UmobiTalk: Ubiquitous Mobile Speech Based Translator for Sesotho Language

John Nyetanyane and Muthoni Masinde(✉)

Unit for Research on Informatics for Drought, Central University of Technology,
Free State, Bloemfontein, South Africa
jnyetanyane@cut.ac.za, muthonimasinde@yahoo.com

Abstract. The need to conserve the under-resourced languages is becoming more urgent as some of them are becoming extinct; natural language processing can be used to redress this. Currently, most initiatives around language processing technologies are focusing on western languages such as English and French, yet resources for such languages are already available. Sesotho language is one of the under-resourced Bantu languages; it is mostly spoken in Free State province of South Africa and in Lesotho. Like other parts of South Africa, Free State has experienced a high number of non-Sesotho speaking migrants from neighboring provinces and countries. Such people are faced with serious language barrier problems especially in the informal settlements where everyone tends to speak only Sesotho. As a solution to this, we developed a parallel corpus that has English as a source and Sesotho as a target language and packaged it in UmobiTalk - Ubiquitous mobile speech based learning translator. UmobiTalk is a mobile-based tool for learning Sesotho for English speakers. The development of this tool was based on the combination of automatic speech recognition, machine translation and speech synthesis. This application will be used as an analysis tool for testing accuracy and speed of the corpus. We present the development, testing and evaluation of UmobiTalk in this paper.

Keywords: UmobiTalk · Automatic speech recognition (ASR)
Machine translation (MT) · Text to speech (TTS) and parallel corpora

1 Introduction

Under-resourced languages are languages that lack unique writing systems or stable orthography, limited presence on the web and lack of electronic resources [6]. These languages are difficult to computerize through the use of natural language processing because large amount of data is required to train the current recognizers [21]. These kinds of languages are becoming unpopular; less economically viable and doomed to lose currency since the attention placed on them is of limited acknowledgement [3].

The choice of Sesotho language was reached based on the fact that it is one of the under resourced languages mostly spoken in the Free State province of South Africa; it is spoken by approximately 4 million South Africans as a home language [31]. In 2011, the province experienced approximately 35 000 migrants coming from outside South Africa with in-migration of approximately 128 000 of population from different

R. Zitouni and M. Agueh (Eds.): AFRICATEK 2018, LNICST 260, pp. 93–106, 2019.
https://doi.org/10.1007/978-3-030-05198-3_9

provinces [31]. One of the reasons South Africa experiences higher migration rates is because it is one of Africa's economic giants, thus acting as a catalyst to motivate immigrants from the neighbouring countries, especially those that are facing socio-economic challenges [30].

The tool described in this paper was motivated by the problems brought by the language barrier that exists between Sesotho speakers and non-Sesotho speakers in Free State. Despite the advancement and proliferation of speech-to-speech technologies and tools, there exist no mobile phone based tool (known to the authors) that can effectively aid learning of the Southern Sotho language. Such a tool is very useful to foreigners, non-Sesotho speakers, migrants as well as people with special needs who are usually faced with a unique challenge on how to integrate themselves to Southern Sotho language speaking society in the province. Although English is seen as an intermediary language that bridges language barriers between different races, people especially in the rural areas of Free State do not know (speak, read and write) English, hence the Sesotho translating tool is seen as an asset. The technological aspects included in UmobiTalk and discussed in this paper are: speech based technology (this caters for both the source and target language), natural language understanding (NLU) modules (such as morphological, syntactic and semantic analyzers), language corpus and machine translation (MT).

We present the details of the design, implementation, testing and evaluation of UmobiTalk. The rest of the paper is structured as follows: Sect. 2 details the underlying theory on which the UmobiTalk is based, Sect. 3 is implementation while the methodology used is in Sect. 4. The details of the tool's evaluation, conclusion and results are in Sect. 5.

2 Literature Review

A speech-to-speech application must have the following components: Automatic Speech Recognition (ASR), Machine Translation (MT) and Text-to-Speech (TTS) [11, 20]. As shown in Fig. 1 below, ASR receives an input (source language), MT converts (processing) a source language to target language and TTS speak (output) the target language.

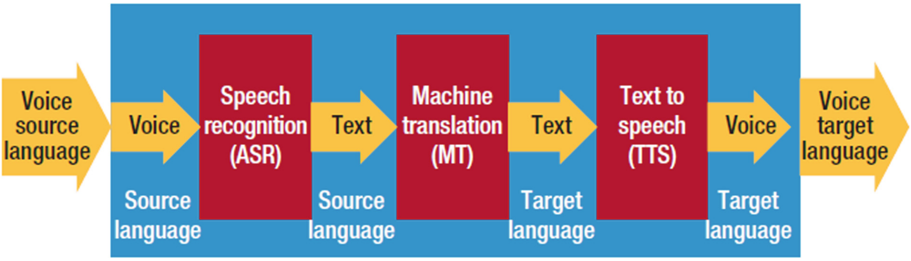


Fig. 1. A speech-to-speech application's events sequence [11, 20].

