



Digitising Kiswahili for Translation Economy

Nabeta K.N. Sangili

Kaimosi Friends University

Uwasilishaji wa Makala

Pokelewa: 2024-08-24

Rekebishwa: 2024-10-20

Kubalika: 2024-10-31

Chapishwa: 2024-11-07

Maneno Muhimu

Artificial Intelligence

Digital

Machine

Service providers

Kiswahili

Jinsi ya Kurejelea Makala:

Sangili, K. N. (2024). Digitising Kiswahili for Translation Economy. *Journal of Kiswahili and Other African Languages*, 2(2), 44-51.

Copyright © 2024



Ikisiri

Translation currently is one of the biggest currency earners globally whose net worth stands in billions of US Dollars. Perhaps, when compared to teaching, translation becomes the second richest single entity in a pool of Language Service Providers. Kiswahili language, in its quest to be among the top languages globally has embraced translation and increasingly expanding its horizon. This paper therefore purposed to find out the state and quality of online translation tools like google translate as used by Kiswahili clients in translation services. The research methodology used was qualitative. Random sampling technique was used to get sample words, phrases and sentences. The results show that Google Translate has demonstrated excellent translation results when it comes to individual lexemes as compared to phrases and sentences. In some cases, the phrases, especially when used figuratively, has potential to mislead and give birth to spurious translation. The results of this research will go a long way in helping improve online translation from and into Kiswahili hence not only improving Kiswahili but also opening potentialities of other African languages.

Utangulizi

Kiswahili is an African indigenous language spoken by natives on the coasts of East Africa, Somalia, Kenya, Tanzania, and Mozambique. It is also spoken as a second language by millions across Africa, Asia, the Middle East, and the world. The language is estimated to command approximately 200 million speakers globally (UNESCO, 2023). Kiswahili is also being taught as a second language in various learning institutions at all levels, from elementary through middle schools to colleges and universities globally. The language has found a home in more than 100 institutions in the USA alone. According to Piedmont Global Language Solutions, a US-based language services provider, approximately 90,000 people speak Swahili in the United States. This number may be higher as many people enrol in the language daily, both in physical classes and online platforms. In mainstream media, Kiswahili is now an integral part of programming in many media companies around the world, including the British Broadcasting Corporation (BBC), Voice of America (VOA), Radio France International (RFI), Radio China, Radio Tehran, and others. Since the 1950s, the language has had a dedicated Station at the UN Media Center. On Social platforms, all search engines are washed with Kiswahili data and metadata. One can now easily find the information needed in Kiswahili. In November 2021, at its 41st Session in Paris, France, the General Conference of UNESCO declared July 7th World Kiswahili Language Day. By Resolution 41 C/61, Member States recognised the critical role played by Kiswahili in promoting cultural diversity, creating awareness, and fostering dialogue among civilisations in the modern era. 7th was chosen because, on 7th July 1954, the founding President of Tanzania, Mwalimu Julius Kambarage Nyerere, through the Tanganyika African Nation Union (TANU), declared Kiswahili as an essential tool in fighting for emancipation from the British Colonial rule (Okombo & Muna, 2017).



Digitizing Kiswahili: An Enhanced Strategy for quick development

What does it mean when we say digitise? Or, simply put, what is language digitalisation? According to Lukamika (2022), it is the conversion of written and printed records into electronic form. The content may be in the form of text, image, audio, or a combination of these, which is known as multimedia. This means that language digitalisation is converting written and printed records into electronic forms in a specific language. It also means rendering a language digitally and increasing its digital footprint for any user anywhere and anytime. According to the Global Language Digitization Initiative, about 1000 native languages out of 7000 are supported online. The main reason is that the 6000 languages are not digitalised. As the world moves towards 100% online connectivity, it also carries over 3 billion people who want to use their languages for various transactions. An online user will want to partially or fully trade, speak, interact, study, teach, and entertain using a language they like and understand most. Increasing the footprint of languages will positively impact languages, vendors, and users.

