TuijinJishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 44 No.6 (2023)

BridgeTalk: A Translator from Sinhala to English and English to Sinhala

E.K. Jayawardhana , T.R.Ranasinghe , S.N.Baalasooriya , Dilshan De Silva , E.Weerasinghe

Department of Software Engineering and Computer Science

Sri Lanka Institute of Information

Technology

Malabe, Sri Lanka

Abstract: **-**This research concentrates on a transfer-based machine translation system that is equipped with translating a grammatically correct Singlish / Sinhala sentence into its corresponding English sentence and vice versa. This approach, Sinhala to English translation has never been addressed before. Because English is the language used by most people to conduct daily business, learning it is now compulsory and very beneficial towards the growth. In the process of translating Sinhala sentences to English the user input can be in either Sinhala or Singlish and the reverse way which is, when translating English sentences to Sinhala input should be in English. The main objective of this translator is to enable a smooth flow translation of sentences and paragraphs to not only to the locals but also to the foreigners and thereby eliminate the language barrier between and within the groups.Application is compatible with both Sinhala and English efficient voice typing capabilities with added benefits. A considerable amount of rules, patterns and words of both languages were used to develop this system. The task of creating a Sinhala to English language translator was undertaken in order to meet this requirement and build sentences that are grammatically sound. This system delivers translations with 90% of accuracy approximately 65 sentences under 1 minute.

Keywords*: English, Sinhala language, transfer based machine translation, translator.*

1. Introduction

Effective language translation technologies have emerged as essential instruments for promoting communication and understanding across many linguistic communities in a world that is becoming more linked and globalized. This research article explores the creation and analysis of a ground-breaking machine translation system intended to translate between English and Sinhala, two languages of substantial cultural, regional, and international significance.

The study addresses the critical need for a reliable and precise Sinhala to English translation system, meeting the linguistic and communication needs of the country's approximately sixteen million native Sinhala speakers. In a variety of fields, including worldwide business and academics, English is crucial as a global language. Individuals attempting to communicate, learn, and work in the absence of a comprehensive translation system have faced difficulties.

Although many initiatives have been successful in tackling comparable issues, there is still a glaring gap in the field of translation abilities, especially when it comes to Singlish to Sinhala and English. The lack of grammatically accurate translations made possible by voice input commands is particularly notable and the research is aims to fill the void that remains this area.

In order to facilitate precise translation between Sinhala and English, the research introduces the Transfer-based machine translation approach. The system is innovative not only because of its translation powers but also because of the variety of functions it provides, Furthermore, the research team aims to enhance the translator's

capabilities by incorporating the following functionalities: Singlish to Sinhala and Sinhala to English grammatically precise translations, facilitated through both keyboard inputs and voice inputs that make it a complete language solution.

The significance of this research lies in its contribution to overcoming language barriers through the development of a machine translation system. As globalization continues to ability to seamlessly translate between languages becomes more and more important. This study addresses a critical gap in translation capabilities, focusing on the unique challenges posed by the Sinhala language, especially Singlish. By offering accurate translations and innovative features, such as voice input commands, aids in language learners and individuals seeking to engage in multilingual environments.

Development of a robust as well as accurate Sinhala to English machine translation system, addressing the specific challenges, inputs facilitated by both keyboard inputs and voice commands, maintain accuracy, speed, and usability throughout the application are under the primary objectives of this research to explore.

This paper contributes significantly to the field of machine translation by introducing a Sinhala to English translation system. Furthermore, research discusses the challenges, methodologies, and outcomes of building such a system, demonstrating its potential to its finest.

2. Literature Review

Language translation, the process of converting text from one language to another, has gained significant attention due to its vital role in enhancing cross lingual communication and understanding of electronic content. With the increasing globalization and interconnectedness of societies, the need for accurate and efficient language translation tools has become essential, and Sinhala is among world's top 100 languages and ranked in 62nd position [1].