ASR technology makes life easier because spoken words can be used to communicate with the machine. According to [24], speech is the easiest way to communicate and is faster than typing and more expressive than clicking. Still on the same topic, recent research of [29] reveals that speech to text applications can also improve system accessibility by providing data entry options for blind, dyslexic, deaf or physically challenged users. The most recent example here is ‘Be My Eyes’; iPhone app that lets blind people contact a network of sighted volunteers for help with live video chat (http://www.bemyeyes.org/). Research into speech processing and communication for the most part, was motivated by people’s desire to build a mechanical model to emulate human verbal communication capabilities [4]. Speech is the most natural form of human communication; ASR has made it possible for computer to follow human voice commands and also understand the human languages.

Evidence from literature demonstrates the fact that, unlike entering input on a fully-sized keyboard, entering text on a mobile device is often sluggish and error prone [1]. Additionally, using a touch screen on small mobile device to input data is time consuming and frustrating. The use of ASR in mobile devices is more effective and flexible than in desktops because they can be used while a person is “on the move” [24]. Researchers have suggested that ASR is very important especially for users with low literacy or little script knowledge such as those in the developing regions.

2.1 Machine Translation

Machine translation (MT) technology is a process of substituting a source text with a target text, but because of language implications, a well-constructed parallel corpus is essential to handle text and phrase translation processes [24]. The viability of machine translator has been experimented by several researchers. According to [27], MT has been seen as an asset that can be used by learners to learn foreign languages. MT translator has been used by most applications such as UmobiTalk, Google translate, dictionary applications, and many more. Earliest research of [19] expounds the idea of using machines to translate started at 1940s and it was seen as an indispensable technology on the basis that it is economically viable compare to human translators, but on the other hand it was posing a threat to professional human translator as it was taking all over. Although the machine translators are quick, inexpensive, always available, and language independent that correlates highly with human translators they will always have flaws compared to professional human translators [15].

To obtain an outmost performance and quality of machine translation, collaboration of machines and human performing post editing of language translation is fundamental [15]. The measuring of the machine translator is determined by its closeness to a professional human translator. According to [10], as they were using the MT to translate from Arabic to English they concluded an MT as an analysis tool to evaluate the parallel corpus’ correctness, robustness, reliability, accuracy, flexibility and many more.

2.2 Language Corpus

In order to build any speech engine, whether speech recognition or speech synthesis engine, one needs a corpus. Language corpus is a collection of pieces of language text in electronic form selected according to external criteria to represent as far as possible a language as a source of data for linguistic research [22, 23]. Corpus build raised rapidly from 1960 to 1980 and corpus usage is not only becoming an important foundation of modern linguistic studies, but as other specialized academic research in the field of medicine, architecture, technology, law, English and other premises [35]. Earliest research of [2] explain that corpus usage has extended to areas such as translation studies, stylistics, and grammar and dictionary developments. In addition, [14] expounded a corpus as a tool that can be used for many diverse language technology applications such as word sense disambiguation, anaphora resolution, information extraction, statistical machine translation, grammar projection, unsupervised part of speech tagging or learning multilingual semantic translation. There are different types of corpora (collection of corpus) which are general, specialized, parallel, historical, multimodal, and learner corpus [13, 36].

General corpus can be spoken (speech corpora) or written corpora aim to provide knowledge for the whole language and is considered as a very large monolingual corpus with millions of words that will be used to match the input [13]. General corpora also known as sample corpora or reference corpora can be used to capture the language variety such as Britain English and American English or Lesotho's Sesotho and South African Sesotho [36]. It is reference corpora in such a way it can be used as a snapshot of a language collected at a particular point in time [36].

Specialized corpus is restricted to a certain domain and is compiled for a specific purpose and represents a particular context, genre, text or discourse and subject matter or topic [13]. [35] define specialized corpora as collecting a particular field of corpora to build ideal library collection. On the other side, [36] explain a specialized corpora as a tool that captures the specific type of a language use and dwell on it by using highly contextualized terminologies. Learner corpus is another type of a specialized corpus focusing on some basic aspects of a language and is used specifically by non-native speakers of a language represented to facilitate the teaching and learning processes and material.

