A Bicolano-to-Tagalog Transfer-Based Machine Translation System

Ria Ambrocio Sagum

Abstract—The Bicolano-Tagalog Transfer-based Machine Translation System is a unidirectional machine translator for languages Bicolano and Tagalog. The transfer-based approach is divided into three phase: Pre-Processing Analysis, Morphological Transfer, and Sentence Generation. The system analyze first the source language (Bicolano) input to create some internal representation. This includes the tokenizer, stemmer, POS tag and parser. Through transfer rules, it then typically manipulates this internal representation to transfer parsed source language syntactic structure into target language syntactic structure. Finally, the system generates Tagalog sentence from own morphological and syntactic information. Each phase will undergo training and evaluation test for the competence of end-results. Overall performance shows a 71.71% accuracy rate.

Keywords: Machine Translation, Transfer-based, POS tagging, Morphological Transfer, language model, Language Translation.

I. INTRODUCTION

Philippines is considered as one of the countries with rich culture when it comes to language. The reason behind this is because it is composed of many islands and regions that allows the people in this country to speak different dialects that only locals can communicate with [1].

To date Philippines have listed 187 languages. Of these, four are already extinct and 25 are endangered to be extinct. The reason behind this can be the non-usage of these languages by their natives since instead of teaching their children to speak their native language they teach them a language that is commonly used by many to easily communicate with others [2].

With this, researchers focus on exploring dialects and device solutions to allow a clear communication amongst Filipinos. The rich language diversity of the Philippines is interesting and should be focused on by the researchers in the field of computing. Using these dialects as the domain of Machine Translation (MT) techniques may help us gain more insights on language structure, discover language issues that have been addressed and come up with a solution that may contribute to MT in general.

Different issues arise when developing a translator. [3][4] One main problem with machine translation (MT) for different dialects is the difficulty in acquiring dictionaries and grammar rules. So is the determination of file system to be used [5]. Lack of structure rules and the long amount of time it will take to acquire knowledge about a language. This issue can be attributed to the reason why MT researchers opted to use statistical based MT systems.

In the development of MT system for dialects, a researcher needs to consider different strategy [6][7]: whether translation from source language to target language takes place in a single stage (direct translation), two stages (via an 'interlingua'), or via the 'transfer' approach, where the translation proceeds in the three stages [8][9][10].

The use of transfer-based approach does not require large amount of word aligned data for translation that is not easily available for Philippine dialects. In transfer-based machine translation, the text input is transformed into an abstract representation, and then its equivalent in the target language is generated using bilingual dictionaries and grammar rules [11].

Morphology is important area in translation. Different research projects on morphological analysis and stemming present in the field are the proofs of its importance [12]. The most spoken dialect of the Philippines, Tagalog, there are different studies in analyzing its morphology. A study called TAGMA (Tagalog Morphological Analyzer), is a morphological analyzer, based on the Optimal Theory and level two morphology that handles Tagalog verbs. [13]. For Tagalog dialect a research called TagSa, a Tagalog Stemming Algorithm, was developed for all forms of Tagalog words. It can be used specifically for morphological analysis to derive roots [14]. A good accuracy of these tools can lead to a better translation accuracy.

There are various means for evaluating the output quality of machine translation systems. The oldest is the use of human judges to assess a translation's quality [15]. Even though human evaluation is time-consuming, it is still the most reliable method to compare different systems such as rule-based and statistical systems [16]. Some researchers use BLEU [17] although some researchers still use the quality testing thru human evaluation.

This study evaluated different morphological and transfer approaches to determine its appropriateness to Tagalog and Bicolano translation. In this study translation will focus on building morphological tools for analysis of the source and translation tools to be used in the translation to the target language. As a follow up in the previous unpublished paper, it will focus on the different tools that will be used starting from the lexical analysis, syntax and semantic analysis until

Revised Version Manuscript Received on August 19, 2019.

RiaAmbrocioSagum, Department of Computer Science,College of Computer and Information Sciences, Polytechnic University of the Philipines, Manila, Philippines.

the translation phase. The source language is Bicolano and target language will be Tagalog.

