REQUIREMENTS FOR AN ENGLISH-SINHALA SMART BILINGUAL DICTIONARY: A REVIEW

L Samarawickrama[1] and B Hettige

Department of Computer Science, Faculty of Computing, Sir John Kotelawala Defence University, Sri Lanka

[1]lasikasamarawickrama@gmail.com

Abstract- This paper presents a review on existing English-Sinhala Dictionaries while considering today's requirements for enhancing English language skills for users who are not fluent in English . Dictionaries are becoming popular as a means for interactive communication between languages. Dictionaries, and language professionals, in general, have long claimed that dictionaries are deficient, especially regarding accessing and updating content. Some authors have also noted that these deficiencies are compounded by the fact that language professionals do not receive proper training in dictionary use, and due to those deficiencies, they do not fully benefit from them. Electronic dictionaries include new search capabilities, not found in traditional dictionaries, that could meet users' needs efficiently. So, this review, it's discusses the features of the dictionaries with a comparison to existing dictionaries. Finally, its state the features that should be added to those existing ones to become a smart bilingual dictionary.

Keywords- Bilingual- Dictionary, Morphology, Context Representation

I. INTRODUCTION

The EF English Proficiency Index("EF English Proficiency Index - Sri Lanka," 2017) ranks countries by the average level of English language skills. According to the EF English Proficiency Index Sri Lanka has a very low proficiency. To enhance English language skills English to Sinhala bilingual dictionary can be used and it significantly effects many activities. In addition to the above, most of the people use a mobile phones and hand-held devices. So, developing an English Sinhala dictionary as a mobile application or that can access through the mobile devices will be worthful.

A dictionary, sometimes known as a guide, is a collection of words in one or more specific languages. ("Dictionary," 2017a). A dictionary can give you a clear definition, grammatical form, comprehensive information on pronunciation, and most importantly carefully chosen examples that show how to use the word. Learners, teachers and textbook designers should be aware of the importance of the dictionary as an important learning device. A dictionary used as a reference, a guide for language teachers and learners with highly confidential help in teaching. It is not for decoding meaning, but encoding is another function of a dictionary.

There are several types of dictionaries. In many languages, words can appear in many different forms, but only the underlined or unconjugated form appears as the headword in most dictionaries. Dictionaries are most commonly found in the form of a book, but some newer dictionaries, like StarDict ("StarDict," 2017) and the New Oxford American Dictionary ("New Oxford American Dictionary," 2017) are dictionary softwares running on PDAs or computers. There are also many online dictionaries accessible via the Internet.

The online dictionaries are limited in some way due to various reasons. Typical limitations are lack of recorded pronunciations (Longman, Oxford, Cambridge), missing phonetic transcriptions (Longman), a smaller number of example sentences. ("Comparative review of dictionaries for English learners | Antimoon," 2017)

Accessing Web dictionaries is lower than a dictionary that's installed on your computer. It's not only about the page loading time and server problems. Often, the interface requires more looking around and clicking. When you look up dozens of words time to time every day , it can really hurt your motivation if your dictionary slows you down. So that it reduces the learning of an language.

Therefore, web dictionaries are a useful resource when you want to get some extra information about a word (e.g. get some more examples, double-check the pronunciation). So with the development of the technology developing a dictionary application which can access through the mobile phone by everyone will be worthful to reduce the distance between languages.

There are a number of online and offline dictionaries available for English-Sinhala language pairs including Malalasekara, Godage, Madura etc. These dictionaries show their features and weakness. Thus, this paper reports an overview of the existing dictionaries including their approaches. In addition, this paper also discusses what are the required technologies and features for a new dictionary that can be used as a tool for enhancing English language skills.

The rest of the paper is organized as follows. Section 2 reports a summary of the historical development of the dictionaries. Section 3 reviews existing dictionaries and their approaches including their features and limitations. Finally, section 4 discusses requirements for a new smart dictionary including new features that can be included.

II. HISTORICAL DEVELOPMENT

History of dictionaries dates back to early centuries. The Akkadian wordlists which was discovered in Ebla (modern Syria) and dated roughly 2300 BCE. A Chinese dictionary in the 3rd century BCE was the earliest surviving monolingual dictionary. The first Sanskrit dictionary, the Amarakośa, written around 4th century CE. Written in verse and it's listed around 10,000 words. The first Japanese dictionary was Niina glossary of Chinese characters. The oldest existing Japanese dictionary, Tenrei Banshō Meigi, was also a glossary of written Chinese. ("Dictionary," 2017b)

A. Dictionaries found in Britain:

The earliest dictionaries in the English language were glossaries of French, Spanish or Latin words along with their definitions written in English. The word "dictionary" was invented by an Englishman called John of Garland , he had written a book Dictionarius to help with Latin "diction". An early non-alphabetical list of 8000 English words was the Elementary, created by Richard Mulcaster. The first purely English alphabetical dictionary was A Table Alphabeticall, written by English schoolteacher Robert Cawdrey.