Parallel corpus is a widely used multilingual translation containing two or more language text samples aligned at sentence level in which one language represents the source and another one represents the target [28, 36]. This technology is one of the indispensable resources that emerges wide range of multilingual applications such as machine translation and cross lingual information extraction [28]. Additionally, parallel corpus can be explained as a valuable resource for cross-language information retrieval and data-driven natural language processing systems especially for statistical machine translation (SMT) [26, 34]. The translation flow can either be unidirectional (one way direction from source to target) or bidirectional (from source to target and vice-versa) [32].

Historical corpus known as diachronic corpus is a corpus that can unhide or track the language and language writing used from centuries (ancient orthogonal) that can be compared with the currently used language with the aim to obtain rational finale on how language evolves [36]. Example on Sesotho orthography which can be explained as a way of writing or text spelling. The convention of South African Sesotho orthography used in ancient times such as old testimonial bibles (20th centuries ago) is quite different from the one used lately [9]. Lesotho still retains an ancient original orthography while the one used in South Africa has evolved [9].

According to [18], the development of historical corpus in Arabia assists linguistic and Arabic language learners to effectively explore, understand and discover interesting knowledge hidden in millions of instances of language use. A historical corpus of electronic art music has been successfully developed and is now available online from UbuWeb art resource site [8]. Despite its flaws in terms of bias whereby male composers are dominant, it provides an interesting test ground for automated electronic music analysis in terms of historical and cultural coverage [8].

Multimodal corpus is a corpus that is done through audio and video recording normally during the discussion meeting [16]. Multimodal corpus includes transcripts that are aligned or synchronized with the original audio or visual recording [13]. Earliest research of [7] said multimodal corpus was developed based on multimodal communicative behavior and can be recorded through visual display which can be writable such as speech or non-writable such as shoulder orientation, gesture, head orientation, and gaze relate to spoken content. Sign language is a good example of why multimodal corpus is necessary where it represents non-verbal language and non-verbal aspects of the language [16].

2.3 The Sesotho Language

Sotho or Southern Sotho language is one of the 11 official South African languages. According to statistics, Sesotho language is primarily spoken by 1 717 881 people in Free State and secondly 1 395 089 in Gauteng province [31]. Southern Sotho, Northern Sotho (Setswana) and Western Sotho (Sepedi) are all derived from Sotho languages and all the speakers are called Basotho [9]. The Sotho languages are closely related to Southern Bantu language such as Nguni languages that comprise of Xhoza, Zulu, Swazi, Hlubi, Phuthi and Ndebele.

Sesotho language is considered as a highly morphological language because a single word is formed by the concatenation of morphemes [9]. A morpheme is an immature or an undeveloped word normally called linguistic unit with a minimal meaning [9], and proper concatenation of them result into normal word. These morphemes include a head morpheme called prefix, a stem morpheme called an infix and a tail morpheme called suffix [17]. A stem morpheme can derive on many new words once different affixes are attached to it [9].

2.4 Related Work

[12] presented the rapid development of an Afrikaans-English speech to speech translator which is a prototype that incorporates the use of ASR, MT and TTS and designed to run on laptops or desktops computers using a close talking head-set microphone. Google translate and Google Translate Android App are Google applications that helps with the learning of multi-languages [20]. [5] described Lwazi corpus

for ASR which is a new telephone speech corpus for nine South African Bantu languages; this corpus aims to facilitate the development of the applications that will enhance education, speech enabled software and information dissemination through media. [33] expressed the TTS application for call centre automation; here, TTS engine monitors live call centre calls between live callers and live operators and detects certain key words that are spoken; these keywords are used to facilitate the report about call issues and real-time assistance of the call centre operator. [6] postulates automatic speech recognition for under resourced languages technology which focuses on integrating Bantu languages (considered as under-resourced languages) and technologies that make use of speech recognition such as In Car Messaging application.

3 Research Methodology

Given the nature of the proposed solution, qualitative research was deemed most applicable because it enabled elicitation of ideas and views of the phenomena in order to get descriptive and accurate findings. Prototyping was applied in the development of the mobile application system prototype while experimentation was used to evaluate the usability of this prototype. Given the enormous scope of developing a corpus, case study research design was adopted; for this purpose, only, a selected representation of the Southern Sesotho language was modelled and used to develop and evaluate the system prototype.

In evaluating the usability of the system prototype, sample method (involves taking a representative selection of the population and using data collected as research information [25]) was applied. The results obtained from the sample were generalized to the entire population. This research's sample was based on the population in Free State. Purposive sampling was used because the participants have some defining characteristics that make them the holders of the data needed for the study, e.g. foreigners and non-Sesotho speakers that are faced with the problem of integration to Sesotho speaking population. Once the sample was determined, the tools to obtain the data were selected. Open-ended interview was conducted; there was a conversation between the researcher and the participants, and the aim was to extract the ideas or views about the proposed application. As a qualitative data gathering technique, observations were conducted; users were videotaped while installing and using the application on their mobiles. The aim was to obtain the behavioral patterns of the participants without necessarily questioning or communicating with them. The aspects that were being evaluated using observation were: speed, robustness, compatibility and usability of the application.