But among many translators there are only few Sinhala to English and English to Sinhala translators. Main reason for this is when translating Sinhala to English, translators encounter, culture bound terms and dialects in literary translations. The outcome of the translation should be a compromise of the source language culture and the target language culture. But in case of Sinhala and English cultural communities, they hardly share parallel cultural aspects. Apparently, Sinhala and English communities belong to two different worlds. [2] Currently the best Sinhala to English and English to Sinhala converter is Google Translator. Google translator is a neural machine translation service that enables the translation of text, documents, and websites from one language to another.

Source [1] cites that Sinhala to English translation used by Google Translator ,the most common language translator used within the community is not 100% accurate. Google translator is incapable of producing grammatically sound full translations although it is the most renowned and trusted translator in the industry.

Source [3] cites Sinhala to English Translator which is capable of interpreting Sinhalese words written in English characters (Singlish) to Sinhala characters, and an English grammar and spell checker, with a full translation accuracy of 87%.

Another approach to translate Singlish to English in cited in source [4], using Trigram and Rule base models. The studyfocuses on language transliteration, particularly addressing Singlish, a form of communication where Sinhala words are represented using English letters. It evaluates various methods for converting Singlish to Sinhala, specifically highlighting a trigram-based approach. This method combines a trigram model with an early UCSC rule-based system. The trigram model is trained using Singlish YouTube comments and their corresponding Sinhala transliterations, employing a trigram tagger implementation similar to POS (Part-of-Speech) tag datasets. By combining the Trigram and Rule-based models, the study achieves a significant improvement in transliteration accuracy, reaching 77% letter-wise accuracy. Moreover, the approach emphasizes a user-centric design, enhancing the user experience for capturing Singlish content effectively.

Some of the available converters are the Easy Unicode Converter [5], the Google Input Tools [6] and the converter found in [7]. Several dictionaries have been introduced with the ability to provide the English meaning of any word typed in Singlish and the Sinhala meaning of any word typed in English. An example of such a

dictionary is the Madura dictionary [8], which offers an extensive collection of definitions for both English and Sinhala languages. Another dictionary app is Sinhala Dictionary Offline, for android users.

The development of the Sinhala to English translator uses HTML and CSS because of the chosen Sinhala to English translation APIs do not support React framework. This choice produces the adaptability to intricate language translation tasks. The methodology employs modular frontends powered by HTML, CSS and JS, and backends orchestrated by JS resulting intuitive user interfaces.

Challenges that User encounters

The importance of this research project is its ability to greatly influence the field of bilingual education as users often feel inconvenienced and frustrated when using a foreign language for educational and other purposes. This inconvenience includes challenges such as,

Language barriers

difficulty in understanding course materials

• reduced engagement,

which can hinder effective learning and communication. The development of an adaptable Sinhala to English translator, aims to address these issues by providing more accessible and user-friendly language translation tools, thus enhancing the overall learning experience for individuals who may struggle with language differences.

BridgeTalk emerges as a comprehensive solution addressing the challenges as well as an innovation. This platform empowers novice programmers to navigate coding complexities more efficiently and intuitively by seamlessly integrating advanced AI technology. Moreover, BridgeTalk is a bridge, fostering effective communication and interaction between personals, thus enriching the educational journey.

Bridge Talk, Sinhala-English Translator is equipped with a range of functions designed to effectively fulfill its intended objectives Such as,

Sinhala To English Translator : Bridge Talk's primary function is the translation of Sinhala text and speech into English.

Singlish to Sinhala Conversion : This feature is specifically built specifically to convert Singlish, a blend of Sinhala and English, into Sinhala. This conversion process aids in preserving cultural nuances and ensuring accurate / seamless communication.

Translation Process : This section of the paper will provide a detailed analysis of the algorithms and methodologies used by BridgeTalk to ensure precise and accurate translations.

Sinhala Voice Typing : Since Voice typing has emerged as rather more convenient method of inputting data. BridgeTalk incorporates a Sinhala to English voice typing feature to the its own system, which will be explored in-depth in this paper. This is one of the features which has the potential to enhance accessibility and user experience in cross-linguistic communication.