B. Sinhala Language Dictionaries:

Sinhala is a Sanskrit term that is corresponding to Middle Indo-Aryan word Sīhala. There are four stages in the development of Sinhala language as Sinhalese Prakrit, Proto-Sinhalese, Medieval Sinhalese and Modern Sinhalese. The Sinhalese alphabet, Sinhala Hadiya, is based on the ancient Brahmi script, as are most Indian scripts. The Sinhalese alphabet is closely related to South Indian Grantha alphabet and Khmer alphabet. ("Sinhalese language," 2018)

The Sinhala Dictionary in respect of the Sinhala Language that has evolved over a long period of time through the two thousand years, with various changes and influences. Compilation of Sinhala dictionaries is not an unusual exercise in the Sinhala language. Several dictionaries which are written in the glossary style in Pali and Sanskrit languages are available in the Sinhala literature. Several sinhala dictionaries were written in alphabetical order by scholars such as Kataragama Rajaguru Bandara, Veragama Punchi Bandara, and Ven. Boruggamuve Revata Thera is found in the history of lexicographical works, they are not comprehensive and complete dictionaries. It was with the advent of European nations in the 16th Century, that the need for a modern, scientific dictionary was felt. The initiative was taken by the British. nationals such as Benjamin Clough, Charles Carter and John Callaway in compiling Sinhala - English dictionaries is seen as a result. ("History," n.d.)

III. TECHNIQUES BEHIND THE DICTIONARIES

This section briefly describes some required features and concepts for a Bilingual-Dictionary.

A. Technology

The development of new technologies and the Internet have progressively modified the concept of the dictionary.

Most of the paper dictionaries have been converted to electronic formats and new online dictionaries have been implemented.Based on the technical evaluation, the author distinguishes between online or offline bilingual dictionaries. Offline dictionaries comprise books, pocket electronic dictionaries (PEDs) and PC dictionaries. PC dictionaries include dictionaries in CDROM, floppy disk, and other formats. Based on meta(lexicographic) evaluation,("Lexicography," 2017) this typology distinguishes between electronic dictionaries based on their paper versions, and newly developed electronic dictionaries with different newly added features.As for the online dictionaries, web applications ,mobile applications can be taken.

B. Quality of example sentences

Example sentences are the most important feature of a dictionary. By looking at an example sentences anyone can figure out the answers to three critical questions: 1 What does this word mean? 2 In what situations do native speakers use this word? 3 What words and grammatical patterns typically appear with this word? Examples also give you ready-made phrases that you can select and use in your own sentences.

C. Quality of definitions

In a learner's dictionary,("Merriam-Webster's Learner's Dictionary," n.d.) definitions should tell the learner about the most common situations and contexts in which a word is used.It should give each and every meaning that could be with a word. By having a definition is easier to understand and more useful than a more general, abstract one.

D. Accuracy and completeness of phonetic transcriptions

Dictionaries should list all the common pronunciations of a word in both British English and American English (General American).It will be easier for the user to get through.

E. Quality and coverage of recordings

 In most of the online dictionaries, this feature cannot be identified. Audio recordings should sound clear with no technical defects. It should be indicated which

transcription is pronounced and the recording should match this transcription. Ideally, there should be a recording for each British and American transcription, as well as for word forms (-s, -ed, etc.)

F. User friendliness

This is an important feature which is necessary for the user.Since every user seeks for an user-friendly interface software should enable the user to perform all the common tasks (searching up a word, playing a recording, etc.) easily and without unnecessary clicking. It should offer all the expected features like selected text input (like copy and paste) and should support all the expected input operations .

G. Search engine

Searching is part of "User friendliness", because technically user seeks for that feature first. But it is so important in a dictionary that it deserves its own category. If a word has many spellings (e.g. hard-pressed/hard pressed), it shouldn't reason out which one you search for.

H. Layout

Definitions, phonetic transcriptions, and example sentences should be laid out in a clean and readable way. The user should not feel overwhelmed by the interface element.

