susan Dell</p>	<p>gore baithuti ba rona ba – mme segolobogolo bao ba ba mo matshosetsing – ka go feleletsa ngwaga ya akatemi,” oketsa de Villiers.</p> <p>5. Yunibesiti e ka bonwa ka ntlha ya go reka dilepothopo tsa lekgotshwana baithuti ba ba humanegileng le tshedimose tso di rekilweng go baithuti botlhe go fitlhelela inthanete.</p> <p>6. De Villiers o rile gore Yunibesiti e ’ S thulaganyo ya tsamaiso ya thuto e ne e setse e maatlafaditswe go dirisa topo e jaanong e leng gore bontsi jwa ditiro tse di diragalang mo inthaneteng.</p> <p>7. “mo godimo ga moo, Yunibesiti e ’ S Senthara ya Go newa Kgakololo ga Baithuti le Katiso e e mo inthaneteng le go tlamela ka ditirelo tsa mo mogaleng go bao ba e tlhokang, maikutlo kgotsa thuso ya tshoganyetso mo go tsona.</p> <p>8. “kabo e e neelwang ke Michael & Susan dell Foundation, us-based philanthropic sa Setheo, e tla dira go le gontsi mo go tsa akatemi le kemonokeng e e diretsweng jwa ya baithuti ba ba nang le mo kotsing ya go gola ga kgotsa ka go bona ya ngwaga e e latelang ya thuto. Thebolelo e tla gagamatsa Kemonokeng ya Akatemi go tlhomamisa gore moithuti mongwe le mongwe go dula mo khosong, le fa go na le dikgwetlho tse di ka maitemogelo.</p> <p>9. “re itumelela go Michael & Susan dell motheo, e e</p>	<p>ithuta mo nakong eno. Ke ka moo re dirang ka thata go netefatsa gore baithuti botlhe ba rona, segolobogolo bao ba leng mo kotsing ba kgone go fetsa dithuto tsa bona tsa ngwaga” go tlaleletsa De Villiers.</p> <p>4. Yunibesithi e dirile leano la go nna teng ga dilepothopo tse di tlaa adimiwang baithuti ba ba humanegileng, mme ba fane gape le ka deitha gore baithuti botlhe ba kgone go fitlhelela inthanete.</p> <p>6. De Villiers o buile gore batsamaisi ba dithuto tsa Yunibesithi ba tlhatlhositse go tshwara kgodiso e e kwa godimo jaanong ka ntlha ya bontsi jwa ditiro di dirwa ka kgo lagano ya inthanete mo moweng.</p> <p>7. “Go tlaleletsa, bogare jwa tshegetso le tlhabololo ya Yunibesithi bo tlamela ka tirelo ka mogala le ka tiriso ya mowa ka inthanete go ba ba tlhokang thuso ka tsa thutego, maikutlo kgotsa tshoganyetso mo dinako tseno”.</p> <p>8. “Kabelo e e neilweng ke ba Michael & Susan Dell, setheo sa ba ditiro tsa bopelontle ba go direla ba bangwe dilo tse di molemo e e ikaegileng kwa US, e tla ya kwa dithuto tseo di tshwanelang le kwa tshegetso ya tsa tlhologanyo go baithuti ba ba amegang le ba ba leng mo kotsing ya go se aloge kgotsa go tlhatlhogela kwa ngwageng o mongwe wa go ithuta.</p> <p>9. “Kabelo e, e tlile go tshegetsa dithuto tsa baithuti go netefatsa gore morutwana mongwe le mongwe o tswela pele ka dithuto tsa gagwe le fa ba kopana le dikgwetlho”.</p>
---	---	--

Foundation, whose support strengthens our interventions to help our students succeed,” he added.	nonotshang go tshegetsa ditsereganyo tsa rona tsa go thusa baithuti ba rona ba atlega,” a tlaaletsa.	10. “Re leboga ba motheo wa Michael le Susan Dell, ka thuso ya bona e nonofisang kgato ya rona ya go tshegetsa baithuti ba rona go tswela pele.
--	--	---

The original article has nine short paragraphs and a total of 323 tokens. The MT system output from English into Setswana was produced by the MTWS in October 2020. It consists of nine paragraphs and contains 414 tokens. The HT was translated manually by a Setswana master’s student from the Tshwane University of Technology using dictionaries and other linguistic materials. It is ten paragraphs long and has 433 tokens. According to Prinsloo and De Schryver (2002), both English and Setswana are written disjunctively, resulting in a slight difference in the word count.

