Machine Translation Systems as Enablers of Multilingualism and Access to Information in South Africa

Muzi Matfunjwa
South Africa Centre for Digital Language Resources
North-West University
Potchefstroom, South Africa

Respect Mlambo
South Africa Centre for Digital Language Resources
North-West University
Potchefstroom, South Africa

Introduction

Access to all kinds of information in all official South African languages (SALs) is a challenge even though the Constitution of the Republic of South Africa (1996) declares that all languages are equal. Information in South Africa is mainly available in English. This is detrimental to people who are not proficient in this dominant language as they cannot access crucial information in their home languages. Practical measures and interventions are essential to promote access to information in all South African languages. This is essential in the quest to promote multilingualism. This chapter discusses Machine Translation Tools that support SALs and how they can facilitate access to information and enhance multilingualism. Three Translation tools, namely Google Translate, isiZulu.net, and Siswati to English Translation Model, are employed to demonstrate how these tools can be utilised to promote access to information in SALs.

South Africa is a multilingual country and has twelve official languages, namely English, Afrikaans, isiNdebele, Sepedi, Sesotho, Siswati, Xitsonga, Setswana, Tshivenda, isiXhosa, isiZulu and the South African Sign Language. These South African Languages (SALs) must all be included in Human Language Technology (HLT) to ensure they are not digitally and technologically neglected. HLT tools are designed to process either text or speech. Text applications include text categorisation, translation technologies, text summarisation, text indexing, text retrieval, information extraction, question answering, report generation, data fusion and text data mining. Speech applications, on the other hand, consist of audiobooks, transcription and dictation, speech recognition, speech synthesis, spoken dialogue systems and speech to speech translation (Uszkoreit 2000; Grover et al., 2011).

There is disparity in the use of HLT among the SALs, which impedes access to information and multilingualism. Grover et al., (2011) established that Afrikaans was the most developed language, followed by South African English in terms of HLT for language resources and applications while Tshivenda, Xitsonga, isiNdebele, and Siswati, as minority languages, have the least resources. This causes these languages to lag behind in the development of HLT. Based on the South African Human Language Technology Audit (SAHLTA) that was conducted by Grover et al., (2011), it was discovered that Machine Translation (MT) was the field that needed dire attention and urgent development, especially for minority languages.

Over the past decade, since the SAHLTA was carried out, there has been a significant development of MT tools for SALs. MT is a contemporary form of translation that is facilitated by computers (Lin & Chien, 2009). MT tools execute automated translations of text and speech from a source language to a target language. The primary purpose of this is to get translations of good quality that are equivalent to human translation (Poibeau, 2017). The significance of MT tools is that they enable language processing, ensuring that languages exist digitally, and are preserved (Motlani. 2016). They also allow individuals globally to interact in their home languages

while finding information in a language they do not understand (Vieira et al. 2021). This shows that MT is essential in the modern era to serve the linguistic needs of language users. The creation of MT tools for SALs is a positive drive in raising their status and executing the directive of the Constitution of the Republic of South Africa (1996: 4), which states that "… the state [the government of South Africa] must take practical and positive measures to elevate the status and advance the use of these languages".

In a multilingual country like South Africa, MT tools provide an ideal environment for information accessibility and multilingualism to thrive. Currently, most official SALs are represented in various MT applications such as Google Translate (GT), isiZulu.net (IZN), and Siswati to English Translation Model (SETM). These applications are open access to every language user. This chapter aims to discuss these MT applications and describe how they could be utilised to promote multilingualism and access to information in various SALs.

Hutchins (1995) provided a comprehensive history of MT from its inception in the 1950s to the present. The early work on MT was reported to be mostly focused on rule-based systems. These systems could translate simple texts but struggled with more complicated texts containing idiomatic or cultural phrases. There was a noticeable shift towards statistical MT systems in the 1970s. These systems produced more accurate translations than rule-based systems since they were based on the analysis of large parallel corpora. Statistical MT systems, however, were unable to create translations that were as accurate or genuine as those produced by human translators. The advanced computer processing power and natural language processing fostered renewed interest in MT in the 1990s. This resulted in the creation of neural MT systems also based on large corpora of data. Neural MT systems have significantly improved translation accuracy and fluency, and they are now the commonly used form of MT. Hutchins added that while MT is still in its early stages, it has the potential to transform the way humans communicate across languages. The author predicted that MT would become more precise and fluent and eventually

produce translations that are identical to those generated by human translators.

Bowker and Ciro (2015) investigated how MT can meet the linguistic needs of newcomers at the Ottawa Public Library in Canada. The authors adopted a community-based participatory research method to engage Spanish-speaking newcomers in the study to ensure that the findings are valuable and beneficial to the community. The study was divided into three phases: assessing the MT system, surveying the target community, and evaluating the recipients. Three free MT systems, namely Google Translate, Reverso, and Systran were evaluated to determine which one produced the most accurate translations of library-related texts. A survey of Spanish-speaking newcomers was conducted to understand their information needs and preferences of different translation services. The authors conducted a recipient evaluation to assess the utility and acceptability of the different types of MT output to Spanish-speaking newcomers. The study revealed that Google Translate was the most effective MT in meeting the linguistic needs of newcomers at Ottawa Public Library. However, it was not always perfect and post-edited output from a human translator was needed.

Kituku et al., (2016) explored three MT system approaches: rule-based MT, data-driven MT, and hybrid MT. It was established that rule-based MT systems use a set of rules to translate text from one language to another. These rules are typically based on the grammar and vocabulary of the two languages. Rule-based MTs are relatively easy to develop, but they do not produce accurate translations. Meanwhile, the data-driven MT, also known as corpus-based MT, relies on aligned parallel corpora. Parallel corpora are aligned through annotation, and a classifier is generated using supervised, semi-supervised, unsupervised, or bootstrapping learning methods with artificial intelligence that can use statistical probability, clustering, or classification approaches. This study determined that this approach is simple to construct and less expensive, but requires large parallel corpora, which may not exist for under-resourced languages unless they are first created. The

authors established that the hybrid MT comes in to exploit the merits of both rule-based and data-driven MT in terms of coverage and accuracy. In this approach, rule-based MT applies grammar rules that analyse syntax, semantics, morphology, part of speech tagging, and orthographic aspects of the source and target languages. The rules are then implemented into the corpus system from data-driven MT, either at the pre/post-processing stage or at the system's core model.