I. Other features

Direction :Direction should be highly concerned whether it's unidirectional, bi-directional or both. Most of the present world dictionaries contain an only unidirectional translation.

Pronunciation: Another feature that will be easier for the user so that user can get the word.

Grammer information: For those who learn English as a language this feature will be more beneficial because grammar is a necessary component in each and every language.

Mode: Most of the dictionaries are Unicode and as an additional content font, images can be used as the mode.

Morphological support: Adding a particular form, shape, structure or an image will be an additional feature which can be included in a smart bilingual dictionary.

Freely availability: Nowadays most of the web and mobile applications are freely available to use and if it's an online dictionary, it will be beneficial if the data usage for one search is less.

IV. EXISTING DICTIONARIES

There are so many existing dictionaries, only a few were discussed below.

Malalasekara dictionary was one of the oldest dictionaries and now in the present, it can be taken as an online dictionary too. (" Malalasekera," 2017). Most of the other dictionaries follow words and meaning available in the Malasekara dictionary to generate their own dictionaries since it contains higher number of definitions.

Derana English to Sinhala Dictionary is a feature-rich dictionary which is developed by the support of the Sinhala community. Derana dictionary consists of more than 300000 definitions. Technical terms from medicine, science, law, engineering, accounts, arts and many other sources can be found in this dictionary. Some additional features of this dictionary are pronunciations, listing most used word at the beginning, and relevant sentences for additional reference (beta). Transliteration facilitates to type Sinhala Unicode in an easy manner. Derana dictionary is the first dictionary in this caliber for the Sinhala language. ("Diana English Sinhala Dictionary," n.d.)

Madura English-Sinhala Dictionary ("Madura English-Sinhala Dictionary," 2017)is a free electronic dictionary service developed by Madura Kulatunga. It is available as computer software, an online website, and an android app. The dictionary contains over 230,000 definitions including various technical terms.

EnSiTip(Wasala & Weerasinghe, 2008) is an English-Sinhala translation tool. The primary purpose of the tool is to assist users who have a basic knowledge of English but are unable to grasp much of what the largely English web has to offer. EnSiTip is an easy-to-use, non-intrusive popup dictionary add-on for Firefox. It mostly

lies apparently dormant, until an English lexical item unfamiliar to the user is 'hovered over' in a website.

Bhasha Sinhala Dictionary("Bhasha Dictionary | Sinhala-English Dictionary," n.d.) is Sri Lanka's Official Sinhala-English Dictionary app for Android . It provides you offline dictionary service for English to Sinhala word translations and online dictionary service for both English to Sinhala and Sinhala to English word translations.

Buddi Dictionary("Buddi dictionary," n.d.) is an English to Sinhala Bilingual dictionary developed to provide Sinhala meaning for the given English term, Buddi Dictionary comes with several features. The dictionary operates with three modes namely Unicode mode, Singlish mode, and Image mode. In the Unicode mode,
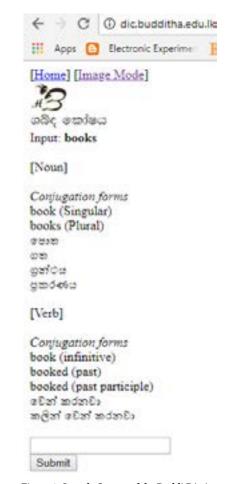

Figure 1. Sample Output of the Buddi Dictionary

Table 1.Comparision the available features on the commonly selected dictionaries

Dictionary	Malalasek	Madura	Derana	Ensitip	Basha	Godage	Buddi
Technology	Book	Web	Web	Web	Mobile	Book	Web
Direction	E to S	both	E to S	E to S	E to S	E to S	E to S
Pronounciation	No	No	No	No	Yes	No	Yes
Grammer Info	Yes	No	No	No	No	Yes	Yes
Mode	Unicode	Unicode	Unicode	Unicode	Unicode	Unicode	Unicode,Font,Image
Morphological supp	No	No	No	No	No	No	Yes
Freely Available	No	Yes	Yes	Yes	No	No	No

results provide through the Sinhala Unicode, and Singlish mode provides transliterated text for Sinhala. For that, it uses Sinhala to English transliteration method. In the image mode results are generated through the images. Therefore any simple mobile phone (normal cell phone) can display this information easily. Technically, Buddi dictionary is a kind of word translator than a dictionary. In addition to the Sinhala meaning, it shows conjugation form of the English word. This feature is useful to uses to get more idea on the English word. In addition, this feature helps users to take more grammatical knowledge on that word., To provide these features, the system uses an English morphological analyzing tool that can identify each English word with its grammatical coronary. Figure 1 shows the result of the input word 'books'. According to its simple design data usage is very low than the others. (less than a kilobyte for a single dictionary search)