Due to its widespread use, the Kiswahili language has demonstrated to the world that it can wrestle enough terabytes of online space through various global projects. Some of these projects have been there for a long time, perhaps since the late 1990s and early 2000s. Some projects that commenced and ended successfully include the Microsoft Kiswahili Localization Project. The localisation project fundamentally involved translating Microsoft Office and Windows into the Kiswahili language. The Microsoft Kiswahili localisation project was divided into two major phases – phase I and II (Kamau, 2008). The second project was the Linux Kiswahili Localization Project. Just like Microsoft, the project ended successfully, and it aimed to produce a Linux program that could be installed in Kiswahili. The third project was the Kiswahili Text to Speech (TTS) Program. This program was designed to enable the computer to read aloud Kiswahili texts. The fourth project was the Kiswahili Version of the Google Search Engine (Kiswahili Google). This project is intended to allow the end user to access information using the Kiswahili language. All major tech companies and software producers have enabled access to their products and services in Kiswahili. The undertakings by IT experts, linguists, and Kiswahili lexicographers to ensure Kiswahili makes strides in technology have immensely contributed to its digitisation.

Translation as a field outside East Africa has made great strides over the centuries since its inception. However, due to delayed written forms of communication, East Africa began experiencing serious translation during and after colonialism, first as a practice and second as an academic field. Since then, translation has contributed significantly to the Kiswahili language, literature and linguistics. Much of Kiswahili literature, linguistic aspects and terminologies, and legal, business and economic jargon have begun developing rapidly due to enhanced efforts in translation after the introduction of translation studies. Through translation, Kiswahili can comfortably access novels, plays, and poems from European, Caribbean, and Asian literature. Besides literature, currently, Kiswahili has English-Kiswahili dictionaries and science and technology. However, it should be noted that this progress has depended on human translation. Computer-aided translation in Kiswahili has not been covered enough as a practice in translation and as an academic endeavour.

Many universities in Kenya and Tanzania teach translation studies at the undergraduate level. At the postgraduate level, in 2010, the University of Nairobi collaborated with the United Nations to establish a Center for Translation and Interpretation (CTI), which allows one to advance Translation or Interpretation at the Postgraduate Level. Moi University also offers a Master of Arts in Translation, allowing one to build on the Bachelor's knowledge. These will make Translation and Interpretation in Kenya and East Africa more professional and Scholarly in the coming years. Low research output on Computer Aided Translation (CAT) and, generally, machine translation can be excused now, but not later, as scholars are taking keen interest in this area. CAT Tools are also gaining eminence among



Translators within the region. However, it should be noted that currently, there is a global upsurge in Kiswahili Language Service provision, including Translation and Interpretation, brought about by the rapid globalisation of the Kiswahili language. Massive investment projects in the region by China, Japan, the USA, the United Kingdom, and other nations have necessitated Translation Interpretation and teaching of Kiswahili as a second language. This explains why Kiswahili is rapidly spreading in China and around the globe.

Abridged History of Machine Translation

The first notion of 'mechanical translation' appeared approximately four hundred years ago. In 1629, Rene Descartes had an idea of representing language by codes. This meant trying to find one code in one language (Source Language) and consequently see the equivalent in another language (Target Language) with the same meaning. Only after the Second World War (1939-1945) did experts begin exploring the possibility of using stored programs in computers (Deleveney, 1960). This was a huge development at that time (Quah, 2006). Just before the Second World War, two researchers, George Artsrouni and Petr Smirnov-Troyanskii, developed an idea that suggested automatic translation must have three stages: an editor who analyses and converts the words from SL; a machine which finds lexical equivalents from SL to TL; and a second editor who was to be familiar with TL and weeded out errors in the translation. In 1954, the University of Georgetown and IBM developed a project on this matter. The first machine translation versions were based on bilingual dictionaries with even specific word order. There was a need for synthetic and semantic functions, so the developers tried to improve these kinds of programs. Unfortunately, the researchers began to be pessimistic because of semantic barriers, which became an inseparable obstacle of MT (Quah, 2006). IBM Corporation and the University of Washington introduced Mark II, an unsuccessful operating system because it did not fully perform semantic and syntactic functions. In 1964, the US government became interested in these incapable translation programs and formed ALPAC (Automatic Language Processing Advisory Committee) to evaluate these systems in 1966.

Drugan (2013), citing Boucau (2006), observes that since the 1990s, translation as a commercial activity has become a global business whose growth outstripped world trade. The rapid expansion of the internet became a significant factor in allowing even smaller companies to market and sell their products internationally. In addition, the demand from consumers for product information, software, user manuals, games, educational materials and so on in their language fueled, in turn, the demand for translation. With the turn of the 21st century and the new millennium, translation services have become part of technology. Currently, technology has taken over translation and language services.