Castilho (2016) explored end-user acceptability measures for machine-translated enterprise content. The author assessed the end-user acceptability of MT system output in terms of usability, quality, and satisfaction. The study used English as a source language and German, Japanese, and Simplified Chinese as target languages. It was ascertained that the usability of MT outputs in these languages might be evaluated through tasks that are recorded by utilising eye tracker. Meanwhile, the quality of MT outputs can be verified using a translation quality assessment questionnaire completed by experienced human translators, and the source content is evaluated using metrics such as readability and syntactic complexity. The satisfaction can be monitored using surveys, post-task questionnaires, and the rankings of human translators. The author also assessed the acceptability of various levels of post-editing for the three target languages, and the results show that implementing light post-editing has a direct and positive influence on acceptability for German and Simplified Chinese languages, but not for Japanese. It was concluded that the aspect of post-editing had a substantial effect on acceptability, with the light post-edited versions displaying greater levels of acceptability when compared to the raw machine-translated versions in terms of usability, quality, and satisfaction.

Nurminen and Koponen (2020) examined the potential of MT to improve fair access to information for underserved groups. The authors showed the importance of access to information as a human right. They argued that language can be an obstacle to access to information, especially for people who speak minority languages or who have limited literacy skills. MT can help to overcome this obstacle by providing access to

information in different languages. The authors also offered an overview of examples of initiatives that used MT to increase fair access to information for underprivileged groups in three contexts: civic participation, public health, and safety, as well as media and culture. They added that MT systems are being used to translate government documents, election materials, and other information necessary for civic participation. Meanwhile, in public health and safety, MT systems are being utilised to translate health information, safety warnings, and other critical information.

In the fields of media and culture, the MT systems are used to translate news, books, movies, and other cultural products so that anyone can enjoy them regardless of their language ability. The authors also examined some of the ethical difficulties that occur when using MT for fair access to information, such as quality, acceptability, and stakeholder involvement. These difficulties arise because MT output might be mistranslated which impacts information reliability. It was recommended that the stakeholders, particularly those disadvantaged by traditional translating methods, should be involved in the creation of MT systems to ensure quality and accuracy.

Andrabi and Wahid (2021) investigated MT approaches utilised for South Asian poor-resourced languages to identify their needs and challenges. Their study determined that MT approaches are classified as rule-based, corpus-based, hybrid, and knowledge-based. The rule-based MTs which do not require a large parallel corpus, were mostly utilised in South Asian poor-resourced languages due to the non-availability of parallel data to build various systems. It was argued that the advancements of the internet are required for users of these poor-resourced languages to acquire information in their native languages with the help of MT systems. The study also addressed the challenges associated with MT systems for poor-resourced languages such as lack of parallel corpus, lexical ambiguity, referential ambiguity, word order issues, sentence alignment, prepositions and postpositions, and morphological variance.

Steigerwald et al., (2022) discussed the challenges and opportunities of using MT systems in academia to overcome language barriers. They highlighted the importance of multilingualism in science and the barriers faced by non-English speaking scientists due to the dominance of English. They explored the potential benefits of using MT to overcome language barriers in academia, such as increasing scientific knowledge accessibility to a wider audience, promoting international collaboration in research, facilitating the training of scientists from underrepresented groups, and supporting the development of multilingual scientific communities. The current state of MT tools was recognised, and it was observed that their recent advancements allowed more accurate and fluent translations of scientific texts. The authors also noted the difficulties associated with using MT in academia as these tools are not perfect and can produce inadequate and erroneous translations. MT should not be viewed as a replacement for human translation, but rather as a tool to assist human translation efforts. The authors proposed that language barriers in academia can be overcome by creating high-quality MT systems specifically for scientific texts, providing effective training to universities and research institutions, and developing policies and practices that support MT use in academia.

Tuominen et al., (2023) conducted a study to investigate the potential of automated subtitling for linguistic accessibility. Automated subtitling is the use of MT to generate subtitles for video footage. The study investigated English-speaking viewers watching video clips of Finnish news and current affairs with automated English subtitles. The authors contended that linguistic accessibility in media is vital for a variety of reasons. They include ensuring that individuals with limited language ability can fully participate in society, promoting cultural understanding and tolerance, and providing access to information and entertainment. They highlighted that the need for automated subtitling stems from the challenges of providing linguistic accessibility to media content, such as the cost of translating and subtitling information into multiple languages, as well as the difficulty of acquiring qualified translators. The

study found that viewers could understand the gist of the content using automatic subtitles, but there were a few deficiencies in the quality of the subtitles. The authors argued that automated subtitling using MT could enhance multilingual accessibility for Finnish broadcaster Yle by supporting underserved languages, but issues like subtitle quality and user feedback need to be addressed before they can be widely adopted.

From the literature reviewed, we observed how MT applications are developed to access information in target languages. Related work shows that MT has been used to promote multilingual access to information. However, to the best of our knowledge, there has been no study on the use of MT tools that support SALs to make information openly accessible in various languages. Thus, this chapter aims to discuss MT applications namely GT, IZN and SETM, and how these applications could be utilised to promote multilingualism and access to information in various SALs.

Discussion

Machine Translation for South African Languages

This section describes MT applications, namely GT, IZN, and SETM, which have the most official SALs.

Google Translate

GT is a free online translation application that enables users to translate words, sentences, documents, and websites between different languages (van Lieshout & Cardoso, 2022). It was founded in 2006. Currently, it provides translation services for over 100 languages, including well-resourced and under-resourced languages (Johnson et al., 2017). GT was developed using very large corpora from various fields, rather than sets of rules for a specific language (Lotz & van Rensburg 2014). With the recent advancement of artificial intelligence, GT has transitioned from operating as a statistical MT to a neural MT system, which has substantially increased translation accuracy and quality (Sun et al., 2022). GT can be accessed freely using a

web interface, a mobile app, or integrated into other applications through the Google Translate Application Programming Interface (API) with an internet connection (Seçkin, 2023).