4 Implementation

4.1 System Framework

UmobiTalk's development was based on the framework in Fig. 2 below; it focuses on three important aspects of the speech based machine translator: ASR, MT and TTS.

Parallel corpus was first developed and UmobiTalk app was used as a tool to evaluate the functionality of the corpus.

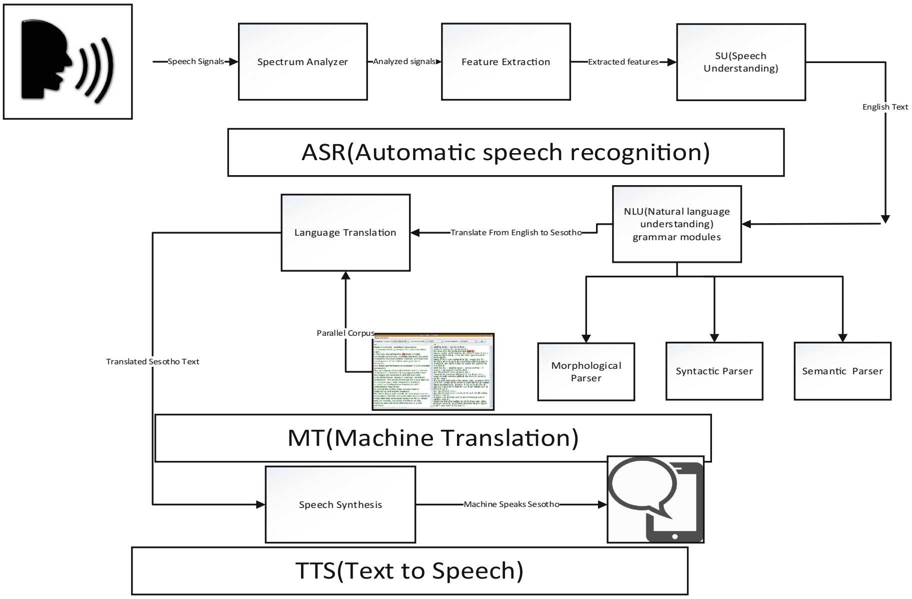


Fig. 2. Summary of processes taking place during the development of UmobiTalk

4.2 Parallel Corpus Collection

We conducted the quantitative research with the aim to obtain basic Sesotho language that migrants need to know; collected data was stored in two different files aligned at a sentence level. The first file known as the source file containing English texts and the second file known as a target file containing Sesotho translated texts. The corpus was then narrowed to only six domain aspects of the language learning: greetings, small talk, tourism (food, culture, places etc.), business, emergency words and etiquette.

4.3 Corpus Annotation

Firstly, each word forming a sentence was assigned its part of speech tag; this made it easy for the machine to understand and categorize their word class. Words that are considered to have grammatical ambiguity or words with more than one grammatical features were assigned more than one part of speech tag e.g. book_NN/VB. The phrase annotation layer was added in which sentences were broken down in to phrases; each phrase was assigned an appropriate phrase tag, forming constituents (syntactically analyzed phrases). The Sesotho texts were tagged using numerical codes that uniquely identify a word or a phrase in a sentence. These numerical codes were used to link the

Sesotho translations with English constituents in a corpus. For each two parallel lines in a corpus, the number of constituents from the source language was equal to the number of Sesotho translations from the target corpus.

4.4 Corpus Analysis

Corpus analysis is a stage in which a corpus's effectiveness, robustness, correctness etc. are evaluated by the software that will manipulate it. Normally, a monolingual corpus is evaluated in a different manner than bilingual corpus. In a huge monolingual corpus, one can use existing software analysis tools that can perform frequency list analysis, concordance and collocation, keywords and n-Grams.

In case of bilingual corpus, the UmobiTalk was used to train the corpus; the speed and translation accuracy were variables that were used to measure the corpus effectiveness. The application was tested from single word translation to multiple complicated words translation.

4.5 Umobitalk Developments

UmobiTalk was developed using ASR, MT and TTS as described in Fig. 1.

Automatic Speech Recognition. The ASR is implemented such that a user is expected to speak the word or phrase they wish to translate. The speech is in a form of acoustic signals, and is transmitted through sound waves. The spectrum analyzer of the recipient machine, which acts like an ear, analyzes and maps the speech signals on a spectrum. Feature extractor then extracts phonemes from analyzed speech signals and finally the speech understanding technology converts the extracted phonemes in to text depending on the language of interest such as English. In UmobiTalk, the recognized English text is displayed on a screen for user verification and edits. This way, the user is certain that the translated word is the correct word he/she is looking for. In Fig. 3 below, the user is prompted to speak to the machine; the spoken words will be displayed on a text box.