4. Data analysis

This section details a comparative analysis of linguistic issues found in the two translations. In the examples provided, ST refers to the text from the original article, MT is the text generated by the MTWS, and HT represents the output created by the recruited student. The analysis is presented in terms of case sensitivity, word omission, word addition, untranslated words, mistranslation, and terminological inconsistencies.

4.1 Case sensitivity

Case sensitivity is fundamental to and characteristic of written communication. It distinguishes between capital and lowercase characters. Capitalization not only can signify the beginning of a phrase and highlight specific sorts of words in a sentence, it also assures the precision of meaning and the efficiency of writing (Pathan 2021:65). According to Swick (2005), words can be capitalized in a variety of contexts, including at the beginning of sentences, for the official name of businesses or institutions, for proper names, for personal titles, to indicate the day and month, and for all the letters of an acronym. Using the correct case is essential in translation as it is one of the primary criteria that determines translation quality. While there were no case sensitivity errors in the HT, the MTWS made multiple case sensitivity mistakes in its translation, thus affecting readability. Here is an example where case sensitivity was compromised:

ST: Rector and vice-chancellor Professor Wim de Villiers said the Covid-19 pandemic had brought with it many new challenges and risks the university had been working hard to mitigate.

MT: *Moreketoro le Motlatsamokanseliri Moporofesara wim de Villiers o rile covid-19 (e) tlisitse leroborobo e ne e le dikgwethlo tse diša le mathata a a yunibesithi e ne e setse e dira ka natla go fokotsa.*

HT: *Moreketoro le Motlatsamokanseliri wa Yunibesithi Wim de Villiers o buile gore bolwetsi jwa Covid-19 bo tlisitse seemo se se bokete le kotsi tseo Yunibesithi e dirang ka thata go di fokotsa.*

The proper nouns “Professor Wim de Villiers” and “Covid-19” are all capitalized in the ST. The MTWS, however, did not recognize these as proper nouns; therefore, it failed to capitalize the first letters; that is, the “w” in “Wim” and the “c” in “Covid-19”. This failure showed the MTWS's limitation in terms of case sensitivity. By comparison, the human translator was able to detect and capitalize these proper nouns as expected. Other examples of the MTWS's failure to capitalize proper nouns are found in paragraph 8: the “D” in Susan Dell's surname and “US”, the abbreviation for the United States. The MTWS should have produced a translation with these proper nouns capitalized identically to the ST and HT, irrespective of their positioning in the sentence.

The MTWS also demonstrated inconsistencies in capitalization by omitting necessary capital letters and improperly capitalizing words that should have remained lowercase. In paragraph 1, the words *Akatemi* and *Kemonokeng* should be in lowercase as they are not proper nouns and do not begin a sentence. The system began sentences using lowercase instead of capital letters as shown in paragraphs 4, 7, 8 and 9. These examples show that the lack of human cognitive competence causes the MTWS to be incapable of recognizing the grammar rules for case sensitivity (Wu et al. 2016). Shi, Huang, Jian and Tang (2020) asserted that such errors may be caused by the translation performance of the MT system, which declines significantly when case-sensitive assessment metrics are introduced. Shi et al. (2020) recommended inserting case tokens into word sequences to convey nearby case information and highlight the use of an additional decoder for case prediction as it is crucial for addressing case sensitivity in automatic translation.

4.2 Word omissions

One of the most common translation methods that should be used carefully by translators is word omission. Although the technique of omission in translation might be used to overcome non-equivalence across languages, how it is utilized, particularly by translators, is questionable (Dulari & Amarasinghe 2020:56). Baker (1992) advised that this technique should be employed only when the meaning given by a specific item or word is not critical to the development of the text. Baker further asserted that when an item does not convey the meaning from the source language to the target language, the translator can then remove that specific term. This technique should, therefore, not alter the meaning of the sentence or deprive the intended readers of important information from the original text. Word omission may be described as intolerable when significant information is omitted from the translation and the entire sentence from the original language loses its meaning (Mmaboko 2005). The following example shows where the MTWS employed word omission:

ST: Stellenbosch University has promised to strengthen academic and psychosocial support for its students. The university will utilise collaborative efforts to help undergraduate students during virtual learning.