In addition to English, GT supports the following SALs; Afrikaans, isiZulu, isiXhosa, Sesotho, Xitsonga, and Sepedi as shown in Figure 1 below. Afrikaans was the first SAL to be made available in GT (Muller, 2009). IsiZulu, isiXhosa, and Sesotho then followed while Xitsonga and Sepedi were added later to GT (Illidge, 2022).

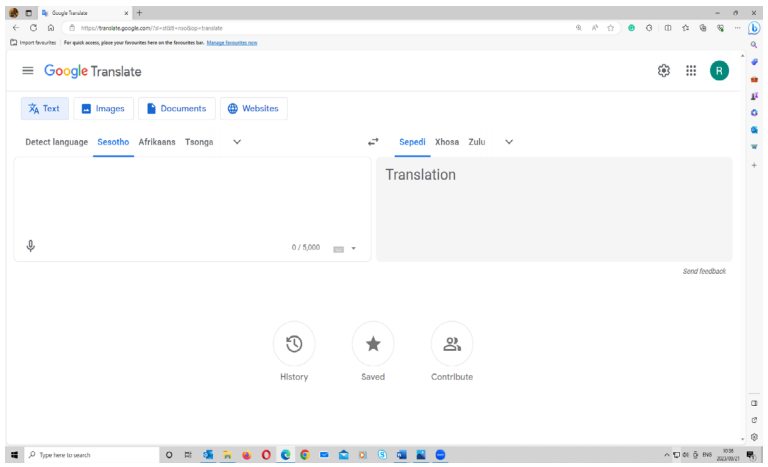

Figure 1: Google Translate interface with South African Languages

Figure 1 shows the SALs available on GT, namely Afrikaans, isiZulu, isiXhosa, Sesotho, Xitsonga, and Sepedi. The availability of these languages provides language users with an opportunity to translate from any source language of their choice in GT to a target language.

IsiZulu.net (IZN)

IZN is an open-access online bilingual isiZulu-English dictionary that offers word and phrase translation equivalents from isiZulu to English. The IZN system also offers "bidirectional word, automatic morphological decomposition, conjugation and

phonetic transcription of isiZulu words, spell checking, as well as grammar and pronunciation basics" (https://isizulu.net/). Prinsloo (2010:184) emphasised that in the IZN system, users are afforded comprehensive services on isiZulu such as word formations, translations, pronunciation guides, word categories, word usage, and definitions as illustrated in Figure 2.

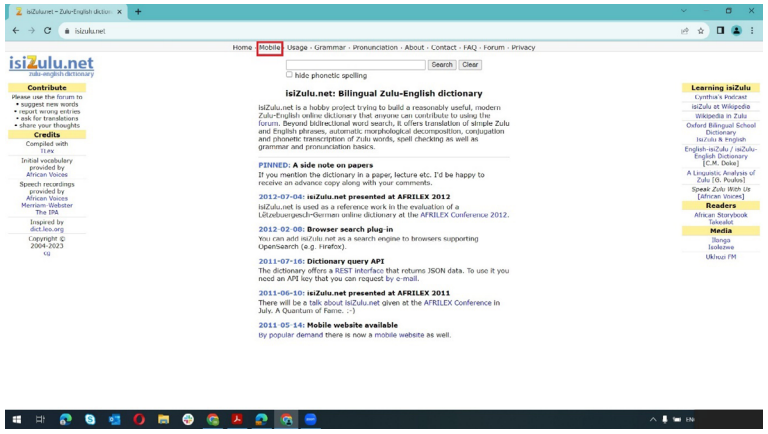

Figure 2: isiZulu.net interface

Figure 2 shows the IZN interface. Since IZN is a multifunctional system, the isiZulu-English Dictionary is accessed through the 'Mobile' feature which is highlighted in red in the Figure.

Siswati to English Translation Model (SETM)

SETM is an MT system that was created using an aligned Siswati and English parallel corpus that was retrieved from the South African Government website (https://huggingface.co/dsfsi/ss-en-m2m100-gov). The SETM can only translate from Siswati to English and it is still under refinement. This model is shown in Figure 3.

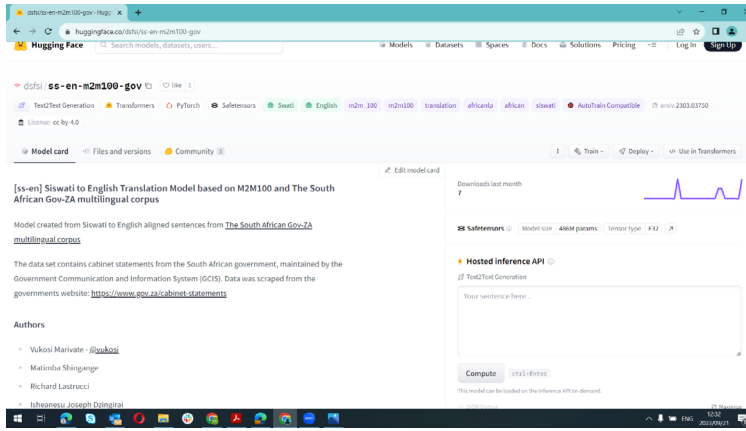

Figure 3: Siswati to English Translation Model interface

In Figure 3, the SETM graphical user interface is shown in which Siswati text can be put under 'Hosted inference API' text input, and the feature 'Compute' is used to automatically obtain the English translation.

Access to Information and Multilingualism Through Machine Translation

MT plays a vital role in modern societies where monolingualism is no longer the norm. MT provides opportunities in which access to information and multilingualism can be achieved. The aim of MT is to provide language users with high-quality translations that are equivalent to human translations but are not meant to replace human translators (Poibeau, 2017). Despite the imperfections of MT, it plays a crucial role in our communities, particularly in facilitating information access and promoting multilingualism. In this section, the researchers discuss how MT applications such as GT, IZN, and SETM can be utilised to facilitate access to information and promote multilingualism in various SALs. Access to information and multilingualism can be achieved through numerous ways such as breaking down language barriers, language learning, enhanced multilingual content creation, and localisation of websites.