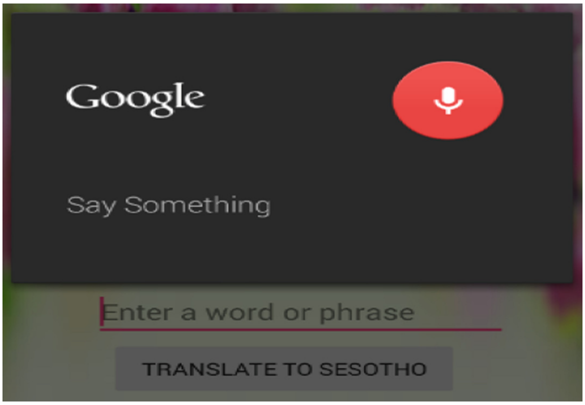
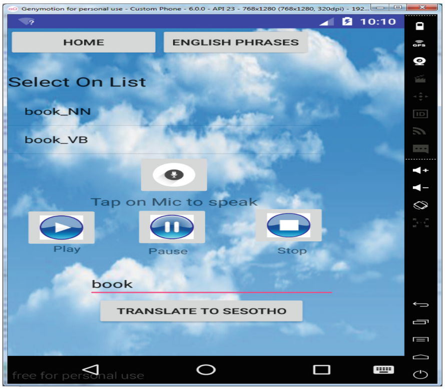


Fig. 3. Allow the user to speak to the machine

Machine Translator. Before the machine translates the inputted text, it needs to ‘understand’ the grammatical features of the text. In order to make this possible, a NLU (natural language understanding) module was developed; it comprises of morphological, syntactic and semantic analyzers. The morphological analyzer uses the tokenizer method to break down the sentence in to words known as elements, each element is then passed to the part of speech tagger which uses the machine dictionary to check the legality of word and assign it a relevant part of speech tag. The tagged words are sent to the syntactic analyzer which uses the phrase structure rules to analyze and group the words based on their grammatical dependency. The phrase structure rules are applied in a form of a parse tree, where by the analytical procedure analyzes the sentence from top to bottom and from left to right. Finally, the semantic analyzer tries to figure out the changes of meaning of words. Semantic analyzer is further explained below.

Semantic Analyzer. The semantic analyzer focuses on identifying and solving word sense ambiguity. A word sense ambiguity is a word having more than one meaning, its correct meaning is determined by its location in a sentence. Those words are labeled with more than one part of speech tag such as book_NN/VB in a machine dictionary. The semantic analyzer is encapsulated with the word sense disambiguation (WSD) module that can detect inputted single words ambiguity (Fig. 4) and in-text word ambiguity (Fig. 5).

The screenshots below demonstrate how UmobiTalk respond to above mentioned ambiguities.


Genymotion for personal use - Custom Phone - 6.0.0 - API 23 - 768x1280 (768x1280, 320dpi) - 192...
10:10
HOME
ENGLISH PHRASES
Tap on Mic to speak
Play
Pause
Stop
TRANSLATE TO SESOTHO
free for personal use

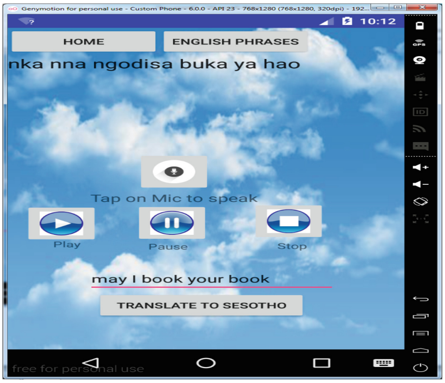
Select On List

book_NN

book_VB

book

Fig. 4. Single word ambiguity


Genymotion for personal use - Custom Phone - 6.0.0 - API 23 - 768x1280 (768x1280, 320dpi) - 192...
10:12
HOME
ENGLISH PHRASES
Tap on Mic to speak
Play
Pause
Stop
TRANSLATE TO SESOTHO
free for personal use

nka nna ngodisa buka ya hao

may I book your book

Fig. 5. In-text word ambiguity

As shown in Fig. 4 above, a user will select from the list which book is he/she referring to. From Fig. 5, WSD module is provided with the set of rules that can identify the position of word with ambiguity in a given context, and disambiguate the word based on what precedes and follows it. It should be noted that the machine ‘noticed’ that “may” is not a “may” of a month because it is followed by the noun instead of a verb; secondly, the first “book” has been recognized as a verb because it is followed by a noun instead of a verb or possessive part of speech. The last book is recognized as a noun because it is followed by the possessive part of speech.