MT: *Yunibesiti ya Stellenbosch e solofeditse go nonotsha boemo jwa tlhaloganyo le mo go tsa Akatemi le Kemonokeng ya baithuti. Yunibesiti e tla dirisa nne maiteko a go thusa baithuti ba pele ga kalogo ba ka nako e go ithuta.*

HT: *Yunibesithi ya Stellenbosch e tshepeditse go tiisa tshegetso ya dithuto mmogo le ya tsa maikutlo le ditlhaloganyo tsa baithuti ba sekolo. Yunibesithi e*

tlile go dirisana mmogo ka go leka go thusa baithuti ba basenadikirii ka nako eno ya go rutwa ka tiriso ya mafaratlhatlha.

The ST words “virtual” and “collaborative” were accurately translated in the HT but were omitted by the MTWS, which led to the intended message losing its original meaning. Such omission is viewed as an intolerable omission of words in the text and is considered a high-risk practice for MT systems (Lotz & van Rensburg 2016:78). An MT system's difficulties in linking words to contexts leads to word omission since the meaning of those words varies depending on the context in which they are used (Precup-Stiegelbauer 2013). Furthermore, most of these words are omitted by MT systems, because accurate translation requires knowledge of context along with the structure and rules of a language. Such difficulties may be addressed simply by incorporating humans into the MT's post-editing of the TT (Wu et al. 2016; Skosana & Mlambo 2021).

4.3 Word additions

When faced with obstacles of lexical equivalents and pragmatics, one of the most common strategies used by translators is word addition (Khanmohammad & Aminzad 2015:8). Other reasons for adopting this strategy include structural and grammatical problems encountered by translators. Such situations may prompt translators to add words to the text to make it more understandable and to ensure it aligns more closely with the cultural elements of the target language. Khanmohammad and Aminzad (2015:8) pointed out that languages are neither systematically nor culturally identical; therefore, word addition is unavoidable in the translation. Moreover, Makamu (2017) classified this strategy as “foreignising” as it is used commonly when translating cultural words from the source to the target language. Although the human translator did not adopt this strategy, the following examples demonstrate how the MTWS was unable to use word additions effectively:

ST: The university will utilise collaborative efforts to help undergraduate students during virtual learning.

MT: *Yunibesiti e tla dirisa nne maiteko a go thusa baithuti ba pele ga kalogo ba ka nako e go ithuta.*

HT: *Yunibesithi e tlile go dirisa ditsela tsa go dirisana mmogo go leka go thusa baithuti ba basenadikirii ka nako eno ya go rutwa ka tiriso ya mafatllharatlhara.*

The MTWS distorted the meaning of the TT by inserting numerous unwanted words. The system added the word *nne*, which translates to “four” in Setswana. There is no reference to the number four in the ST; therefore, the added word constitutes an error as it makes the TT inaccurate. Other examples found in paragraphs 3, 6 and 7, were “a”, “e”, and “S”, which were added unnecessarily and repeatedly. These words do not contribute to the meaning of the sentences and instead create confusion for the target readers. Wu et al. (2016) suggested three phases of post-editing to deal with any confusion or uncertainty produced by the MT systems. In the first phase, the translator runs over the TT and finds the MT output errors. The translator in the second phase focuses on fixing translation errors such as word additions that may contribute to text confusion or uncertainty. Lastly, the translator proofreads by double-checking

the TT to confirm that any MT errors have been corrected. These phases demonstrate the significance of post-editing to improve the outputs of MT systems.

4.4 Untranslated words

Words that are not translated from a source language to a target language are usually associated with a translation using loan words. Makamu (2017) observed that this strategy involves human translators preserving cultural terminology by retaining it in its original source language rather than translating it. This approach is ideal when the translator has difficulty translating culturally unique terms, modern concepts, or buzzwords. Makamu (2017) emphasized that providing explanations alongside loan words enhances readers' understanding. This translation strategy is particularly effective for unfamiliar loan words in the target culture and works even better when accompanied by contextual clarification (Mmaboko 2005). Untranslated words by the MTWS, on the other hand, were the result of unfamiliar or unknown terminology from training corpora for the MT system. According to Gulcehre, Ahn, Nallapati, Zhou and Bengio (2016:266), the problem of unfamiliar or unknown words is a significant issue that might affect the performance of various NLP systems, including MTWS. The following example shows words that were left untranslated by the MTWS:

ST: “A grant provided by the Michael & Susan Dell Foundation, a US-based philanthropic non-profit organisation, will go towards tailored academic and psychosocial support for students who are vulnerable and at risk of not graduating or moving on to their next year of study.