Breaking Down Language Barriers

Language can serve as both a bridge and a barrier in communication, preventing people from conveying simple, clear, and accurate information (Abuarqoub, 2019). Information mainly exists in English and peer-reviewed articles, such as scientific publications, are dominantly in this language (Liu, 2017). This is only good for sharing global information among people who understand English. However, people who are not fluent in or have minimal understanding of English are deprived of crucial knowledge as they cannot fully comprehend the information written in this language. This obstacle is caused by differences in language competency or capability, as well as improper language levels, jargon, or slang that the recipient does not understand (Rani, 2016). Abuarqoub (2019) further asserted that language barriers hinder relationships, leading to conflict, frustration, offense, violence, hurt feelings, wasted time, effort, and money. Language barriers in South Africa significantly hinder access to essential services like healthcare, business, and education (Hussey, 2012). These problems can be alleviated by using MT tools such as GT and SETM that produce translations between SALs. GT is crucial for language users to easily access and understand information that was originally in a foreign language, thus breaking barriers as shown in Figure 4a and Figure 4b in which the English healthcare article retrieved from (https://www.cancer.gov/about-cancer/understanding/what-is-cancer) is translated into isiZulu. Additionally, in Figure 5a and Figure 5b, an abstract written in English which was retrieved from an article titled 'The extraction of terminology list using ParaConc for creating a quadrilingual dictionary' (Mlambo et al., 2020) is translated into Afrikaans and Xitsonga in that order.

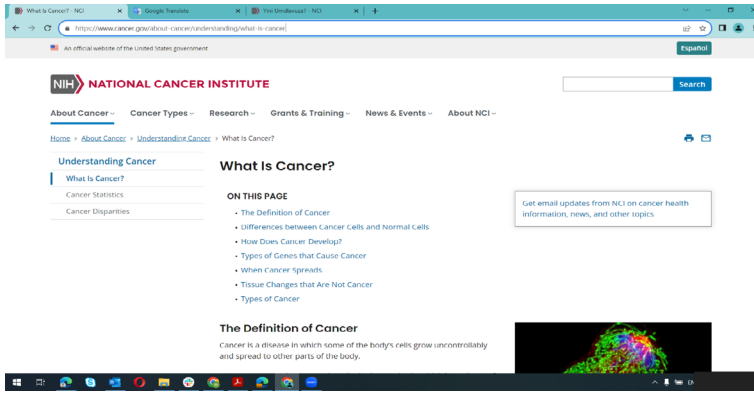

Figure 4a: English cancer awareness article

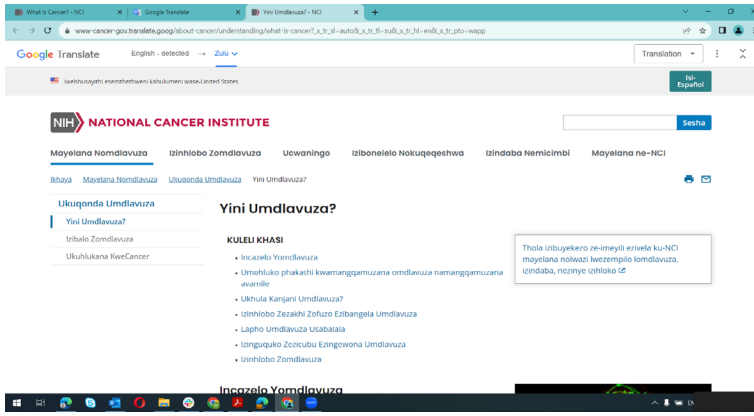

Figure 4b: IsiZulu article translated by Google Translate from English

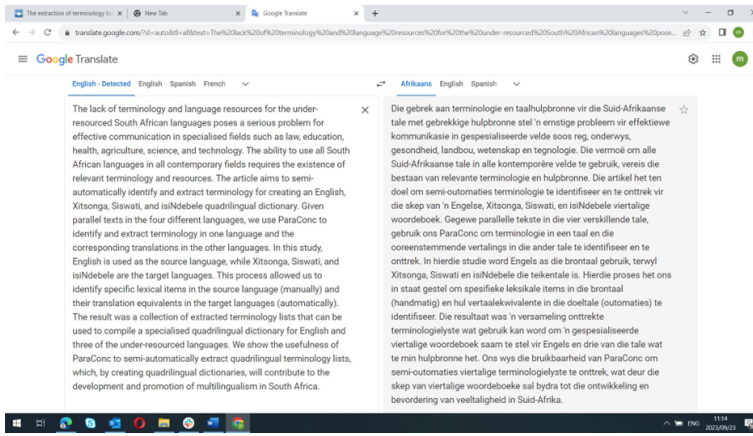

Figure 5a: English abstract translated into Afrikaans by GT

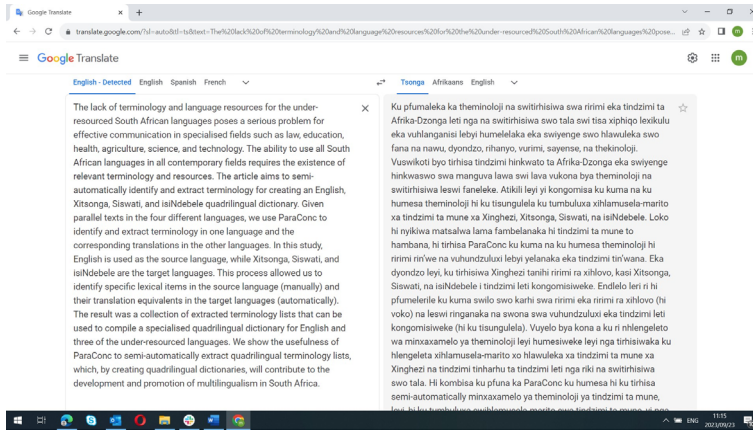

Figure 5b: Abstract written in English translated by GT into Xitsonga

In Figure 4a, the English article about cancer from the National Cancer Institute was translated by GT to produce the isiZulu version in Figure 4b. While in Figure 5a, the English article was translated into Afrikaans and in Figure 5b, it was translated into Xitsonga. This is evidence that using GT helps break language barriers. GT enables users who are only fluent in isiZulu, Afrikaans, and Xitsonga to access the information in their home

language. This would also promote multilingualism in the country in the sense that speakers of other languages can access information presented in any English document in their home languages by using MT tools.