In summary: The evolving landscape of language translation, crucial for global communication, spotlights the significance of accurate tools in bridging linguistic divides. Among the world's top 100 languages, Sinhala holds the 62nd position, yet there's a major shortage of robust translators catering to Sinhala and English translation. Challenges emerge when cultural terms and dialects encounter translation barriers, leading to compromises in preserving the essence of both source and target languages. Notably, the stark cultural differences between the Sinhala and English communities hinder seamless translation.

While Google Translator remains a widely used tool, its Sinhala to English translations fall short of complete accuracy, prompting the search for alternative solutions. Source [3] introduces a translator boasting an impressive 87% accuracy in interpreting Sinhalese words written in English characters (Singlish) to Sinhala

characters, while Source [4] delves into a trigram-based approach for Singlish to Sinhala transliteration, achieving a notable 77% letter-wise accuracy.

Various converters and dictionaries, such as Easy Unicode Converter, Google Input Tools, and others, have emerged to aid in translation tasks and offer extensive definitions for both languages. The technological limitations of chosen translation APIs have led to the utilization of HTML, CSS, and JS in developing the Sinhala to English translator, ensuring adaptability and intuitive user interfaces.

This research project aims to alleviate language barriers in education by providing accessible and user-friendly translation tools, enhancing learning experiences for individuals grappling with language differences. BridgeTalk, the proposed comprehensive solution, merges AI technology and intuitive interfaces to facilitate effective communication. Its functionalities span Sinhala to English translation, Singlish to Sinhala conversion, detailed algorithmic analysis, and even Sinhala voice typing, promising improved accessibility and user experience in cross-linguistic communication.

3. Methodology

BridgeTalk offers several functions regarding language translation process. Each of these functions will be examined below in details to gain a comprehensive understanding of BridgeTalk's capabilities and their implications for effective language translation and communication.

3.1 Sinhala To English Translator

This module's main goal is to make it easier to translate content from Sinhala into its equivalent English text. Users start off this process by typing in Singlish. To ensure proper representation and comprehension, the system automatically changes this input and effortlessly converts it into the Sinhala script. A grammar checker then confirms the supplied Sinhalese sentence(s) or paragraph(s)'s grammar. Then it will recognize the user's supplied phrases or units. Full stops are used as delimiters to define the boundaries of the units. The SEES system limits input in this situation to sentences that end in a full stop, enabling methodical unit processing.

Sinhala to English Translation Process:

- *Input:*Users begin by providing content in Sinhala.
- *Conversion:*The system automatically converts Sinhala input into English text for accurate representation and comprehension.
- *Grammar verification:*A grammar checker verifies the grammar of the provided Sinhalese sentences or paragraphs.
- *Phrase/Segment Recognition:*The system identifies user-supplied phrases or units, delimited by full stops, ensuring methodical unit processing.
- Full stops are utilized as delimiters to establish boundaries between units for systematic processing.
- The SEES system restricts input to sentences ending with a full stop, facilitating organized and systematic processing.
- Return the converted English text.

And for English to Sinhala Translation, translation process works in vise versa.

In addition to that system provides options to interchange languages between the translator, action of copying to clip board in both languages and option to translations to be downloaded as following Fig.5.

Fig. 5. Sinhala-English Translation

3.2 Singlish to Sinhala Conversion

This study offers an automatic method for turning text written in Singlish, a Romanized version of Sinhala, into Sinhala characters. The suggested strategy uses a methodical procedure to bring about this conversion. The following lists the essential steps:

- Initialize arrays for Sinhala vowels, consonants, special characters, and their corresponding Unicode representations.
- Define mappings between the Unicode characters for Sinhala and the characters used in Singaporean. For instance, "oo" in Singlish corresponds to " ෝ “in Sinhala.
- Define special consonants and characters that require unique handling, such as '◌◌ං ' and '$' + '\u200D'.
- Implement a function to take Singlish text as input.
- Consonant and special character combinations should be iterated through in the function, and their equivalent Sinhala Unicode characters should be used in their stead.
- Process consonants and vowels in the Singlish text and replace them with the appropriate Sinhala representation.
- Replace vowel modifiers that follow consonants with Sinhala Unicode.
- Substitute Sinhala Unicode counterparts for consonants, followed by the character HAL.
- Return the converted Sinhala Unicode text.