MT: “*kabo e e neelwang ke Michael & Susan dell Foundation, us-based philanthropic sa Setheo, e tla dira go le gontsi mo go tsa akatemi le kemonokeng e e diretsweng jwa ya baithuti ba ba nang le mo kotsing ya go gola ga kgotsa ka go bona ya ngwaga e e latelang ya thuto.*

HT: “*Kabelo e e neilweng ke ba Michael & Susan Dell, setheo sa ba ditiro tsa bopelontle ba go direla ba bangwe dilo tse di molemo e e ikaegileng kwa US, e tla ya kwa dithuto tseo di tshwanelang le kwa tshegetso ya tsa tlhaloganyo go baithuti ba ba amegang le ba ba leng mo kotsing ya go se aloge kgotsa go tlhatlhogela kwa ngwageng o mongwe wa go ithuta.*

In the example, “US-based philanthropic” was translated correctly in the HT and left untranslated by the MTWS. Similarly, the system did not translate the word “lockdown” in the fourth paragraph. These untranslated areas show insufficient vocabulary input into the MTWS. In other words, the MTWS needs to be updated with new training corpora from diverse fields as it is unfamiliar with words that are beyond the training data. The expectation is that incorporating more diverse training data when updating MT systems will reduce the occurrence of unfamiliar or unknown words in these systems. Turganbayeva and Tukeyev (2021:215) suggested that such issues may also be overcome by using input data during the MT systems’ pre-processing, which substitutes the unknown words with synonyms. For such a technique to be effective, a database of synonyms for each word in the TT should be compiled first and followed by corpus segmentation. As a result, when unfamiliar words are searched for in the training data, they can be substituted with similar words from the synonyms dictionary. This technique has worked for resource-scarce languages (Turganbayeva & Tukeyev 2021:220).

Therefore, it has potential to be effective for the MT systems for South African indigenous languages.

4.5 Mistranslations

Ezeafulukwe (2019:140) defined mistranslation as a lack of faithfulness in the translated text; that is, the translation has lost information from the ST or acquired information that was not in the ST. Zhou (2015) stated that mistranslation may be caused by three factors: the translator's concession to the clients, misunderstanding of the original text, and cultural ineptitude on the part of the translator. Zhou (2015) further pointed out that a successful translation can be done only with a thorough understanding of the original text, which places a great demand on the translator's abilities. Unfortunately, many translators do not attempt to include helpful explanations in their translations, resulting in mistranslations. Translators who are unfamiliar with other cultures typically use common sense to understand the original material, which also results in mistranslation. Consider the following example where the MTWS mistranslated the text:

ST: The university has made available laptops on a loan basis to financially disadvantaged students and data bundles have been procured for all students to access the internet.

MT: *Yunibesiti e ka bonwa ka ntlha ya go reka dilepothopo tsa lekgotshwana baithuti ba ba humanegileng le tshedimose tso di rekilweng go baithuti botlhe go fitlhelela inthanete.*

HT: *Yunibesithi e dirile leano la go nna teng ga dilepothopo tse di tlaa adimiwang baithuti ba ba humanegileng, mme ba fane gape le ka deitha gore baithuti botlhe ba kgone go fitlhelela inthanete.*

The words “data bundles”, which refers to a package that gives one access to the internet, was mistranslated by the MTWS as “information”, whereas the HT used a borrowed word and translated it as *deitha*. This discrepancy occurred because the human translator performed translation based on the knowledge of the language, but the MT systems operated solely with artificial intelligence. Li and Lu (2021) confirmed that MT systems can translate simple sentences but have difficulties translating long sentences or sentences with a slightly complex structure. Skosana and Mlambo (2021:7) proposed that the MTWS’s shortcomings of providing incorrect vocabulary, spelling, tenses, and sentence structuring, which can result in mistranslations, can be prevented by involving language experts in the development and pre-editing of the MT systems’ training data. MT systems’ difficulties are also caused by the fact that the systems tend to translate the sentence according to the grammar rules of the original text. In these instances of mistranslation, the systems translate each word as it occurs in the ST, resulting in confused sentence structures and unclear meaning, thus affecting readability (Li and Lu 2021). As Wu et al. (2016) submitted, post-editing of the MT systems’ outputs can improve the quality of the translation.