Similarly, the SETM can also be used to promote access to information in SALs. For instance, many government documents in South Africa are written in English as the dominant language. SETM, which was created using governmental corpora, can assist individuals access these documents in their target languages, as presented in Figure 6.

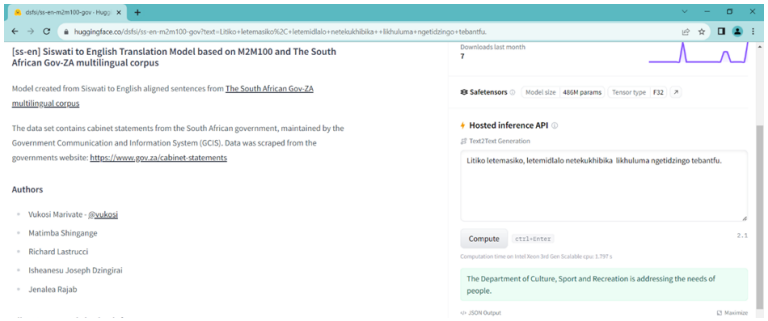

Figure 6: Siswati to English Translation Model output

In Figure 6, the information from a the South African government document has been translated into Siswati. Siswati speakers can use GT to understand what the English sentences mean in their home language. GT helps facilitate access to information in Siswati. Mager et al., (2023) support that MT tools are crucial for overcoming language barriers and enabling people to continue using their native languages. In such instances, GT is not only useful to indigenous language speakers in South Africa and their ability to access information, but also to English speakers who want to know terms in other SALs. This will thus enable them to converse with speakers of other languages. Therefore, MT tools greatly improve access to multilingual materials for underserved languages (Turner et al., 2015).

Language Learning

Language learning involves acquiring knowledge, skills, values, and attitudes to understand, speak, read, and write in other languages. It can be achieved through formal education, immersion in a linguistic environment, self-learning, or a combination of these methods (Hoque 2017). It includes the learning of grammar, spelling, pronunciation, and the capacity to communicate in a variety of contexts. MT systems can be valuable language learning resources with benefits such as enhanced writing, grammar, vocabulary, pronunciation, and spelling (Almusharraf & Bailey, 2023). The tools assist language users by offering translations, enabling them to comprehend and practice the languages they are attempting to learn. For example, IZN can be used for language learning purposes as presented in Figure 7.

Figure 7: isiZulu.net as language learning tool

In Figure 7, the IZN has translated the isiZulu phrase Umfazi wami to 'my woman'. Correct translation equivalents of each word umfazi 'woman' and wami 'my' are provided. Since IZN can be used for more than basic translation purposes, language learners can utilise it to learn the meanings of words and phrases in isiZulu and English. They can also acquire knowledge of word categories and word constructions in both languages. Language learners can also improve their language learning process and access information that their teachers may not be able to provide (Riasati et al., 2012). Therefore, using the IZN can close the language gap between isiZulu and English learning, and it is an effective catalyst for multilingualism.

Moreover, MT can be utilised to translate learning material into a variety of target languages and make language learning easier for students.

Enhanced Creation of Multilingual Content

Multilingual content creation is the process of developing and distributing information, media, or creative materials in different languages. This entails developing interfaces and content that are meant to be accessible and understandable to audiences that speak various languages (McDevitt et al., 2004). Multilingual content creation is a strategic communication approach that acknowledges and embraces linguistic diversity. It allows individuals and organisations to engage with a global audience and promots meaningful interactions across borders and cultures (Stein-Smith, 2021). In a multilingual country like South Africa, MT tools such as GT can significantly facilitate multilingual content creation, as depicted in Figures 8a, 8b, and 8c.