II. RELATED WORKS

Machine Translation (MT) is defined as "translation from one natural language (source language (SL)) to another language (target language (TL)) using computerized systems and, with or without human assistance" [18]. Different applications can benefit in translations [19][20], thus, NLP researchers devote their time towards creation of a good translator. These were done in application to different language since each language has their own unique structure.

In 2005, a Hindi to English translation system was develop that focused on designing a system that translate the document from Hindi to English using transfer-based approach. This system takes an input text check its structure through parsing. Reordering rules are used to generate the text in target language. It is better than Corpus Based MTS because Corpus Based MTS require large amount of word aligned data for translation that is not available for many languages while Transfer Based MTS requires only knowledge of both the languages (source language and target language) to make transfer rules [21]. Another research in this language translation was developed and experimented on solving the word ambiguities to get a high accuracy in the translation [22]. This study also uses a parsing technique for the development of their system. It is known that a good parsing technique is helpful in the development of a translator [23][24][25].

Another study on MT focuses on Telugu and Tamil, major Dravidian languages with rich literary tradition sharing indubitable linguistic similarities and dissimilarities were created. An MT between them may be viewed as a bridge to understand and share the richness of both the languages. This study allows the researchers to look for the features needed to consider in a rich language since Tagalog belongs to this classification. The Telugu-Tamil MT system is an assembly of various linguistic modules run on specific engines whose output is sequentially maneuverer and modified by a series of modules till the output is generated. It employs three stage architecture: Stage 1: Source language analysis; Stage 2: Source language to target language transfer; Stage 3: Target language generation [26].

CARLA: (Computer Assisted Related Language Adaptation) is a system that allows the user to write linguistic rules to do automated morphological parsing and then transfers the text morpheme by morpheme to produce a rough draft of the input text in related language. It works one sentence at a time. CARLA gives very literal translation from SL to TL [27]. The effect of using linguistic rules to produce translation of a language to its target language is being highlighted in this study. Linguistic rules can be written to help in translation. The use of multitape automata may be implemented this [28].

The current state of Philippine linguistic resources, which includes formal grammars, electronic dictionaries and corpora are not yet significant to address industrial- strength language technologies. According to the research, Language Formalism for Multi-Lingual Machine Translation, the use of computational approach can be used in automatically estimating constituent structures from a corpus. An example of this is the unsupervised probabilistic approaches [29].

Their method in creating their translation was done with a whole engine that includes the analysis to f- structure, transfer of source to target f-structure, and generation from f-structure. Initial linguistic resources were established to test the engine and to develop the full bidirectional English-Filipino machine translator system. These linguistic resources include the formal grammar rules for the English and Filipino language, mono-lingual dictionaries for both languages and the transfer dictionaries, which include transfer rules (structural level) and transfer dictionary (word level). Testing involved subjecting the system to different sentences and sentence constructions in both languages (English and Filipino). Results show that translation quality is extremely dependent on the available linguistic resources [29]. This result shows that translation from Tagalog to English is possible.

As stated about the demographic description of the Philippines, this causes the country to have many dialects and the communication sometimes become a problem. Researches in the field of Natural Language Processing (NLP) are trying to solve this by developing a translator that can be used not only for English to Tagalog but Tagalog to different dialects [30][31][32]. A unidirectional machine translator for languages Tagalog and Cebuano, dialect translator was studied and developed years ago [8][33]. This research of Yara made used of morphological analysis that is based on TagSA (Tagalog Stemming Algorithm) and is focused on an affix correspondence-based POS (part-of-speech) tagger. The rules used in morphological synthesis are reverse of the rules used in morphological analysis. A bilingual dictionary from Tagalog to Cebuano has been developed and is used by the different components of the system. The machine translator has been evaluated, with the Book of Genesis as input, using GTM (General Text Matcher). Result of the evaluation gives a score of good performance 0.8027 or 80.27% precision and 0.7992 or 79.92% recall [33].

This study will take the different morphological and transfer approaches and determine its appropriateness to Tagalog and Bicolano. A system will be developed and be evaluated to test the quality and correctness of the translation. Specifically, translation will focus on building morphological information during analysis and Tagalog word generation will use the information extracted from analysis phase.

III. METHODOLOGY

The developed system uses the transfer-based approach in the translation it went through different modules which was presented in this section.