Machine Translation in the 21st Century

The 21st century is arguably the era of high-tech technology. Technological developments have been quick, powerful, and extremely useful and cover the entirety of human professions. With the increased presence of telecommunication companies and the introduction of Smartphones, (mini) iPads, tablets and the subsequent rollout of 3G, 4G and 5G networks, internet penetration has increased the uptake and consumption of end-user translation programs (Oyelaran-Oyeyinka and Adeya, 2002 & Internet Society, 2014). All one needs is a smartphone operating on an IOS or Android system-and everything else is downloadable with a tap on the screen. These apps make translating on the go extremely fast and easy and will help bridge any communication gap; advantages and disadvantages for these apps should abound.

MT has been argued to be advantageous to humanity: increased speed of translation; translator no longer needs to remember how she/he translated a word/phrase previously; reduced costs, among other notable benefits. But it has disadvantages too: One, it has slashed their profits; two, difficulties in coping with phenomena which require linguistic knowledge, like morphology, syntactic functions,



and word order; and three, they lose adequacy due to missing or spurious translations. Among the various existing Machine Translation approaches, Statistical Machine Translation is currently considered a significant breakthrough in the modern age of translation and machines. Statistical Machine Translator (SMT) is an application of machine-based translation that collects knowledge from statistics from previous experience. In other words, it is an application that collects data on pre-translated texts and uses the same to generate a translation. SMT is based on the concept of probability. The translation is chosen from the highest probability. The probability score is obtained from previous data from training the SMT with human-translated documents. The probability score is obtained from a mathematical model, including the language and translation models. The source language text is pre-processed before applying the language and global search models and preprocessed again for the final presentation in the target language (Syahrina, 2011).

Computer and mobile translation applications

An application is software that allows one to perform a specific task. In IT, an application uses a technology, system, or product. The term application or app is a shorter form of application program. They are designed to perform a specific function directly for the user- what is instead known as the end-user. An end-user refers to anybody who will use the program after it has been offloaded into the applications market and vendored by the likes of Google Play Store and Apple. One can download apps like Opera Mini, Opera Max, Facebook, hand Scanner, My Market, hangouts, play Newsstand, WhatsApp, Kingsoft Office and translation applications for mobile devices like Smartphones and tablets, which opens up many new possibilities for modern communicators. Technologically developed languages like Indo-European and Asiatic families, having interacted with technologies for a long time, can render almost perfect translations using these mobile applications. Apart from Kiswahili, IsiZulu, Yoruba and Hausa, these applications cannot functionally translate African languages. Though these four African languages have been mentioned to be technology-enabled, it is worth noting that they do not match the sharpness and accuracy of European and Asian languages in translation. They are in what we can essentially refer to as the developing stage, something that can take up to a century or more to be almost fully developed. However, the pace at which African languages are interacting with technology, the keenness of translators to technologies, the sprouting of the young generation interested in emerging technologies and the development of computational and Corpus Linguistics may just catalyse some Cross-border and Vehicular languages to be at par with technologically enabled global languages like English, French, Chinese, Spanish, Japanese and German.

Translation applications for mobile phones include Google Translate, iTranslate, Talking Translator, and translated4.eu; English Swahili translator with voice; Jibbig; mantaphrase; wayGo; iHandy; Voice translate; Bing translate; voice translate pro; tourist language learn and speak. Kiswahili language features in many of these programs, for instance in Google Translate, iTranslate, Talking Translator, Lonely Plane offline translator, English Kiswahili translator with voice, Navita translator, Jibbig, among others, but is not one of the component languages in trippo Mondo voice translator; sayHi Translate and Mantaphrase.

Results and discussion: Google Translate and Kiswahili

This is probably the biggest application for accessibility and usage by being a product of the largest internet company. Google Translate enjoys global coverage. It offers a free multilingual statistical MT service. It has text, speech, and image translation tools to help end-users translate phrases, words and real-time video from SL to TL. Currently, it supports over 90 languages, and according to Google, more than 500 million people use Google Translate every month, making at least one billion daily translations. A user can enter a text by on-screen recognition or handwriting recognition. Once a word, phrase or sentence is keyed in, the app statistically searches and displays the closest translation.