4.6 Terminological inconsistencies

Gašpar, Seljan and Kučič (2022:2) defined terminological inconsistency as a usage of two or more words used to translate a single term in a ST. The inconsistent use of words across many languages continues to be a challenge for language users, machine translators and human translators. Terminology inconsistency in MT systems can be caused by various factors, including a lack of standardized terminology and poor machine training data (Itagaki, Aikawa & He 2007:1). The inconsistencies in translations can be attributed to the lack of effort by language experts and terminologists in dealing with foreign or new terminology. Terminology inconsistencies often arise when language specialists create new terms without consulting end-users, leading to rejection due to perceived inaccuracy (Sarairoh 2001). Additionally, software developers and human translators may struggle with excessive terminologies, unsure of their proper usage. Examples of compromised consistency are as follows::

ST: So, we are doing everything we can to ensure that all of our students – and especially those who are vulnerable – can successfully complete the academic year,” **added De Villiers.**

MT: *Ka jalo, re dira sengwe le sengwe se re ka se kgonang go netefatsa gore baithuti ba rona ba – mme segolobogolo bao ba ba mo matshosetsing – ka go feleletsa ngwaga ya akatemi,” oketsa de Villiers.*

HT: *Ke ka moo re dirang ka thata go netefatsa gore baithuti botlhe ba rona, segolobogolo bao ba leng mo kotsing ba kgone go fetsa dithuto tsa bona tsa ngwaga” go tlaleletsa De Villiers.*

ST: We are grateful to the Michael & Susan Dell Foundation, whose support strengthens our interventions to help our students succeed,” he added.

MT: *“re itumelela go Michael & Susan dell motheo, e e nonotshang go tshegetsatsa ditsereganyo tsa rona tsa go thusa baithuti ba rona ba atlega,” a tlaleletsatsa.*

HT: *“Re leboga ba motheo wa Michael le Susan Dell, ka thuso ya bona e nonofisang kgato ya rona ya go tshegetsatsa baithuti ba rona go tswela pele.*

The MTWS was inconsistent in translating the word “added” in the ST. In the first instance, the word “added” was translated as *oketsa*; in the second it was translated as *tlaleletsatsa*. Such inconsistent usage of the word may lead to confusion, ambiguity and misunderstanding. Itagaki et al. (2007:1) identified developers’ errors in MT system design as a source of terminology inconsistencies, often resulting from using translation memory tools that recycle translations without considering context. Itagaki et al. (2007:6) also affirmed that to reduce instances of terminological inconsistency in MT systems, developers should design a translation consistency index tool to verify the quality of translation data. The human translator, on the other hand, being flexible, shows the translation of the word “added” as *tlaleletsatsa* in the first occurrence and omitted in the second.

5. Conclusion

In this study, we compared the translation outputs of an MTWS and a human translator. The findings of this study revealed that the role of human translator in providing high-quality translation outputs still surpasses that of the MTWS as an MT system. The MTWS was unable to provide accurate translations in terms of case sensitivity and terminology consistency. It omitted and added words that changed the intended meaning of the ST, and left certain words untranslated. The MT system also showed it is prone to mistranslating words since it does not consider the context in which the words occur. The human cognitive competence, intelligence and flexibility were identified as abilities that aided the human translator in producing better translations. These findings indicate that human translators are more effective and efficient than MT systems, which often translate texts without considering the intended context or purpose of the target language. The study also demonstrated the significance of post-editing the MT outputs to ensure readable and acceptable text. Therefore, all efforts to replace human translators with MT systems in rendering translation will fail, since the latter is incapable of interpreting and providing quality outputs.