Transfer-Based Approach

The system will first analyze the source language (Bicolano) input to create some internal representation. Transfer and generation phase will process word per word translation. In transfer phase, each source language

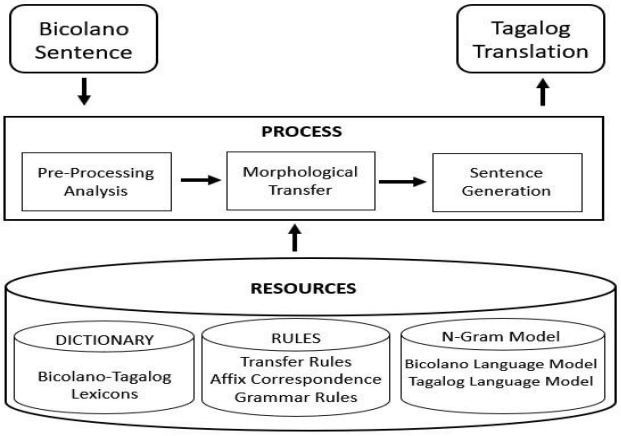
(Bicolano) morpheme is being translated. The output of the translation will be used in morpheme
combination with the most word count in the trained Tagalog corpus, which will be the final Tagalog translation. Figure 1 shows the architectural design of the BicolanoTo Tagalog Transfer Based Machine Translation System.

The Processing Phase consists of three parts (Fig. 1): Pre-processing Analysis, Morphological Transfer, and Sentence Generation. The following architecture shows the detailed part of each processes.

Pre-Processing Analysis

The pre-processing analysis has three parts: Once the input was entered in the system then sentence splitting will be done, after which tokenizetion takes place. The system will check the structure of the input and make necessary changes to make sure that the input is correctly spelled and grammatically correct. This phase is called normalization. Stemming is being use to get the root word of a lexeme in POS tagging. (See Fig. 2)

Fig.1System Architecture of Bicolano-Tagalog



Transfer-based Machine Translation

Bicolano-Tagalog Lexicon

The Bicolano-Tagalog Lexicons were built to be used as its dictionary. This took some time since the resources for Bicolano dialect is limitted. For this study to build a dictionary for the translation each word in Bicolano language was translated manually to its English equivalent then using English-Tagalog translations was done, the part of speech was also manually done to be used as look up dictionary.

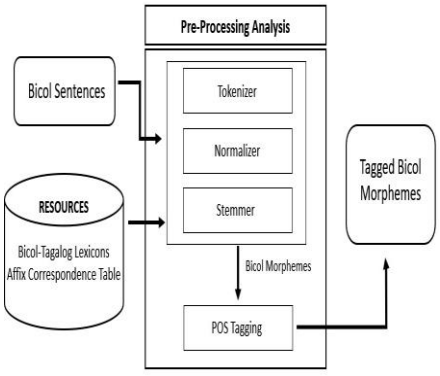


Fig. 2 Pre-Processing

Stemming

This process (Fig.3) will output the stemmed word from the tokenization stage. The stemmer process includes the ff: prefix, infix and suffix stemming. It includes the removal of derivational affixes, to get the base word or stem.

The researcher used Affix-Stemmer for stemming. As it receives the input word/s it will stem possible affix then it will check if the word is registered in the dictionary if no, it will again stem possible affix and then check the dictionary the process will continue until the root word/stem is found. During the development of the system n-gram was used but it performance is not acceptable as when Affix-Stemmer was used.

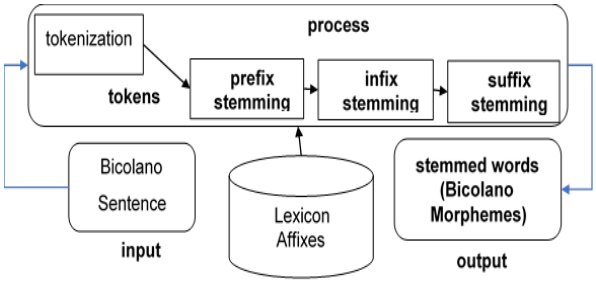


Fig. 3 Tokenization and Stemming

POS Tagger

In the previous work the POS tagger yields low accuracy. That is why the researcher opted to use another algorithm to solve the problem and come accross with using Averaged Perceptron tagger. This tagger (Fig. 4) uses a set of features and weights, to predict the specific tag for each word. These features are suffixes, prefixes, previous and next tags of the next and previous words, and the word itself. On the other hand, weights are used to balance the correct and wrong prediction.