The following Fig. 1 and Fig. 2 show examples of the use of each combination of characters in Sinhala when entering text in Singlish.

අ (a)	ආ (aa)	ඇ (A)	ඈ (Aa)	ඉ (i)
ඊ (ee)	උ (u)	ඌ (uu)	ඍ (\R)	එ (e)
ඒ (ea)	ඓ (I)	ඔ (o)	ඕ (oe)	ඖ (au)
අ (a\n)	අඃ (a\h)	ක (ka)	ඛ (kha)	ග (ga)
ඝ (G)	ඞ (\N)	ඟ (nnga)	ච (cha)	ඡ (Cha)

ජ (ja)	ඤ (q)	ඤ (KNa)	`c (nnja)	ට (ta)
ඨ (Ta)	ඩ (da)	ඪ (Da)	ණ (Na)	ඬ (nnda)
ත (tha)	ථ (Tha)	ද (dha)	ධ (Dha)	න (na)
ඳ (nndha)	ප (pa)	ඵ (Pa)	බ (ba)	භ (bha)
ම (ma)	ඹ (Ba)	ය (ya)	ර (ra)	ල (la)
ව (wa)	ශ (sha)	ෂ (Sha)	ස (sa)	හ (ha)
ළ (La)	ෆ (fa)			

Fig. 1. Sinhala Alphabet

j (k)	ක (ka)	කා (kaa)	කෑ (kA)
m (ki)	n (kie)	o (ku)	p (kuu)
කj (kei)	කකා (ko)	කෙකr(koe)	කෙකෟ(kau)
ත්‍රා(kraa)	ත්‍රෑ (krA)	ත්‍රූ(krAa)	v (kri)
කෘ (kru)	කෘෘ(kruu)	කෲt (kre)	කෙෲ (krei)
කෙ#$(kroe)	z (\rka)	කය (kYa)	කක (kI)

Fig. 2. Letter Variations

3.3Translation Process

The core of this Sinhala to English translator system are translation between Sinhala and English, Singlish to English. The system's adaptability and utility can only be improved with this bidirectional translation feature. The study incorporates the commonly used machine translation tools, such as:

- Google Translate API
- Uni Code -Developer API into the framework in order to build this application.

Text may be both seamlessly and effortlessly translated between the two languages using the above-mentioned APIs, which offer a powerful and user-friendly platform. It is used in the context of this research to support several essential capabilities.

The technology can convert Sinhala text into English using its initial feature. Users who enter text in Sinhala and need an English translation may find this feature especially helpful.

Contrarily, the system can convert text into Sinhala thanks to the second feature. Users that submit content in English can easily get a translation into Sinhala, ensuring that the system can serve a wide range of users.

The system also provides the option to convert Singlish sentences to be converted into Sinhala to ensure the ease in handling the Sinhala language.

And the also provides predefined letters Fig.3 with special characters as buttons which are not commonly used and inconvenient to type for better user experience. These buttons are generated using JavaScript and when user clicked on a certain letter, that letter imports into the input section.

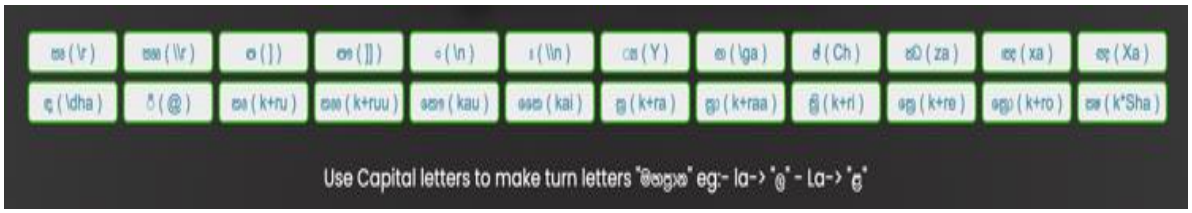

Fig. 3. Letter Variations

All above-mentioned capabilities are equipped with text copy option to provide the user with time efficiency when dealing with translation.