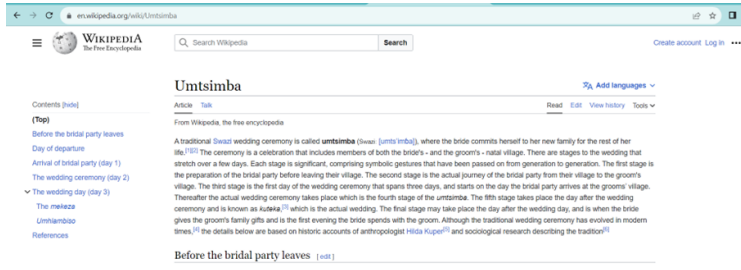

Figure 8a: Article written English in Wikipedia

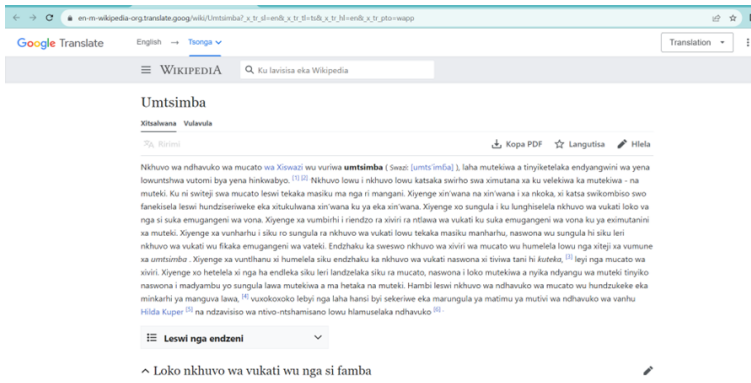

Figure 8b: English article translated into Xitsonga in Wikipedia

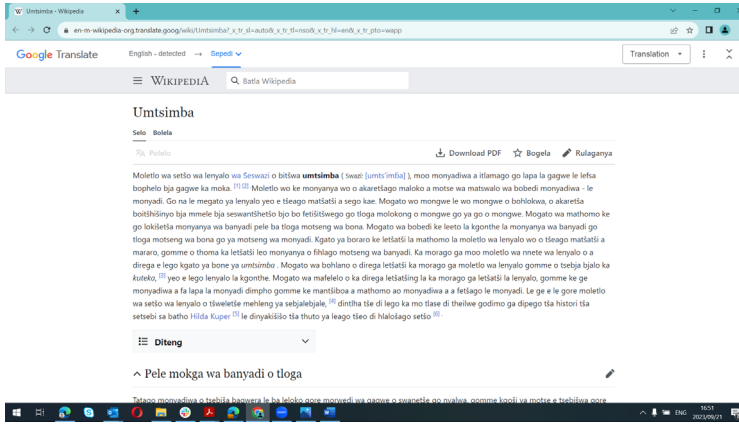

Figure 8c: English article translated into Sepedi in Wikipedia

In Figures 8b and 8c, the English article (8a) has been translated to create content in Xitsonga and Sepedi. GT has provided the means to create multilingual content in these languages. Therefore, MT tools allow content creators and organisations to expand their audience coverage by translating their content into various languages. MT tools also ensure inclusivity and accessibility of information to diverse language communities.

Localisation of websites

Sandrini (2008) described website localisation as the modification of an existing website to make it accessible, usable, and culturally appropriate to its intended audience. This process involves translating content and design into multiple languages to ensure accessibility and relevance to consumers across different regions or countries (Pym & Windle, 2011; Jiménez-Crespo 2013). According to Bowker and Ciro (2018), the World Wide Web and its associated navigation tools have developed into a resource that is currently unmatched as a means of providing users across the globe with access to information. The process of website localisation is vital for reaching a global audience and for enhancing users/speakers of marginalised languages from diverse cultural backgrounds to use websites and access information. In South Africa, MT can be a cost-effective solution for localising websites and enable access to information in various SALs. For example, in Figures 9a, 9b, 9c and 9d, GT has been used to localise the English website of the South African Centre for Digital Language Resources (https://sadilar.org/index.php/en/).