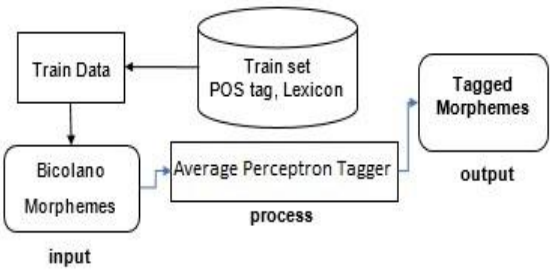


Fig. 4 Part-Of-Speech Tagging

Normalizer

Normalizer includes: vowel removal, phonetic style, repetition of characters, accent style and group repeating units. The module is trained and test using Bicolano sentences. Shown on figure 5 is the architecture for the normalizer.

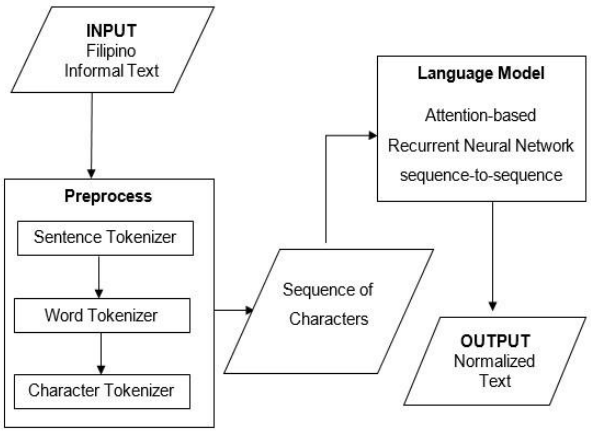


Fig. 5 Sentence Normalizer

When several ambiguous words occur together, the possibilities multiply. However, it is easy to enumerate every combination and to assign a relative probability to each one, by multiplying together the probability of each choice in turn. The combination with highest probability is then chosen. A European group developed CLAWS[34] a tagging program that did exactly this, and achieved accuracy in the 93-95% range.

Figure 6 shows the second phase which is the *Morphological Transfer*, input to transfer is the tagged words and then stream of morphemes to transfer to the target language. The program will iterate over every morpheme and evaluate its corresponding tag (Ex. RW - Root Word). It will do whatever is appropriate. For root and free morphemes, a direct transfer is done where in the morpheme is searched in the dictionary. If there are multiple morphemes available of the same tag, the program chooses the word with the most frequency using the unigram model.

In the case of affixes, an affix rules table is used to get the target language equivalent of a source affix. To make sure that the affixes used for the stem follows the rules, the POS tags of a word to which each an affix may apply is stated. If the affix is not found on the table, it is already implied to use the same affix for target language. These two dialects are very similar so we kept it simple.

If there are multiple affixes applicable to a word, the program generates all possible words then choose the one mostly used. Of course, if there are proper nouns, punctuation marks, and anything where transfer is unnecessary, the program doesn't go on and returns the same output as input then wait for next word to translate.

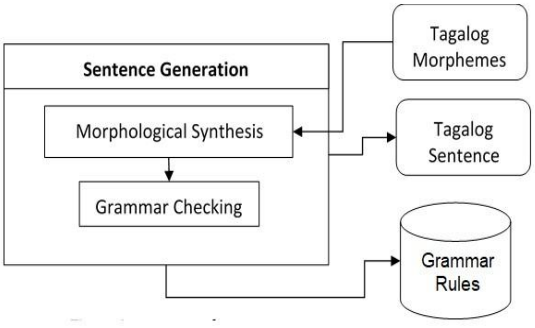


Fig. 6 Morphological Transfer

Lastly, if there are words not in the dictionary, the program checks first if it might actually already be in target language. If not, we use a bigram model for prediction. For every unknown word, the program takes the preceding word and checks if there are bigrams of it. If there are, the code proceeds to sentence combinations. Each possible word from the bigrams will replace the unknown word in the sentence. It then calculates the overall probability of the sentence generated. After all the iterations, the sentence combination with the highest probability is chosen and the unknown word is predicted. The output of this phase will be the final list of the words translated (in order of where it appeared in the sentence) and the hash map for the POS tags used in Sentence.