References

- Abbott, Jade Z. and Laura Martinus. 2019. Benchmarking neural machine translation for southern African languages. *Proceedings of the 2019 Workshop on Widening NLP*. Florence, Italy. pp. 98-101.
- Adegbola, Tunde. 2009. Building capacities in human language technology for African languages. *Proceedings of the EACL 2009 Workshop on Language Technologies for African Languages*. Athens, Greece. pp. 53-58. <https://doi.org/10.3115/1564508.1564519>
- Ahrenberg, Lars. 2017. Comparing machine translation and human translation: A case study. *Proceedings of the 1st Workshop on Human-Informed Translation and Interpreting Technology*. Varna, Bulgaria. pp. 21-28. https://doi.org/10.26615/978-954-452-042-7_003
- Arvianti, Gilang Fadhilia. 2018. Human translation versus machine translation of Instagram's captions: Who is the best? *Proceedings 2nd English Language and Literature International Conference (ELLiC)*. Semarang, Indonesia. pp. 531-536.
- Baker, Mona. 1992. *In other words*. London: Routledge. <https://doi.org/10.4324/9780203133590>
- De Pauw, Guy, Gilles-Maurice De Schryver, Laurette Pretorius and Lori Levin. 2011. Introduction to the Special Issue on African Language Technology. *Language Resources and Evaluation* 45:263-269. <https://doi.org/10.1007/s10579-011-9157-9>
- Dulari, Achira and Harsha Amarasinghe. 2020. Investigation on the usage of omission as a method of translation: with special reference to English translation of “Heart of a dog” by Mikhail Bulgakov and its Sinhala Translation ‘Ballekuge Hadawtha’ by Gamini Viyangoda. *Iconic Research and Engineering Journals* 4(4):56–61. <https://irejournals.com/formatedpaper/1702512.pdf>

- Puttkammer, Martin and Jakobus S. Du Toit. 2022. Canonical segmentation and syntactic morpheme tagging of four resource-scarce Nguni languages. *Journal of the Digital Humanities Association of Southern Africa* 3(03):1-7. <https://doi.org/10.55492/dhasa.v3i03.3818>
- Ezeafulukwe, Olivia Udo. 2019. Justifying mistranslation. *New Journal of African Studies* 15:138-155. <https://doi.org/10.4314/og.v15i1.8>
- Finlayson, Rosalie and Mbulungeni Madiba. 2002. The intellectualisation of the indigenous languages of South Africa: Challenges and prospects. *Current Issues in Language Planning* 3(1):40-61. <https://doi.org/10.1080/14664200208668036>
- Gašpar, Angelina, Sanja Seljan and Vlasta Kučiš. 2022. Measuring terminology consistency in translated corpora: Implementation of the Herfindahl-Hirshman Index. *Information* 13:1-14. <https://doi.org/10.3390/info13020043>
- Gulcehre, Caglar, Sungjin Ahn, Ramesh Nallapati, Bowen Zhou and Yoshua Bengio. 2016. Pointing the unknown words. *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*. Berlin, Germany. pp. 140-149. <https://doi.org/10.18653/v1/p16-1014>
- Itagaki, Masaki, Takako Aikawa and Xiaodong He. 2007. Automatic validation of terminology translation consistency with statistical method. *Proceedings of Machine Translation Summit XI: Papers*. Copenhagen, Denmark. pp. 269-274. <https://aclanthology.org/2007.mtsummit-papers.36.pdf>
- Joffe, David and Gilles-Maurice De Schryver. 2004. TshwaneLex-a state-of-the-art dictionary compilation program. In Geoffrey Williams and Sandra Vessier (Eds.), *Proceedings of the eleventh EURALEX international congress*. Lorient, France. pp. 99-104.
- Khanmohammad, Hajar and Ali Aminzad. 2015. A study of additions and omissions in English translations of Persian-based social, cultural and political news headlines: A case study of ISNA news agency in Arab spring: Islamic Azad University. *T&I Review* 5:7-27. <http://doi.org/10.22962/tnirvw.2015.5..001>
- Li, Yuting and Xiuying Lu. 2021. Study on post-editing for machine translation of railway engineering texts. *SHS Web of Conferences*. 96:05001. EDP Sciences. <https://doi.org/10.1051/shsconf/20219605001>
- Lotz, Susan and Alta Van Rensburg. 2016. Omission and other sins: Tracking the quality of online machine translation output over four years. *Stellenbosch Papers in Linguistics* 46:77-97. <https://doi.org/10.5774/46-0-223>
- Makamu, Thembheka A. 2017. *Translation and interpretation of cultural concepts from Xitsonga into English*. Unpublished Doctoral Thesis, University of South Africa.
- Makgatho, Mack, Vukosi Marivate, Tshephisho Sefara and Valencia Wagner. 2022. Training cross-lingual embeddings for Setswana and Sepedi. *Journal of the Digital Humanities Association of Southern Africa* 3(03). <https://doi.org/10.55492/dhasa.v3i03.3822>