3.4 Sinhala Voice Typing

The voice typing functionality enables users to input Sinhala text through speech recognition, with subsequent translation into English using the Web speech API. The research methodology for this component is structured into several key phases.

The core functionality hinges on the integration of the browser's built-in Speech Recognition API (with a fallback to web kit-Speech-Recognition for compatibility). This API enables the application to capture spoken words or phrases and convert them into textual form, which is then displayed to the user.

And the system provides Fig. 4 options as copy to clip board, delete text, start/ stop listening and to change the listening language as English, Sinhala.

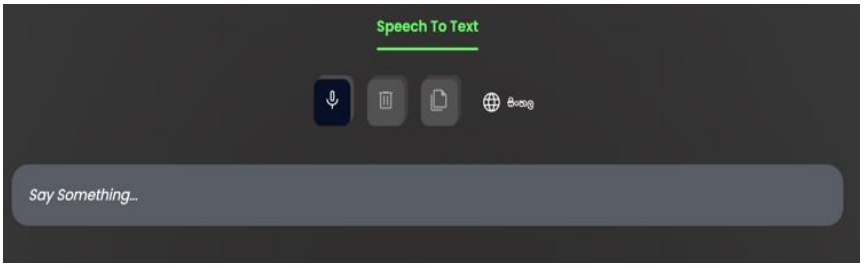

Fig. 4. Letter Variations

3.4.1Data Collection.

The success of the voice typing feature heavily depends on the quality and diversity of the training data used for speech recognition. Data collection for voice typing encompasse the following steps:

- *Sinhala Speech Data Collection:* A diverse dataset of Sinhala speech samples was collected from various speakers, ensuring variability in accents, tones, and dialects. These samples were recorded under controlled acoustic conditions to minimize noise.
- *Data Preprocessing:* Collected audio recordings were preprocessed, which included noise reduction, audio file format conversion, and segmentation into smaller speech units.

3.4.2 Speech Recognition Model Development

The core of the voice typing feature is the speech recognition model, which converts spoken Sinhala into text. The model development process involved the following steps:

Language Model Integration: A language model for Sinhala was integrated into the speech recognition system to improve context-awareness and accuracy.

3.4.3 Integration with Web Speech API based Sinhala and English Speech Recognition

The integration of the Web speech API was a pivotal step in enabling the translation of recognized Sinhala speech into English. This phase included the following tasks:

API Integration: The Web speech API was integrated into the voice typing system to enable real-time translation of recognized Sinhala speech into English text.

3.4.4 User Interface and Interaction Design

The user interface for voice typing was designed to facilitate a seamless user experience. Key considerations during this phase included:

- *Voice Input Interface:* A user-friendly voice input interface was developed, allowing users to initiate voice typing easily.
- *Feedback and Confirmation:* The interface provided real-time feedback on recognizedspeech and allowed users to confirm or edit the transcribed text.

4. Conclusion and Future Work

In conclusion, this research endeavors to address the significant gap in Sinhala-English translation by introducing a transfer-based machine translation system. By focusing on the translation of grammatically correct Singlish/Sinhala sentences into their corresponding English counterparts and vice versa, this study pioneers a novel approach in the field of language translation. The importance of English as a global language necessitates accessible and accurate translation tools, and this system aims to bridge linguistic barriers prevalent not only within local communities but also between diverse language groups.

The developed system, with its compatibility for both Sinhala and English voice typing capabilities, showcases a commitment to inclusivity and ease of use. Drawing upon a comprehensive database of language rules, patterns, and vocabulary from both languages, this system strives to produce translations that are both linguistically accurate and grammatically sound.