Language Model

The bi-gram models were used in word prediction this phase. The Book of Genesis Tagalog Literature Domain were used as sentence corpus for the model. LM is also used in POS tagging under pre-processing phase.

Sentence Generation

The last phase of the process is the *Sentence Generation* as shown in Figure 7. This phase combine into sentence the list of words translated by the transfer phase corresponds to the order of words in the source sentence. The formed sentence is the Tagalog equivalent translation of the input Bicolano Sentence. The system will then check the grammar using Rule-based Approach, of the output sentence based on Tagalog grammar rules.

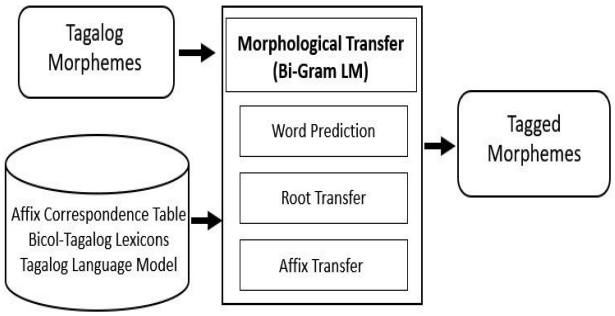


Fig. 7 Sentence Generation Architecture

Grammar Checker

The input will be the tagalog words with its corresponding POS tags, then it will be put in the list of tuples – these are the built in type of python language which are written in square brackets [] for list and open and close parenthesis for tuples.

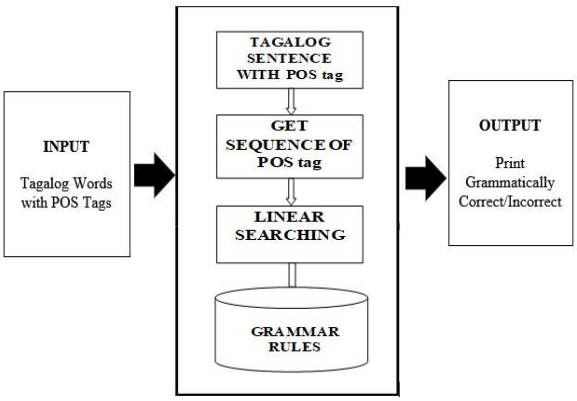


Fig. 8 Grammar Checker Architecture

Sentence level is the process of the grammar checker. It will get one sentence or equivalent to one list of tuples [("word","POS tag"), … , ("word","POS tag")]. We have 9 POS tags in our modules; noun as PANGNGALAN, pronoun as PANGHALIP, verb as PANDIWA, adjective as PANG-URI, adverb as PANG –ABAY, determiner as PANTUKOY, preposition as PANG – ANGKOP, linking verb or lexical marker as PANGAWING and punctuation as BANTAS. Next, it will filter out to get the sequence of POS tags. The sequence of POS tags will be checked if there are two or more consecutive POS tags in the sequence, if there are consecutive

POS tags it will be reduce or group or one POS tags, for example a list of tuples of a sequence POS tags:

[('PANG-ABAY'), ('PANDIWA'), ('PANTUKOY'), ('PANTUKOY'), ('PANGNGALAN'), ('PANGNGALAN'), ('PANG-UGNAY'), ('PANGNGALAN'), ('PANGHALIP'), ('PANGHALIP')]

Will reduce into

[('PANG-ABAY'), ('PANDIWA'), ('PANTUKOY'), ('PANGNGALAN'), ('PANG-UGNAY'), ('PANGNGALAN'), ('PANGHALIP'), ('PANGHALIP')]

After that the new sequence of POS tag will be checked if the sequence is grammatically correct by using the grammar rules in a text file. The grammar rules are consisting of 3000+ grammar set. The module implemented linear search algorithm to check and determine if the sequence of POS tags of the input is match in one of the grammar set in the grammar rules.

The output will print out the sentence with POS tags, the sequence of POS tags, and if it is grammatically correct or incorrect. If it is grammatically correct it will store in a file.