When the language has been analyzed, the machine translator will take over. The machine translator is implemented to run all the lines inside a corpus by comparing each inputted constituent against the list of existing constituents in a source corpus to find the best match; if the match is found, then the specific Sesotho translation is extracted from a target corpus and presented on a screen.

TTS Developments. Sesotho words audio files were compiled and incorporated into UmobiTalk; these files are used when reading the Sesotho translations. The UmobiTalk's TTS operation compares each translated text with the set of audio files to find the best match. The selected audio files are stored in a media player list that is later manipulated. Time interval between the successive audio files is set to approximately 10 s to allow successive audio files to be executed immediately after the preceding one has complete. Without setting the time interval, the audio files will be executed at once or concurrently making it difficult to hear the words. The shortcoming of Sesotho TTS approach is that it has been designed to predict existing Sesotho words; therefore, new words (words not in corpus such as proper nouns) cannot be read by the application. This is currently addressed by integrating the customized Sesotho TTS with the existing English TTS to read unknown words in English.

5 Evaluation, Conclusion and Further Work

5.1 Evaluation of Umobitalk Speed and Accuracy

The system testing and evaluation was conducted by the researcher through experiments, to meticulously observe and document the language translation accuracy and the response time. The response time was basically based on the time it takes from when button translate is clicked and to when the output is displayed on a screen. To record the response time, we modified the application to have timer that will determine the speed of the language translation. The experiments were conducted to determine the overall translation speed of the application. The experiments were based on translation of word patterns known as n-Gram (number of words in a sentence) inputted by the user. These words were inputted in a form of phrases that were analyzed as constituents. Constituent is word(s) in a sentence that can be grouped and analyzed together based on their grammatical dependencies. Based on the experimentation, we concluded that the translation speed is determined by the response time of the machine, the shorter the response time the higher the translation speed. The response time is dependent on the number of constituents that are inputted by the user. The more the inputted constituents the longer the response time. Translation speed was analyzed in a form of a graph below (Fig. 6).

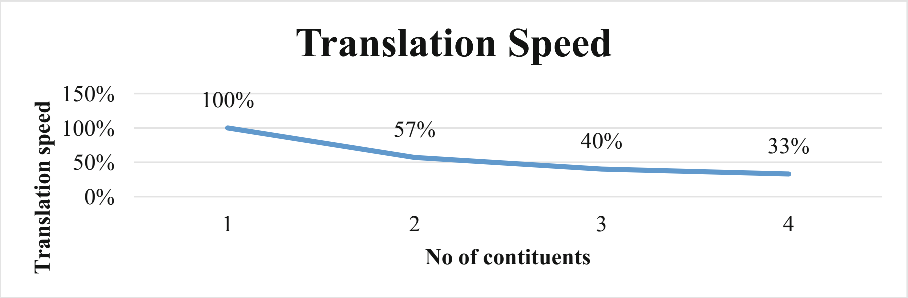


Fig. 6. Modelling of translation speed using the line graph

The overall translation speed of the machine is 57.5%.

We tested the translation accuracy in the same manner with the translation speed, by evaluating it against the number of constituents inputted. We ensured that constituents inputted abides to language grammatical rules. The translation accuracy was calculated by comparing the machine translated text with the correct Sesotho text determined by the Sesotho linguistic, to determine the closeness. Therefore, the higher the closeness the higher the accuracy rate (Fig. 7).

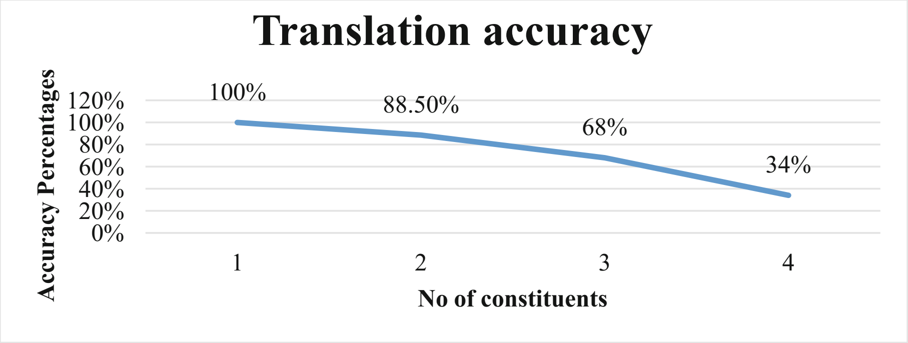


Fig. 7. Modelling of translation accuracy using the line graph

From the graph above, the more the constituents are feed to the system, the less the accuracy. The overall translation accuracy of UmobiTalk is 72.6%, the motive behind achieving such a good score was depending on well-structured parallel corpus. However, main challenge that degrades the accuracy rate was difficulties that were encountered when trying to intertwine two different languages, having two different language structures.