The achieved 90% accuracy rate in translating approximately 65 sentences within a minute demonstrates the system's efficiency and potential for practical application. This research serves as a foundational step towards enabling smoother communication, fostering better understanding, and eliminating language barriers in today's interconnected world.

As technology advances and linguistic complexities persist, continuous efforts in refining and enhancing such translation systems will play a pivotal role in facilitating cross-linguistic communication, thereby contributing to a more inclusive and accessible global community.

Future work in this domain holds promising avenues for further enhancement and expansion of Sinhala-English translation systems. To advance the field, the following areas merit attention:

- Refinement of Accuracy
- Incorporating Contextual Understanding
- Enhanced User
- Expansion to Dialects and Domains
- Real-Time Translation

Refrences

[1] “Evaluation Scores of Google Translate in 107 Languages,” Google Docs. https://docs.google.com/spreadsheets/u/0/d/1fJQLMj8O5z3Q7eKDxi1tNNrFipiEL0UDyaEF0fleZ54/edit?pli=1#gid=0 (accessed Sep. 03, 2023).

[2] “Cultural Challenges in Sinhala-to-English Translation. (2020, September[Online].www.dailymirror.lk. https://www.dailymirror.lk/newsfeatures/CulturalChallenges-in-Sinhala-to-English-Translation/131-194973 (accessed Sep. 03, 2023).

[3] SEES[Online]: Available https://ieeexplore.ieee.org/abstract/6421408

[4] W. M. P. Liwera and L. Ranathunga, “Combination of Trigram and Rule-based Model for Singlish to Sinhala Transliteration by Focusing Social Media Text,” 2020 From Innovation to Impact (FITI), 2020, doi: https://doi.org/10.1109/FITI52050.2020.9424880.

[5] “Easy Sinhala Unicode Converter,” www.easysinhalaunicode.com. https://www.easysinhalaunicode.com (accessed Oct. 05, 2023).

[6] “Try Google Input Tools online – Google Input Tools,” www.google.com. https://www.google.com/inputtools/try/

[7] Sinhala Transliteration (Singlish to Sinhala) converter [Online]: Available:ttps://www.t30apps.com/sinhala-transliteration-converter/

[8] “Madura English-Sinhala Dictionary - Online Language Translator,” www.maduraonline.com. https://www.maduraonline.com/ (accessed Oct. 03, 2023).

[9] Sinhala Dictionary Offline [Online]:Available: https://play.google.com/store/apps/details?id=sachith.com.dictionary.offline&hl=en&gl=US&pli=1

[10] L. Wijerathna, W.L.S.L. Somaweera, S.L. Kaduruwana, Y.V. Wijesinghe, D. I. De Silva, K. Pulasinghe, S. Thelijjagoda, "A Translator from Sinhala to English and English to Sinhala (SEES)," Proc. International Conference on Advances in ICT for Emerging Regions (ICTer), Colombo, Sri Lanka, December 2012, pp. 14-18.

[11] De Silva, A. Alahakoon, I. Udayangani, V. Kumara, D. Kolonnage, H. Perera, S. Thelijjagoda, “Sinhala to English Language Translator," 4th International Conference on Information and Automation for Sustainability, Colombo, Sri Lanka, 2008, pp. 419- 424.

[12] D. I. De Silva, P. K. D. A. Alahakoon, P. V. I. Udayangani, D. Kolonnage, M. H. P. Perera, and S. Thelijjagoda. "Application of Transfer based Machine Translations from Sinhala to English," In Proceedings of the 4th SLIIT Research Symposium, vol. 2, pp. 33- 36. 2008.

[13] T.W. Wei-Jing Z. K. Papineni and S. Roukos, "Bleu: a method for automatic evaluation of machine translation," in Proc. 40th Annual Meeting of the Association for Computational Linguistics (ACL), July 2002, Philadelphia.