Evaluation Methodology Tool

The following formula were used to compute for the accuracy of each module of proposed system. The accuracy for each module were computed to know what will be the consequence once a particular module poorly performs. The f-measure (Eq. 1) serves as the basis for acuracy of the modules of the system it is done by computing the true positive, true negative, false positive, false negative. Where: tp is true positive (Correct Conversion); tn is true negative (Correct Absence of Result); fp is false positive (Incorrect Conversion); and fn is false negative (Unrecognized Words).

$$AR = \frac{tp+tn}{tp+tn+fp+fn} \qquad \text{Eq. 1}$$

On the other hand , the system for unidirectional translation of Bicol-Tagalog was evaluated using the BLEU Score. Bilingual Evaluation Understudy or BLEU is an algorithm for evaluating the quality of text which has been machine-translated from one natural language to another. Quality is considered to be the correspondence between a machine's output and that of a human: "the closer a machine translation is to a professional human translation, the better it is" – this is the central idea behind BLEU. BLEU was one of the first metrics to claim a high correlation with human judgements of quality, and remains one of the most popular automated and inexpensive metrics. Scores are calculated for individual translated segments—generally sentences—by comparing them with a set of good quality reference translations. Those scores are then averaged over the whole corpus to reach an estimate of the translation's overall quality. Intelligibility or grammatical correctness are not taken into account.

IV. DISCUSSION OF THE RESULTS

The Bicolano-Tagalog MT is evaluated by phase in terms of accuracy rate. Each phase had undergone training and accuracy testing. Below are presented results from the series of accuracy tests for evaluation of system.

Using series of testing and training of number of sentences, the tables summarizes the accuracy evaluation results of each module.

Table 1 Accuracy Evaluation for each Module: Stemming, POS Tagging and Grammar Checker

Module	Total Input	Accuracy (%)	Error Rate (%)
Normalizer	1691	91	9
Stemming	340	85.33	14.67
POS Tagging	830	80.72	19.28

Table 1 shows the summary of accuracy evaluation results for the Normalizer, Stemming and POS Tagging. Out of 1691 words in 100 sentences, the accuracy rate of the normalizer is 91 with an error rate of 9. The stemming has 340 words with an accuracy rate of 85.33 and an error rate of 14.67. Lastly, the POS tagging has an accuracy rate of 80.72 words with an input of 830 words and an error rate of 19.28. The following modules show improvements from the previous study [1], where the accuracy for stemming and POS tagging were 78.85 and 79.02 respectively. The increase in the accuracy rate is possible with the new approach used for the POS tagging and new rules were added for the Affix Correspondence Table and Grammar Checker. New words and translations were also added to the Bicolano-Tagalog Lexicons translated by the Bicolano experts.

Table 2 Evaluation for the Transfer Module

Module	Total Input	Accuracy (%)	Error Rate (%)
Transfer	358	95.20	4.80
Grammar Checker	100	92.38	7.62

Table 2 shows the summary of the evaluation for the transfer module. The transfer module has an accuracy rate of 95.20 based on 358 words input and an error rate of 4.80 while grammar checker has an accuracy rate of 92. 38 based on 100 words input and an error rate of 7.62.

Table 3 Evaluation of the Bicolano-Tagalog translator

Module	Total Input	BLEU
Integrated	100	.717

Table 3 shows the evaluation for the input in Bicolano-Tagalog Translator. The transfer-based machine translation system for Bicolano-Tagalog Sentences was evaluated using the Bilingual Evaluation Understudy or BLEU. This is an algorithm for evaluating the quality of text which has been machine-translated from one natural language to another. The system was able to obtain a 0.717 BLEU score. The scores are calculated for individual translated segments—generally sentences—by comparing them with a set of good quality reference translations. Those scores are then averaged over the whole corpus to reach an estimate of the translation's overall quality. Intelligibility or grammatical correctness are not taken into account. Each module is dependent to one another. The output from the stemmer will be fed to the POS Tagger and so on and so forth. So, the translated sentence or the final output will be based to each modules output.