Google Translate does not apply grammatical rules since its algorithms are based on statistical rather than traditional rule-based analysis. The program uses English as the intermediating language in that it can have a formula like this one:

Kiswahili (L) English French (L2)
 Unless English is the target or the source language.

Because of English intermediation, the translation as a product has been faulty in some instances. For instance, Google Translate sometimes gets the pronunciation wrong when using a voice transaction tool. Words like “handsome”, “zigzag”, and “hunters”, when uttered, even at close range, do not produce any translation. Instead, it regurgitates the same word, but the voice tries to Swahilize the pronunciation (which is quite embarrassing for anybody who knows how Kiswahili phonology). This observation is made for all the languages tested by other researchers, including Rawlinson. Sometimes, the voice tool confuses the sounds of words like “farming and famine” and does not give the correct translation. When typed with single lexemes like “flow”, the app gets the probability wrong by giving us an equivalent “kati yake.” A Second attempt of the flow as polepole and flowing as “inapita” and “ni inapita”. This indicates that some single lexemes' translation probability is inferior and spurious. What the tool does is combine the translation of “it is “, which is rightly translated as “ni”, and the search for the equivalent of ‘flowing, which is translated as “inapita”. The table below shows the accuracy estimations on individual lexes:

Table 1: Individual Lexemes Translation

Part of Speech	English (SL)	Kiswahili (TL)	Accuracy Estimation
Noun	Man, girl, engineer, doctor, lawyer, jury, commissioner, England, France, horse	Mtu, msichana, mhandisi, daktari, mwanasheria, jury , kamishna, England , Ufaransa, farasi	Very Good
Adverb	Nearly, loudly, very, really,	Karibu, kwa sauti, sana, kweli	Excellent
Verb	Run, jump, stop, explore, happen, be, shrink, widen, evolve, seem, have	Kukimbia, kuruka, acha, chunguza, kutokea, kuwa, kunya , panua, toa , kuonekana, kuwa na	Very Good
Adjective	Exciting, green, tidy, beautiful, handsome, technical, Italian, sweet,	Ya kufurahisha, kijani, safi, mrembo, mzuri, kiufundi, Italia, tamu	Very Good
Pronoun	that, something, I, me, mine, you, we, hers, they, them, yours, myself, himself , themselves,	Hiyo, kitu, mimi, mimi, yangu, wewe, sisi, wao, wao, mimi mwenyewe, mwenyewe , wenyewe	Very Good
Preposition	After, in, to, on, with, under, between, over, at, by, without, because of, next to, on top	Baada ya, ndani, kwa, on, na, chini, kati, juu, katika, na, bila, kwa sababu ya, ijayo, juu.	Excellent
Conjunction	And, because, but, for, if, or, when, until,	Na, kwa sababu, lakini, kwa, kama, au, lini, mpaka	Excellent
Exclamation	Excellent!	Bora!	Excellent

From the foregoing, it is clear that Google’s translation app is between very good and excellent when rendering translations from English to Kiswahili. However, a few issues can be noted and generalised: a translator should be keen on specialised terminologies and, second, on names of nations and nationalities. It should be clear that the above translations table did not consider context, and therefore, when translating in context, one should be careful when using the application. For instance, the Swahili word “mtu” for Man can be misleading if we are talking about gender just as handsome for “mzuri”. Google’s translation application omits gender, thereby making the product spurious. Relying on the application without carefully reviewing or editing the translation should not be encouraged.



Apart from translating single lexemes using mobile applications like Google, phrases and sentences even pause, which creates a bigger accuracy problem. Their levels of accuracy dwindle with the complexities of contexts and the effect a phrase or sentence may carry. The table below demonstrates Google translation accuracy:

Table 2: Phrasal/Sentence Translation Accuracy

Sn.	English (SL)	Kiswahili (TL)	Accuracy Estimation	Correct Translation
1	The English teacher	Mwalimu wa Kiingereza	Excellent	Mwalimu wa Kiingereza
2	They don't like joyriders	Hazipendi furaha	Poor	Hawapendi wadoeaji
3	Drug addicts	Walemavu wa dawa za kulevya	Poor	Waraibu wa dawa za kulevya
4	Doctors operating patients	Madaktari wanaofanya kazi wagonjwa	Poor	Madakari wapasuao wagonjwa
5	Tell the students	Waambie wanafunzi	Excellent	Waambie wanafunzi
6	Man who came yesterday is English	Mtu aliyekuja jana ni Kiingereza	Poor	Mwanamume aliyekuja jana ni Mwingereza
7	Iran nuclear reactants	Iran athari za nyuklia	Poor	Vinu vya Kinyuklia vya Iran
8	Coronavirus has shut the economies of the nations	Virusi vya corona imefunga uchumi wa mataifa	Good	Kirusi cha Korona kimefunga uchumi wa mataifa
9	Economic meltdown	Meltdown kiuchumi	Poor	Mwanguko wa kiuchumi
10	Postpartum patients	Wagonjwa wa baada ya kujifungua	Good	wagonjwa wa(naougua) baada ya kujifungua
11	Sanders bowed out of the race	Sanders akainama kutoka kwa mbio	Poor	Sanders alijiiondoa katika kinyang'anyiro

The table above rates phrasal/sentence translation as mostly poor with only two examples coming out perfectly. From it, we can extrapolate several issues and build on them:

- i. Translations whose meanings have totally changed/mistranslation
- ii. Half translations

Translations whose meanings have totally changed include the following statements: they don't like joyriders translated as 'hazipendi furaha' (they (inanimate) do not like joy). The meaning equivalence is lost in these two statements as one indicates that some people (they) don't like joyriders while the other induces meaning hinging on inanimate objects that don't like joy. The meaning shifts on two levels: one, the object in the ST and TT have shifted, and two, the grammatical meanings of the ST and TT have changed. Statement 3 also changes in that whereas ST refers to people who are extremely attached to drugs, TT refers to disabled people (who, it seems, possess prohibitive drugs). The translation, therefore, equates addiction to disability, which is not the case. On statement four, ST's meaning is clear, but the TT loses the concept of operation (surgery). Statement 6 mentions a nationality (English), but the translation miscues the nationality as language. Google Translate cannot distinguish the actual intention of using the word English. Statement 7 completely gets it wrong as the ST refers to a nuclear facility, but the TT renders into Swahili as nuclear effects. In statement 11, Sanders bowed out of race is spuriously translated. Google Translate literally translates a bow as an act of bending forward the head or part of the body instead of disqualifying himself from the race. The literal translation, therefore, makes the translation erroneous. On half of the translation, only



statement 9 was partially translated. GT could not translate meltdown into Kiswahili. This may mean that no parallel corpus exists in Kiswahili for the word meltdown.

Now, the question begs, why is Kiswahili good at translating individual lexemes but poor in phrases and sentences? Is this inherent in technology, especially technology with a short history of interacting with some languages like Kiswahili? Google Translate technology is based on a statistical model. It has two statistical probability models: language model and translation model and massive parallel corpora of STs and TTs. Once a language has developed enough parallel corpora, it can begin to render better or almost perfect translations using machines. According to Dogru, Matin-Mor and Aguilar-Amat (2017), in corpus-based approaches to machine translation, the more specific the training corpus domain, the better the translation output will be. This means that for languages like English-French, French-German and other European pair languages with better output in Google Translate, the specific domain training and input is far better than Kiswahili-English or English-Zulu. The European languages have had more extended interaction with technology than African languages, hence why African languages are still performing poorly in phrase and sentence translations. Some researchers who have done comparative translation using these applications agree that there is a level of improved perfection when the source and target language are English, French, Spanish, and a few other European languages used by the EU as compared to the different languages (Aiken & Balan, 2011). In recent research, Aiken (2019) notes that an improvement has begun to be noticed among some languages, but unfortunately, Kiswahili is decreasing in quality Google Translate translation.