Mlambo, Respect and Muzi Matfunjwa. 2024. The use of technology to preserve indigenous languages of South Africa. *Literator* 45(1), a2007. <https://doi.org/10.4102/lit.v45i1.2007>

Mmaboko, Elliott M. 2005. SABC news in Sotho languages: A case study in translation. Unpublished MA dissertation, University of the Witwatersrand.

Nemutamvuni, Mulalo E. 2018. Investigating the effectiveness of available tools for translating into Tshivenda. Unpublished MA dissertation, University of South Africa.

Pathan, Md. Arif Khan, 2021. The Most Frequent Capitalization Errors Made By the EFL Learners at Undergraduate Level: An Investigation. *Scholars International Journal of Linguistics and Literature* 4(3):65-72. <https://doi.org/10.36348/sijll.2021.v04i03.001>

Precup-Stiegelbauer, Laura-Rebeca. 2013. Automatic translations versus human translations in nowadays world. *Procedia - Social and Behavioral Sciences* 70:1768-1777. <https://doi.org/https://doi.org/10.1016/j.sbspro.2013.01.252>

Prinsloo, D.J. and Gilles-Maurice De Schryver. 2002. Towards an 11x11 array for the degree of conjunctivism/disjunctivism of the South African languages. *Nordic Journal of African Studies* 11(2):249-265. <https://doi.org/10.53228/njas.v11i2.359>

Saraireh, Muhammad A. 2001. Inconsistency in technical terminology: A problem for standardization in Arabic. *Babel* 47(1):10-21. <https://doi.org/10.1075/babel.47.1.03sar>

Seleka, Ntwaagae. 2020. Stellenbosch University to offer academic and psychosocial support to students, News24, 17 August, p. 6. <https://www.news24.com/news24/southafrica/news/stellenbosch-university-to-offer-academic-and-psychosocial-support-to-students-20200817> (Accessed 17 August 2020).

Shi, Xuewen, Heyan Huang, Ping Jian and Yi-Kun Tang. 2020. Case-sensitive neural machine translation. *Proceedings of Pacific-Asia Conference on Knowledge Discovery and Data Mining*. PAKDD 2020, Singapore, May 11–14, 2020. https://doi.org/10.1007/978-3-030-47426-3_51

Skosana, Nomsa J. and Respect Mlambo. 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>

South African Centre for Digital Language Resources (SADiLaR). <https://sadilar.org/>

Swick, Ed. 2005. *English Grammar for ESL Learners*. USA: McGraw-Hill Companies.

Turganbayeva, Aliya and Ualsher Tukeyev. 2021. The solution of the problem of unknown words under neural machine translation of the Kazakh language. *Journal of Information and Telecommunication* 5(2):214-225. <https://doi.org/10.1080/24751839.2020.1838713>

Wu, Fang-fei, Qiu-ping Liang and Gao-da He. 2016. A comparative study of machine translation and human translation - A case study of Kingsoft Powerword. *Proceedings of the*

International Conference on Computational Modelling, Simulation and Applied Mathematics. Bangkok, Thailand, 24–25 July 2016. <https://doi.org/10.12783/dtcese/cmsam2016/3651>

Yin, Robert K. 2003. *Case study research: Design and methods*. Third Edition. Thousand Oaks California: Sage Publications.

Zhou, Jianchuan. 2015. On the mistranslation of cultural words in the English versions of Sun Tzu's Art of war. *Asian Social Science* 11(21):300-305. <https://doi.org/10.5539/ass.v11n21p300>

Zong, Zhaorong. 2018. Research on the relations between machine translation and human translation. *Journal of Physics: Conference Series* 1087(6):1-6. <https://doi.org/10.1088/1742-6596/1087/6/062046>