V. CONCLUSION AND RECOMMENDATION

The researcher was able to develop tools needed for translating languages and integrated it to build a unidirectional Machine Translator for two dialects from Philippine Language, Bicolano and Tagalog using the transfer-based approach that generate sensible results. The system was able to obtain an average 0.717 BLEU score out of 100 Bicolano sentences transferred to Filipino Sentences. However, the study shows that MT systems output translation of the source language was greatly affected by the resources the system used e.g. lexicon, rules, tables and corpus as well as lack of training and testing data can lead to imprecise evaluation. And also, each module affects the final output or the translated sentences. If the output of the stemmer from the start is accurate, the accuracy of the final output will also be affected thus the final sentence will be accurately translated.

It is recommended to further improve the lexicon/dictionary by acquiring more resources for Bicolano Dialect and for more accurate translation of the system, provide more training and testing data. Furthermore, focus on the enhancement of the algorithm and model implementation. For future works, the proponents aim to integrate this system to another MT to develop a bi-directional MT for Bicolano and Tagalog Languages.

VI. ACKNOWLEDGMENT

The researcher would like to acknowledge the help extended by BICTAG group BSCS 4-3 batch 2018.

REFERENCES

1. C. K. Cheng and S. L. See, "The Revised Wordframe Model for the Filipino Language," Journal of Research in Science, Computing and Engineering (JRSCE), vol. 3, no. 2, pp. 17-23, August 2006.
2. R.Raga , Reflections on the Awareness and Progress of Natural Language Processing (NLP) Research in the Philippines. Philippine Computing Journal, 11(1):1-9.
3. G. Stankevičiūtė, R. Kasperavičienė and J. Horbačauskienė, "Issues in Machine Translation: A case of mobile apps in the Lithuanian and English language pair," pp. 75-88, 2017.
4. M. D. Okpor, "Machine Translation Approaches: Issues and Challenge," IJCSI International Journal of Computer Science Issues, vol. 11, no. 5, pp. 159-165, September 2014.
5. A. Magidow, "A relational database model and prototype for storing diverse discrete linguistic data," pp. 27-45, 2015.
6. S. K. Jha, P. P. Singh and V. K. Kaul, "International Journal of Advanced Research and Development," Phrased based T2 model: A review of Google translate, Bing translator &Anusaaraka, vol. 2, no. 6, pp. 407-411, November 2017.
7. A. Godase and S. Govilkar, "MACHINE TRANSLATION DEVELOPMENT FOR INDIAN LANGUAGES AND ITS APPROACHES," International Journal on Natural Language Computing (IJNLC), vol. 4, no. 2, pp. 55-74, April 2015.
8. J. Fat, T2CMT: Tagalog-to-Cebuano Machine Translation. Masters Thesis DLSU
9. A. Ballabh and D. U. C. Jaiswal, "A STUDY OF MACHINE TRANSLATION METHODS AND THEIR CHALLENGES," International Journal of Advance Research In Science And Engineering, vol. 4, no. 01, April 2015.
10. D. Chiang, "Hierarchical Phrase-Based Translation," Association for Computational Linguistics, vol. 33, 2007.
11. A. Borra, A transfer-based analysis Engine for an English to Filipino Machine Translation Software. Manila: University of the Philippines Los Banos, MS Thesis, 1999.
12. C. Jordan, J. E. Mason, J. Healy, V. Keselj and C. Watters, "Swordfish2: Using Kernel Density Estimation to Smooth N-gram Histograms for Morphological Analysis," Journal of Interesting Negative Results in NLP and ML, pp. 1-18, 2008.
13. F. Fortes, A Constraint-based Morphological Analyzer for Concatenative and Non-Concatinative Morphology of Tagalog Verbs, Manila: De La Salle University, MS Thesis, 2002.
14. D. Bonus, A Stemming Algorithm for Tagalog Words, Manila: De La Salle University, MS Thesis, 2003.
15. Morphologic.hu, Comparison of MT systems by Human Evaluation, May 2008, 12 06 2012. [Online].
16. D. D. Anderson, Machine Translation as a tool in a second language learning, CALICO Journal, vol. 1, no. 13, pp. 68-96, 1995.
17. S. M. Mohammad, M. Salameh and S. Kiritchenko, "How Translation Alters Sentiment," Journal of Artificial Intelligence Research, pp. 96-130, 2016.
18. M. A. Chéragui, "Theoretical Overview of Machine translation," Proceedings ICWIT, pp. 160-169, 2012.
19. S. Yao, "Application of Computer-aided Translation in English Teaching," vol. 12, no. 8, 2017.
20. T. Vidhayasai, S. Keyuravong and T. Bunsom, "Investigating the Use of Google Translate in "Terms and Conditions" in an Airline's Official Website: Errors and Implications," vol. 49, January-June 2015.
21. Gehlot, Sharma, Singh and Kumar, Hindi to English Transfer Based Machine Translation System, 2005.
22. S. Mall and U. C. Jaiswal, "Word sense disambiguation in Hindi applied to Hindi-English machine translation," COMPUTER MODELLING & NEW TECHNOLOGIES 2017, vol. 21, no. 2, pp. 56-68, 2017.
23. T. Denkinger, "Chomsky-Schützenberger parsing for weighted multiple context-free languages," Journal of Language Modelling, vol. 5, no. 1, pp. 3-55, 2017.

