ABSTRACT

Nowadays a lot of people migrate or visit different places around the world without knowing their local dialects. Learning a local language helps a person to communicate and integrate with the community.

Studying a foreign language takes time and dedication. This is one reason why people hire a personal translator to do the job. Human translators are quite expensive and mostly common to international business organizations.

Machine translators are a good alternative for human translators as it is much more affordable and provides the basic feature which is to translate a language.

Machine translation (MT) is the automatic translation from one natural language into another using computer. It is an area of applied research that draws ideas and techniques from linguistics, computer science, Artificial Intelligence (AI), translation theory, and statistics. Machine translations have been tried over the past decades. It has remained a key application in the field of natural language processing (NLP).

Some studies had proven that machine translators are not as accurate in translating languages and yet a lot of people use them for such reasons that (a) the meaning of the translated language is much more important rather than the grammar, (b) people are not willing to pay for translating simple sentences, and (c) machines are available for use at any time.

With the growing numbers of translator using machine languages, the researchers decided to propose English to Kankanaey Translator for those people who want to translate English to Kankanaey language particularly the tourists who are visiting the Cordillera Region. This translator is foreseen to help the tourist understand the language and have a good communication to the cordilleran people. It also helps the tourist to learn and speak the Kankanaey language when they visit Cordillera Region.

This paper presents the theories and concepts that were adopted in the design and implementation of the translation algorithm for English to Kankanaey language using the statistical machine translator and dictionary based algorithm. It also discusses the advantages and disadvantages of the developed translation algorithm to other existing translation algorithms in terms of English to Kankanaey.
1. INTRODUCTION

Visiting other places is exciting as much as learning new language or dialect in that particular place. Learning a local language helps a person to communicate and engage with the community. Studying a foreign language takes time and dedication. This is why people hire a personal translator to do the job. Human translators are quite expensive and mostly common to international business organizations.

Machine translation as an idea has been around for a very long time. Even before, philosophers and poets have dreamed of a unifying principle that could knit the different cultures and languages of the world together. Up to this time, several researchers still continue to do researches related to machine translation.

Machine translation (MT) is the automatic translation from one natural language into another using computer. The term ‘machine translation’ has now established itself as the general accepted name for any system which uses an electronic computer to transform a text in one language into some kind of text in another natural language.

Machine translation software is becoming more important in the modern world due to new markets opening up in non English speaking countries. Here in the Asean countries, languages are fundamentally different to English that a deep understanding of their natural language needs to take place before any kind of translation can occur. Two biggest examples of cultures with a big appetite for technology which require the use of machine translation applications and processes is China and India.

Machine translators are a good alternative for human translators as it is much more affordable and provides the basic feature which is to translate a language. It has been proven that machine translators are not as accurate in translating languages and yet a lot of people use them for such reasons that (a) the meaning of the translated language is much more important rather than the grammar, (b) people are not willing to pay for translating simple sentences, and (c) machines are available for use at any time.

In MT, the translation process involves a two step procedure. First, decoding the source material into a meaningful text and then re-encoding it into the target translation language. As mentioned above, the first attempts used simple substitution while modern applications are a lot more sophisticated. Software development needs to combine linguistic rules and artificial intelligence with the knowledge and experience of a real native speaker of the target language in order to be truly successful.

Some existing applications that are related to language translation include Google Translate, Systran, Bing Translator and Moses.

In the Philippines, Kankanaey is one of the indigenous spoken language. Kankanaey is widely used by people from the Cordillera specifically from the Mountain and Benguet Provinces. They do not have any unique hand writing thus a person writes Kankanaey words in english characters and is spelled as pronounced. Learning the language is difficult since there are no published books intended for learning. A person learns Kankanaey through socialization, observation, and analysis which would take an ample time to master. This would be an obstacle for foreigners visiting the Cordillera region as they have limited resources to prepare themselves to learn the culture and language as well..

The Online Kankanaey-English Translator System would serve as a point of interest although it has limited features and questionable events such as:

1. Is it able to store all the Kankanaey and english words?
2. Is it able to translate phrases?
3. Would it display all the Kankanaey version of all English words? If not, then how can a foreigner further express their thoughts?
4. Language is dynamic wherein new words are being added and meanings are changed. Moreover, a certain word is used in different certain events. What happens when a new word is introduced? What happens when a meaning has been changed? How would an update happen? Is update going to be manual or automatic? Lastly, is there any algorithm that would identify which appropriate word to use in a certain event?
5. What happens when the database of english and kankanaey words have a big amount of figures? Is linear searching algorithm still appropriate?
6. It is possible to write a kankanaey word in different spellings as long as the pronunciation is the same, this would be an issue during searching. How would the system address the problem?

With the above mentioned queries, the researchers came up with developing a better translation algorithm specifically from an english language to a kankanaey language.

2. THEORIES AND CONCEPTS OF THE DEVELOPED TRANSLATION ALGORITHM

During the design and implementation of the translation algorithm for English to Kankanaey language, the researchers adapted the theories and concept of Statistical Machine Translator because the researchers have found out that it will give the accurate meaning for the translated word and it doesn’t need to maintain the dictionary.

2.1 Design and Algorithm for English to Kankanaey Translation

Statistical Machine Translation (SMT) draws from many fundamental research areas in computer science, so some knowledge of automata theory, formal languages, search, and data structures is beneficial. Statistical machine translation (SMT) is an approach to MT that is characterized by the use of machine learning methods. Statistical MT models take the view that every sentence in the target language is a translation of the source language sentence with some probability. The best translation, of course, is the sentence that has the highest probability. The key problems in statistical MT are: estimating the probability of a translation, and efficiently finding the sentence with the highest probability. This concept was adopted to develop such project.