So, what is the way forward for the Kiswahili language? There is a need for Kiswahili translators to increase the online upload of translations so that when Google Translate searches for the parallel corpora, it attains the highest probability. Only recently, the English-Kiswahili bilingual dictionary was uploaded online in soft copy by the Tanzanian Institute of Kiswahili (TATAKI), which has enhanced the availability of an accurate single lexeme online and is accessible by Google Translate. This explains why, in our first table, the accuracy of single lexeme translation is much higher than that of phrases or sentences in Table 2. According to Munteanu and Marcu (2006), an excellent way to alleviate this lack of parallel data is to exploit a much more available and diverse resource: comparable non-parallel corpora. Comparable corpora are texts that, while not parallel in the strict sense, are somewhat related and convey overlapping information. Examples are the multilingual news feeds produced by news agencies such as Agence France Presse, Xinhua News, Reuters, CNN, and BBC. As a global language, Kiswahili has already gained prominence, establishing subsidiaries for BBC and RFI, among other international media outlets. Localisation of websites from various languages into Kiswahili also enhances phrase and sentence translations, but it is not enough. However, as Kiswahili grows across the world and alongside it, language services like translation and interpretation, we are likely to see more accurate translations on GT.

Conclusion

Machine translation has opened up more languages to make their footprints visible for academic discourses and criticism. Like other African and Asian languages, Kiswahili is progressing positively in online translation technologies. However, the lack of parallel corpora has hampered the correct translation from and into Kiswahili. Besides the lack of parallel corpora, using statistical probability to choose the closest correct translation has also led to rampant mistranslation. Even though Kiswahili has been given functional space within these apps, much will have to be done to bring it to the level of English, French, Spanish and Italian, the leading languages whose levels of accuracy are far above 70%. This will call for human input from various translators. Localisation of websites and translation of news from domesticated international agencies like BBC, Radio France International (RFI) and other sources will go a long way in alleviating the mistranslations on GT.



References

- Aiken, N. (2019). An Updated Evaluation of Google Translate Accuracy. *Studies in Linguistics and Literature*, 253-260.
- Aiken, M., & Balan, S. (2011). An analysis of Google Translate accuracy. *Translation Journal*, 16(2), <https://translationjournal.net/journal/56google.htm>
- Cheragui, A. M. (2012). Theoretical Overview of Machine Translation. Proceedings ICWIT 2012: pp 160-169. <http://ceur-ws.org/Vol-867/Paper17.pdf> Accessed on 20/4/2020.
- Delavenay, E. (1960). *An Introduction to Machine Translation*. The Camelot Press Ltd.
- Doğru, G., Martín-Mor A., & Aguilar-Amat, A. (2017). Parallel Corpora Preparation for Machine Translation of Low-Resource Languages: Turkish to English Cardiology Corpora https://www.researchgate.net/publication/337331602_Parallel_Corpora_Preparation_for_Machine_Translation_of_Low-Resource_Languages_Turkish_to_English_Cardiology_Corpora.
- Drugan, J. (2013). *Quality in Professional Translation: Assessment and Improvement*. Bloomsbury.
- Internet Society (2014). Internet Society Global report 2014. Geneva. www.internetsociety.org
- Kamau, S. N. (2008). A digital divide: Kiswahili holds the key. *Journal of Language Technology & Entrepreneurship in Africa*, 1(1).
- Lukamika, K. M. (2022). Towards Digitisation of Kiswahili Programmes: Problems and Prospects for the University Libraries in Kenya - A Case Study of Tom Mboya University College. *East African Journal of Information Technology*, 5(1), 76-83.
- Munteanu, D.S., & Marcu, D. (2006). Improving Machine Translation Performance by Exploiting Non-Parallel Corpora. *Computational Linguistics*, 31(4): 477-504.
- Okombo, P.L., Muna, E. (2017). The International Status of Kiswahili: The Parameters of Braj Kachru's Model of World Englishes. *Africology: The Journal of Pan African Studies*, 10(7).
- Oyeleran-Oyeyinka, B., & Adeya, C. (2002). Internet access in Africa: Empirical evidence from Kenya and Nigeria. https://www.researchgate.net/publication/221987324_Internet_Access_in_Africa_Empirical_Evidence_From_Kenya_and_Nigeria.
- Quah, C.K. (2006). *Translation and Technology*. Palgrave Macmillan.
- Syahrina, A. (2011). Online Machine Translator System and Result Comparison: Statistical Machine Translation vs Hybrid Machine Translation. Bachelor's Thesis. University of Boras.
- UNESCO (2023). *World Kiswahili Language Day 2023: Unleashing Kiswahili's Potential in the Digital Era*. UNESCO.