24. M.-J. Nederhof and K. Sima'an, "Parsing and finite-state technologies, introduction to the special issue," Journal of Language Modelling, vol. 4, no. 1, pp. 1-2, 2016.
25. S. Mall and U. C. Jaiswal, "Shallow Parsing and Word Sense Disambiguation Used for Machine Translation from Hindi to English Languages," International Journal of Intelligent Engineering and Systems, vol. 10, no. 3, pp. 381-390, 2017.
26. K. Parameswari, Development of Telugu-Tamil Transfer-Based Machine Translation system, no. With Special reference to Divergence Index.
27. S. White and R. Stone, Introduction to CARLA STUDIO for Philippine Languages. Document version 0.9. (2004)
28. M. Hulden, "Rewrite rule grammars with multitape automata," Journal of Language Modelling, vol. 5, no. 1, pp. 107-130, 2017 2017.
29. R. Roxas, E.Devillers and R. Giganto, Language Formalism for Multi-Lingual Machine Translation of Philippine Dialects. De La Salle University, Manila, 2000.
30. J. Bautista, C. Bayla, K. Fianza, D. Mamis, J. Tangangco, J. Yango, and D. Miguel. Bi-directional Ilocano-English Language Translator Using Customized Moses Statistical Machine Translation System (SMTS). In Proceedings of the 11th Natural Language Processing Research Symposium, pages 18-25, Manila, Philippines, 2009.
31. A.N. Lazaro, N. Oco, R.E. Roxas. Developing a Bi-Directional Ilocano-Enlish Translator for the Travel Domain: Using Domain Adaption Tecniques on Religious Parallel Corpora. Presented at the 11th Inernational Conference of the Asian Association for Lexicography, Guangzhou, China, 2017.
32. D.G. Macabante, J.C. Tambanillo, A. Dela Cruz, N. Ellema, M. Octaviano, R. Rodriguez and R.E. Roxas. Bi-directional English-Hiligaynon statistical machine translation. In Proceedings of TENCON 2017 – 2017 IEEE Region 10 Confernce, pages 2852 – 2853, Penang, Malaysia, 2017.
33. J. Yara, A Tagalog-to-Cebuano Affix-Transfer-based Machine Translator, no. De La Salle University, 2007.
34. G. Leech, R. Garside, and M. Bryant. CLAWS4: The tagging of the Bristish National Corpus. In Proceedings of the 15th International Conference on Computational Linguistics (COLING 94) Kyoto, Japan. Pp. 622-628, 1994.

AUTHOR PROFILE

RiaAmbrocioSagum,was born in Laguna, Philippines on August 31, 1969. She took up Bachelor of Computer Data Processing Management from the Polytechnic University of the Philippines and Professional Education at the EulogioAmang Rodriguez Institute of Science and Technology. She received her master's degree in Computer Science from the De La Salle University in 2012. She is pursuing her postgraduate studies and is taking Doctorate in Information Technology at De La Salle University. She is an Associate Professor and Research Coordinator at the Department of Computer Science, College of Computer and Information Sciences, Polytechnic University of the Philippines in Sta. Mesa, Manila. Her specialization is in the field of Natural Language Processing. Ms. Sagum has been a presenter at different conferences both in International and National level. She is a member of different professional associations including ACM-CSTA and a board member of the Computing Society of the Philippines- Natural Language Processing Special Interest Group.