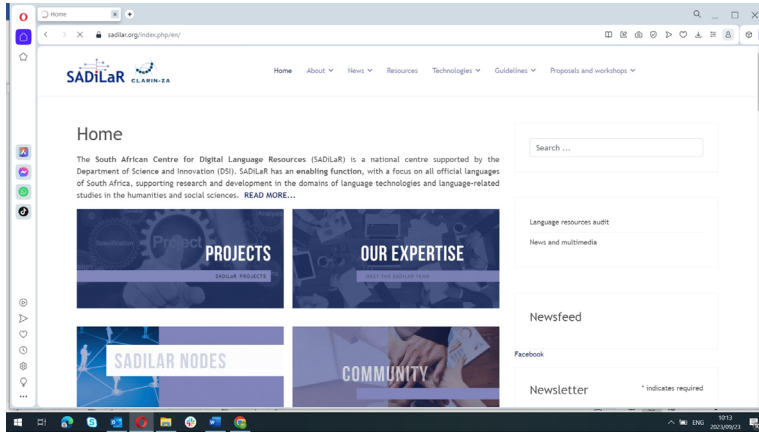

Figure 9a: Original SADiLaR Website in English

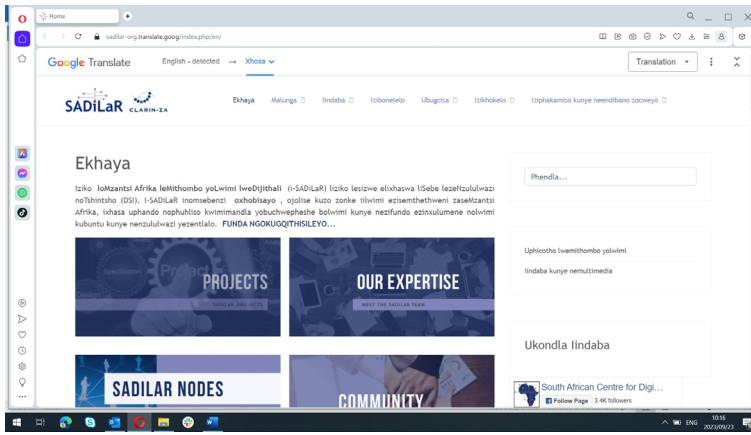

Figure 9b: SADiLaR Localised Website to isiXhosa

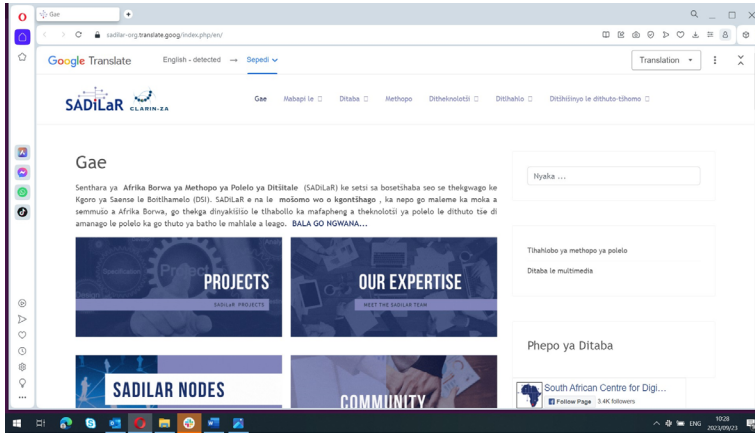

Figure 9c: SADiLaR Localised Website to Sepedi

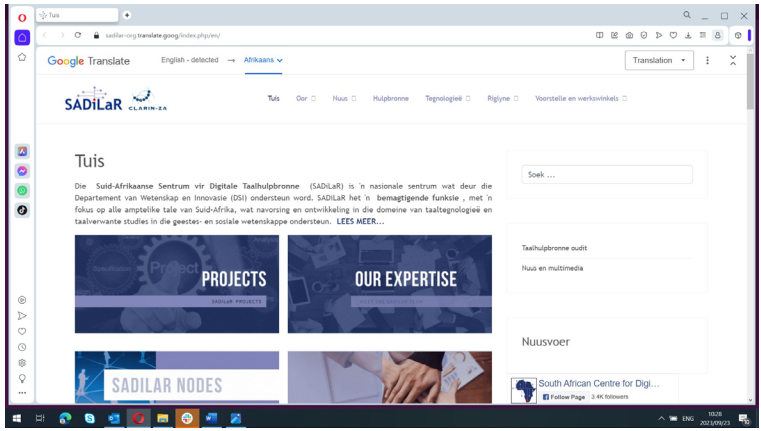

Figure 9d: SADiLaR Localised Website to Afrikaans

The SADiLaR English website in Figure 9a has been localised into isiXhosa in Figure 9b, into Sepedi in Figure 9c, and into Afrikaans in Figure 9d. GT as an MT tool has been aptly used to localise the website. The localisation assists language users to access information in a SAL that they prefer and understand better. Flavián et al., (2006) asserted that the usage of websites in multiple languages addresses different people's linguistic and cultural preferences. Therefore, it is imperative that English websites be localised to other SALs so that people can be able to access them in their home languages as well as in other SALs that they desire to learn. Thus, MT tools are helpful in the promotion of multilingualism and access to information.

Conclusion

This chapter has described three MT tools; Google Translate, isiZulu.net, and Siswati to English Translation Model. It has discussed and demonstrated how these MT tools can be used to promote access to information by breaking down language barriers, encouraging language learning, enhancing the creation of multilingual content, and localising websites. These initiatives will not only facilitate easy access to information but will bolster multilingualism among different speakers of the South African Languages. Therefore, MT provides useful

opportunities that language users must utilise to promote access to information and encourage multilingualism in the diverse linguistic societies of South Africa.

Reference list

Abuarqoub, I.A. 2019. Language Barriers to Effective Communication. *Utopía Y Praxis Latinoamericana,* 24(1), pp.64–77.

Almusharraf, A. and Bailey, D. 2023. Machine Translation in Language Acquisition: A Study on EFL Students' Perceptions and Practices in Saudi Arabia and South Korea. *Journal of Computer Assisted Learning,* pp.1–16. https://doi.org/10.1111/jcal.12857

Andrabi, S.A.B. and Wahid, A. 2021. A Review of Machine Translation for South Asian Low Resource Languages. *Turkish Journal of Computer and Mathematics Education,* 12(5), pp.1134–1147. https://doi.org/10.17762/turcomat.v12i5.1777

Bowker, L. and Ciro, J.B. 2015. Investigating the Usefulness of Machine Translation for Newcomers at the Public Library. *Translation and Interpreting Studies,* 10(2), pp.165–186. https://doi.org/10.1075/tis.10.2.01bow

Bowker, L. and Ciro, J.B. 2018. Localizing Websites using Machine Translation: Exploring Connections between User Experience and Translatability. In: S. Chan (ed.), *The human factor in machine translation.* London: Routledge. pp.7–29. https://doi.org/10.4324/9781315147536-2

Castilho, S. 2016. Measuring Acceptability of Machine Translated Enterprise Content. Unpublished Doctoral dissertation. Dublin: Dublin City University.

Constitution of the Republic of South Africa. 1996. https://www.justice.gov.za/legislation/constitution/

Flavián, C., Guinalíu, M. and Gurrea, R. 2006. The Role Played by Perceived Usability, Satisfaction and Consumer Trust on Website Loyalty. *Information & Management,* 43(1), pp.1–14. https://doi.org/10.1016/j.im.2005.01.002

Grover, A.S., van Huyssteen, G.B. and Pretorius, M.W. 2011. The South African Human Language Technology Audit. *Language Resources & Evaluation,* 45, pp.271–288. https://doi.org/10.1007/s10579-011-9151-2

Hoque, E. 2017. An Introduction to the Second Language Acquisition. In: E. Hoque (ed.), *Language acquisition.* Dhaka, Bangladesh: EDRC. pp.1–23.

Hussey, N. 2012. The Language Barrier: The Overlooked Challenge to Equitable Health Care. *South African Health Review,* (1), pp.189–195.

Hutchins, J. 1995. Machine Translation: A Brief History. In E.F.K. Koerner and R.E. Asher (eds), *Concise History of the Language Sciences: From the Sumerians to the Cognitivists.* Pergamon Press. pp.431–445. https://doi.org/10.1016/B978-0-08-042580-1.50066-0

Siswati to English Translation Model. 2023. [ss-en] Siswati to English Translation Model based on M2M100 and The South African Gov-ZA Multilingual Corpus. Available at: https://huggingface.co/dsfsi/ss-en-m2m100-gov (Accessed on 9 September 2023).

IsiZulu.net. 2011. isiZulu.net: Bilingual Zulu-English Dictionary. Available at: https://isizulu.net/ (Accessed on 29 October 2023).

South African Centre for Digital Language Resources (no date). SADiLaR home. Available at: https://sadilar.org/index.php/en/ (Accessed on 20 September 2023). National Cancer Institute (no date). What is cancer? Available at: https://www.cancer.gov/about-cancer/understanding/what-is-cancer (Accessed on 18 September 2023).