V. DISCUSSION

This paper reported a brief review of existing dictionaries and available features that can be beneficial for people who are not fluent in English. In addition, this review was efforts of some related research in the use of dictionaries in the context and their existing techniques etc. so, Table 1 compares the available features on the commonly available selected dictionaries. In addition, different findings have been obtained, the first of which is the importance of bilingual dictionaries is to have language translation. When it takes two words it should give the best solution or the meaning, therefore it will be easy for the learner to understand. The variation of word's meaning should be clearly identified. As for an example when it takes the word book it has two meanings so it should be clearly mentioned. The findings give an overall beneficial effect of dictionaries to have and idea about what new features should be added. A great attention can be allotted to dictionary research through means of questionnaires,

surveys while accomplishing tasks, to take an idea about the importance of investigating and minimizing the mistakes made. Therefore this behavior will boost up the importance of this tool and encourage to develop warm intimacy with dictionaries and to reduce the language gap. From the review of the literature and the analysis of a set of existing dictionaries, it can be concluded that developing a software-based dictionary will be important. In the present every one uses the mobile phone.so as a further work those online dictionaries can be developed as a mobile application, therefore it will be so much help to get knowledge with regarding difficult words.

As the final conclusion, a simple dictionary incorporating all above mention facilities is required. Especially bidirectional support, selected word support, the pronunciation for the given English word and if the user required more details, then dictionary should be capable to provide sample sentences to get a clear idea about the meaning. As further work of this research, Buddi dictionary can be improved to provide maximum facilities to enhance users English language proficiency.

REFERENCES

Bhasha Dictionary | Sinhala-English Dictionary. (n.d.). Retrieved July 14, 2018, from http://www.bhasha.lk/products/dictionary

Buddi dictionary. (n.d.). Retrieved July 14, 2018, from http://dic.budditha.edu.lk/

Comparative review of dictionaries for English learners | Antimoon. (n.d.). Retrieved October 12, 2017, from http://www.antimoon.com/how/learners-dictionaries-review.htm

Derana English Sinhala Dictionary. (n.d.). Retrieved July 14, 2018, from http://sinhala-english-dictionary.blogspot.com/2017/02/derana-english-sinhala-dictionary.html

Dictionary. (2017a). In Wikipedia. Retrieved from https://en.wikipedia.org/w/index.php?title=Dictionary&oldid=803859323

EF English Proficiency Index - Sri Lanka. (2017, September 22). Retrieved September 22, 2017, from //www.ef.com/epi/regions/asia/sri-lanka/

Gunapala Piyasena Malalasekera. (2017). In Wikipedia. Retrieved from https://en.wikipedia.org/ w/index.php?title =Gunapala_ Piyasena_ Malalasekera &oldid=787984048

History. (n.d.). Retrieved July 26, 2018, from http://www.dictionary.gov.lk/index.php?option=com_content&view=article&id=2&Itemid=104&lang=en

Introduction. (n.d.). Retrieved July 26, 2018, from http://www.dictionary.gov.lk/index.php?option=com_content&view=article&id=1&Itemid=103&lang=en

Lexicography. (2017). In Wikipedia. Retrieved from https://en.wikipedia.org/w/index.php?title = Lexicography&oldid= 801282848

Madura English-Sinhala Dictionary. (2017). In Wikipedia. Retrieved from https:// en.wikipedia.org/ w / index. php ?title=Madura _ English - Sinhala _ Dictionary&oldid=788982008

Merriam-Webster's Learner's Dictionary. (n.d.). Retrieved October 12, 2017, from http://learnersdictionary.com/

New Oxford American Dictionary. (2017). In Wikipedia. Retrieved from https://en.wikipedia.org/w/index.php?title=New_Oxford_American_Dictionary&oldid=799828616

Sinhalese language. (2018). In Wikipedia. Retrieved from https://en.wikipedia.org/w/index.php?title=Sinhalese_language&oldid=851575948

StarDict. (2017). In Wikipedia. Retrieved from https://en.wikipedia.org/w/index.php?title=StarDict&oldid=781598448

Wasala, A., & Weerasinghe, R. (2008). EnSiTip: A Tool to Unlock the English Web.