5.2 Conclusion and Further Work

In this paper the researchers have presented the Umobitalk application, aim to improve the Sesotho language as one of Bantu languages known to receive less attention in the field of natural language processing (NLP) and human computer language (HCL) due to lack of resources. This application was also developed to help the migrants to learn few basic words and phrases, enabling them to have foundation that they can build on to learn the whole Sesotho language. This application is quite flexible and generic enough by allowing the user to play around with existing words to form variety of sentences that can be translated.

UmobiTalk application was formed by three layers

1. Automatic speech recognition (ASR), used for input purposes
2. Machine translation (MT), used to process data
3. Speech synthesizer, used to output data.

This prototype is developed as an analysis tool to assess and improve the parallel corpus implemented; since the quality of the corpus cannot be evaluated directly, hence a tool to manipulate the corpus is prerequisite. This prototype was tested and evaluated by the researcher and group of respondents. The software proven to be quite easy to use, can work without having to access the data connection, unless a user need to use Google speech technology. Using the application lights up some aspects that needs to be further addressed as documented below.

Improvement of Accuracy from Speech Technology. ASR must be able to recognise long or continuous speech. Further work must be done to accommodate people with different accents. The Sesotho speech synthesis needs to be improved in terms of quality audio and implementation. Therefore, the study of Sesotho sounds known as morphological studies have to be revised despite the language's insufficient resources.

Addition of New Words in a Database and in Corpus (Migration from Specialized Parallel Corpus to General Parallel Corpus). Addition of words was limited to a certain number, putting in mind phone storage and processing capabilities. If pursuing a small application that can manipulate a huge corpus with millions of words, a best solution is to adopt the client server approach in which dictionary database and corpus are removed from the application to cloud server. This is an effective approach in terms of improving system functionality. However, it cannot be viable due to cost ineffectiveness, since the user must always be online to operate the system.

Integrate Translation to Other Languages. The functionality level will increase; single application can solve more than one problem. The major challenge will be the level of complexity which will increase not only on a technical basis but also on graphical user interface. The user before performing translation must first use combo box tool to select maybe the source and the target language.

Bidirectional Translation Between Two Languages. Vice versa translation between two languages will require dictionary database and parallel corpus to be extensively modified to allow backward translation from Sesotho to English. This idea will help the Sesotho speaking individuals who wants to learn English language.

Sesotho Pronunciation Learning Technology. A translated Sesotho phrase such as "setjhaba sa Qwaqwa" is difficult to be read and properly articulated by a non-Sesotho speaker such as Indians or Chinese speaking people, thus a need to learn pronunciation is vivid. As a further work UmobiTalk will integrate Sesotho pronunciation training, to enforce a proper pronunciation of Sesotho words for an effective communication principle. The only deterioration that will hinder the successful implementation of this tool is language insufficient resources and speech based technology that require thorough learning of Sesotho morphemes. The predicted advantage of Sesotho speech based technology is to help the non-native Sesotho speakers to learn complex phonological aspects of Sesotho consonants such as "kg, tl, tlh, qwa, qha" and many more that are used most of the time.

References

1. Alumäe, T., Kalijurnd, K.: Open and extendable speech recognition application architecture for mobile environments. In: Spoken Language Technologies for Under-Resourced Languages (2012)
2. Anthony, L.: AntConc: design and development of a freeware corpus analysis toolkit for the technical writing classroom. In: Proceedings International Professional Communication Conference. IEEE (2005)
3. Anodo, T.: Open source spelling checker for Kimiiru language. Doctoral dissertation, University of Nairobi (2013)
4. Anusuya, M., Katti, S.: Speech recognition by machine, a review. arXiv preprint arXiv:1001.2267 (2010)
5. Barnard, E., Davel, M., Van Heerden, C.: ASR corpus design for resource-scarce languages. In: ISCA (2009)
6. Besacier, L., et al.: Automatic speech recognition for under-resourced languages: a survey. Speech Commun. **56**, 85–100 (2014)
7. Chen, L., et al.: VACE multimodal meeting corpus. In: Renals, S., Bengio, S. (eds.) MLMI 2005. LNCS, vol. 3869, pp. 40–51. Springer, Heidelberg (2006). https://doi.org/10.1007/11677482_4
8. Collins, N.: The UbuWeb electronic music corpus: an MIR investigation of a historical database. Organised Sound **20**(1), 122–134 (2015)
9. Demuth, K.: Accessing functional categories in Sesotho: interactions at the morpho-syntax interface. In: Meisel, J.M. (ed.) The Acquisition of Verb Placement. SITP, vol. 16, pp. 83–107. Springer, Dordrecht (1992). https://doi.org/10.1007/978-94-011-2803-2_4
10. Devlin, J., et al.: Fast and robust neural network joint models for statistical machine translation. In: Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), vol. 1 (2014)
11. Duarte, T., et al.: Speech recognition for voice-based machine translation. IEEE Softw. **31** (1), 26–31 (2014)
12. Engelbrecht, H., Schultz, T.: Rapid development of an Afrikaans-English speech-to-speech translator. In: International Workshop on Spoken Language Translation (IWSLT) (2005)
13. Arshavskaya, E.: The Routledge Handbook of Language Learning and Technology, pp. 103–106 (2018)
14. Graën, J., Batinic, D., Volk, M.: Cleaning the Europarl corpus for linguistic applications (2014)