In designing the Statistical Machine Translator, the researchers subdivided the algorithm into three phases. In the first phase of the algorithm, it reads all the Bilingual Corpora that are stored in a text file. The bilingual corpora consist of English...
phrases and its equivalent Kankanaey phrases as illustrated in figure 1.

After reading the bilingual corpora, the algorithm breaks down the English phrase into single words and gathered all the possible Kankanaey phrases as assumed to be its equivalent. After gathering the Kankanaey phrases, each English word is assigned on each of the word on a Kankanaey word and makes as its equivalent and store the result on a text file. The figure below shows the flowchart for the bilingual corpora of the first algorithm.

The second algorithm reads the text file produced from the first algorithm and counts the duplicate word and its equivalent then prints the statistical count on how many times the word has been encountered. This approach used an algorithm known as bilingual word alignment algorithm. A bilingual word alignment algorithm as represented in figure 3, finds the corresponding word-level translation between the source and the target language sentences. It is widely used in the area of natural language processing, like machine translation, cross-language information retrieval, and bilingual dictionary compilation. In this paper, a new algorithm is proposed. By adopting a multilayered matching and disambiguation strategy, the bilingual word-alignment task was transformed to an iterative solution of anchor word-pair.

Figure 1: The Bilingual Corpora of the English to Kankanaey Translator

After reading the bilingual corpora, the algorithm breaks down the English phrase into single words and gathered all the possible Kankanaey phrases as assumed to be its equivalent. After gathering the Kankanaey phrases, each English word is assigned on each of the word on a Kankanaey word and makes as its equivalent and store the result on a text file. The figure below shows the flowchart for the bilingual corpora of the first algorithm.

Figure 2: 1st Algorithm Flowchart

Figure 4: 2nd Algorithm Flowchart
The third algorithm is the input phase where the user will enter an English word or phrase to be translated. The user input will search the entered word on the text file produced by the second algorithm and search the highest frequency as an output.

3. THE SYSTEM

Following the given algorithm above, the first part of the algorithm creates a dictionary using the bilingual corpora. See figure 5.

Figure 5: Screenshot for the 1st algorithm

Figure 6 shows the dictionary that has been created and stored in the first algorithm.

Figure 6: Dictionary Created from the 1st Algorithm

To input any English word/phrase to be translated into Kankanaey word/phrase, refer to figure 7. This is the application of the second algorithm where the inputted value is being processed. It gathers its resources in the dictionary and counts the frequency of the similar words of the inputted value and gets the highest frequency as the output as shown in figure 8.

Figure 7: Input Phase

Upon gathering the highest frequency it gives you the output as shown in figure 9.

Figure 8: Frequency Count

Figure 9: General Output

Upon gathering the highest frequency it gives you the output as shown in figure 9.
4. ADVANTAGES AND DISADVANTAGES OF THE DEVELOPED TRANSLATION ALGORITHM

4.1 Advantages of English to Kankanaey Translator

The advantages of the developed translation algorithm to other existing translation algorithms in terms of English to Kankanaey:

Accurate meaning. The algorithm provides the correct meaning of the translated word.

Example:
- English word/phrase: I went to market
  Kankanaey translation: Saken inmey en palengke

Easily adapts modern language. The algorithm does not care whether the input is old or modern language because it can still provide the right meaning of the translated language.

Example:
- Word ‘Gay’;
  Old meaning refers to feelings of being carefree, happy, bright and showy.
  Modern meaning refers to the people who are homosexuality.

No maintenance for dictionary. The algorithm does not need to maintain the dictionary because the dictionary will depend on the bilingual corpora. The dictionary does not need to be update.

Spelling doesn’t matter. The algorithm ignores the spelling whether it is correct or not because it will still be read as it is with the same meaning.

4.2 Disadvantages of the English to Kankanaey Translator

The disadvantages of the developed translation algorithm to other existing translation algorithms in terms of English to Kankanaey include the following:

It cannot fix grammar. The algorithm does not fix the grammar because it translates the entered value word per word.

Example:
- English word/phrase: Its ours.
  Kankanaey translation: Sitosa tako kuwa.
  Correct translation: Kuwa tako sitosa.

Some Kankanaey is limited to word per word. Kankanaey word is limited and there are kankanaey words that have been used many times but it depends in the usage.

More bilingual corpora are needed. Bilingual corpora are really needed to make the translation more accurate and more reliable.

5. CONCLUSIONS AND RECOMMENDATION

5.1 Conclusions

Designing and developing a language machine translator is a great challenge for every researcher. It needs a lot of learning process before you can do such thing. Thus, a detailed background of machine translation is really needed.

Generally, the output of this research is a translation application, where the system accepts an English word or phrase to be translated in Kankanaey. With this system, anybody could learn and understand one of the dialects of the Cordilleran people. This could be a great advantage in bridging the gap of language barrier especially for tourist in the Cordillera.

5.2 Recommendations

For further development of the study the researchers recommended the following.

1. This study served as a framework to the University of the Cordilleras for further study.
2. The future researchers should improve the design interface of the system using graphical user interface. And to add features which were not incorporated in the English-Kankanaey Translator Algorithm.
3. Moreover, add more bilingual corpora and algorithm word arrangement for the translated Kankanaey phase to rearrange the word and correct the grammar to make it more accurate for the translation process.

6. ACKNOWLEDGEMENT

The researchers would like to extend their gratitude and great appreciation to all the people who have helped and encouraged them to finish and fulfill their research paper:

To the Heavenly Father who guided the researchers throughout this study, who gave them the courage, determination, patience and strength in every decision they made.
This study will not be possible without the undying support of their parents especially on their financial and emotional needs. Without them, the researchers would never be able to manage the thesis expenses.

To their panelists and adviser, for being a great part of their study.
REFERENCES