Illidge, M. 2022. Google Translate adds two new South African Languages. Viewed at https://mybroadband.co.za/news/internet/444186-google-translate-adds-two-new-south-african-languages.html [Accessed on 9 September 2023]

Jiménez-Crespo, M.A. 2013. *Translation and Web Localization.* Routledge. https://doi.org/10.4324/9780203520208

Johnson, M., Schuster, M., Le, Q.V., Krikun, M., Wu, Y., Chen, Z., Thorat, N., Viegas, F., Wattenberg, M., Corrado, G., Hughes, M. and Dean, J. 2017. Google's Multilingual Neural Machine Translation System: Enabling Zero-Shot Translation. *Transactions of the Association for Computational Linguistics,* 5, pp.339–351. https://doi.org/10.1162/tacl_a_00065

Kituku, B., Muchemi, L. and Nganga, W. 2016. A Review on Machine Translation Approaches. *Indonesian Journal of Electrical Engineering and Computer Science,* 1(1), pp.182–190. https://doi.org/10.11591/ijeecs.v1.i1.pp182-190

Lin, G.H. and Chien, P.S.C. 2009. Machine Translation for Academic Purposes. *Proceedings of the International Conference of TESOL and Translation, Taiwan, Changhua.* 11 December 2009. pp.133–148.

Liu, W. 2017. The Changing Role of Non-English Papers in Scholarly Communication: Evidence from Web of Science's Three Journal Citation Indexes. *Learned Publishing,* 30, pp.115–123. https://doi.org/10.1002/leap.1089

Lotz, S. and Van Rensburg, A. 2014. Translation Technology Explored: Has a Three-Year Maturation Period Done Google Translate any Good? *Stellenbosch Papers in Linguistics Plus,* 43, pp.235–259. https://doi.org/10.5842/43-0-205

Mager, M., Bhatnagar, R., Neubig, G., Vu, N.T. and Kann, K. 2023. Neural Machine Translation for the Indigenous Languages of the Americas: An Introduction. arXiv preprint arXiv:2306.06804. https://doi.org/10.18653/v1/2023.americasnlp-1.13

McDevitt, K., Pérez-Quiñones, M.A. and Padilla-Falto, O.I. 2004. Design of a Community-Based Translation Center. arXiv preprint cs/0401007.

Mlambo, R., Skosana, N. and Matfunjwa, M. 2021. The Extraction of Terminology List using ParaConc for Creating a Quadrilingual Dictionary. *Southern African Linguistics and Applied Language Studies,* 39(1), pp.82–91. https://doi.org/10.2989/16073614.2021.1896971

Motlani, R. 2016. Developing Language Technology Tools and Resources for a Resource-Poor Language: Sindhi. *Proceedings of the NAACL Student Research Workshop,* San Diego, California. 13–15 June 2016. pp.51–58. https://doi.org/10.18653/v1/N16-2008

Muller, R. 2009. Google now speaks Afrikaans. My Broadband. Available at: http://mybroadband.co.za/news/internet/9395-google-now-speaks-afrikaans.html [Accessed on 12 September 2023]

Nurminen, M. and Koponen, M. 2020. Machine Translation and Fair Access to Information. *Translation Spaces,* 9(1), pp.150–169. https://doi.org/10.1075/ts.00025.nur

Poibeau, T. 2017. *Machine Translation.* MIT Press. https://doi.org/10.7551/mitpress/11043.001.0001

Prinsloo, D.J. 2010. Internet Dictionaries for African Languages. *Lexicographica,* 26, pp.183–194. https://doi.org/10.1515/9783110223231.2.183

Pym, A. and Windle, K. 2011. Website Localization. In: K. Malmkjaer and K. Windle (eds), *The Oxford Handbook of Translation Studies.* Oxford: Oxford University Press. pp.410–424.

Rani, K.U. 2016. Communication Barriers. *Journal of English Language and Literature,* 3(2), pp.74–76.

Riasati, M.J., Allahyar, N. and Tan, K.E. 2012. Technology in Language Education: Benefits and Barriers. *Journal of Education and Practice,* 3(5), pp.25–30.

Sandrini, P. 2008. Localization and Translation. *MuTra Journal,* 2(1), pp.167–191.

Seçkin, C. 2023. Instructors' Perceptions of Students' Google Translate use in Language Learning. Söylem Filoloji Dergisi, Çeviribilim Özel Sayısı, pp.474–482. https://doi.org/10.29110/soylemdergi.1186593

Steigerwald, E., Ramírez-Castañeda, V., Brandt, D.Y., Báldi, A., Shapiro, J.T., Bowker, L. and Tarvin, R.D. 2022. Overcoming language barriers in academia: Machine Translation Tools and a Vision for a Multilingual Future. *BioScience,* 72(10), pp.988–998. https://doi.org/10.1093/biosci/biac062

Stein-Smith, K. 2021. Multilingualism for Global Solutions and a Better World. *Journal of Language Teaching and Research,* 12(5), pp.671–677. https://doi.org/10.17507/jltr.1205.05

Sun, Y.C., Yang, F.Y. and Liu, H.J. 2022. Exploring Google Translate-Friendly Strategies for Optimizing the Quality of Google Translate in Academic Writing Contexts. *SN Social Sciences,* 2(8), pp.1–18. https://doi.org/10.1007/s43545-022-00455-z

Tuominen, T., Koponen, M., Vitikainen, K., Sulubacak, U. and Tiedemann, J. 2023. Exploring the Gaps in Linguistic Accessibility of Media: The Potential of Automated Subtitling as a Solution. *The Journal of Specialised Translation,* 39, pp.77–98. https://doi.org/10.26034/cm.jostrans.2023.067

Turner, A.M., Brownstein, M.K., Cole, K., Karasz, H. and Kirchhoff, K. 2015. Modeling Workflow to Design Machine Translation Applications for Public Health Practice. *Journal of Biomedical Informatics,* 53, pp.136–146. https://doi.org/10.1016/j.jbi.2014.10.005

Uszkoreit, H. 2000. Language Technology a First Overview. German Research Center for Artificial Intelligence, pp.1-4.

Van Lieshout, C. and Cardoso, W. 2022. Google Translate as a Tool for Self-Directed language learning. *Language Learning & Technology,* 26(1), pp.1–19.

Vieira, L.N., O'Hagan, M. and O'Sullivan, C. 2021. Understanding the Societal Impacts of Machine Translation: A Critical Review of the Literature on Medical and Legal Use Cases. *Information, Communication & Society,* 24(11), pp.1515–1532. https://doi.org/10.1080/1369118X.2020.1776370