15. Green, S., Heer, J., Manning, C.: The efficacy of human post-editing for language translation. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM (2013)
16. Gries, S.T., Berez, A.L.: Linguistic annotation in/for corpus linguistics. In: Ide, N., Pustejovsky, J. (eds.) Handbook of Linguistic Annotation, pp. 379–409. Springer, Dordrecht (2017). https://doi.org/10.1007/978-94-024-0881-2_15
17. Guma, S.: An Outline Structure of Southern Sotho. Shuter and Shooter, South Africa (1971)
18. Hammo, B., et al.: Exploring and exploiting a historical corpus for Arabic. Lang. Resour. Eval. **50**(4), 839–861 (2016)
19. Hutchins, J.: From first conception to first demonstration: the nascent years of machine translation: 1947–1954. A chronology. Mach. Transl. **12**(3), 195–252 (1997)
20. Hyman, P.: Speech-to-speech translations stutter, but researchers see mellifluous future. Commun. ACM **57**(4), 16–19 (2014)
21. Imseng, D., et al.: Impact of deep MLP architecture on different acoustic modeling techniques for under-resourced speech recognition. In: 2013 IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU). IEEE (2013)
22. Jakubíček, M., et al.: The ten ten corpus family. In: 7th International Corpus Linguistics Conference CL (2013)
23. Kennedy, G.: An Introduction to Corpus Linguistics. Routledge, Abingdon (2014)
24. Kumar, A., et al.: Rethinking speech recognition on mobile devices (2011)
25. Latham, B.: Sampling: What is it: Quantitative research methods 5377 (2007)
26. Nakazawa, T., Kurohashi, S., Kobayashi, H., Ishikawa, H., Sassano, M.: 3-step parallel corpus cleaning using monolingual crowd workers. In: Hasida, K., Purwarianti, A. (eds.) Computational Linguistics. CCIS, vol. 593, pp. 79–93. Springer, Singapore (2016). https://doi.org/10.1007/978-981-10-0515-2_6
27. Niño, A.: Machine translation in foreign language learning: language learners' and tutors' perceptions of its advantages and disadvantages. ReCALL **21**(2), 241–258 (2009)
28. Paulussen, H., et al.: Dutch parallel corpus: a balanced parallel corpus for Dutch-English and Dutch-French. In: Spyns, P., Odijk, J. (eds.) Essential Speech and Language Technology for Dutch. NLP, pp. 185–199. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-30910-6_11
29. Reddy, R., Mahender, E.: Speech to text conversion using android platform. Int. J. Eng. Res. Appl. (IJERA) **3**(1), 253–258 (2013)
30. Sibanda, O.: Social ties and the dynamics of integration in the city of Johannesburg among Zimbabwe migrants. J. Sociol. Soc. Anthropol. **1**(1-2), 47–57 (2010)
31. Statistics South Africa. Population Census (2011). [dataset]. http://www.statssa.gov.za/census2011/Products/Census_2011_Census_in_brief.pdf
32. Sundermeyer, M., et al.: Translation modeling with bidirectional recurrent neural networks. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP) (2014)
33. Thenthiruperai, B., Kates, J., Miller, K.: Use of speech recognition engine to track and manage live call center calls: U.S. Patent No. 8,130,937, 6 March (2012)
34. Tian, L., et al.: UM-corpus: a large English-Chinese parallel corpus for statistical machine translation. In: LREC (2014)
35. Yang, L., Zhang, D., Tang, Y.: The realization of food corpus based on database technology. J. Simul. **2**(6), 327–330 (2014)
36. Vaughan, E., O'Keefe, A.: Corpus analysis. In: The International Encyclopedia of Language and Social Interaction (2015)