ENGLISH-MYANMAR BIDIRECTIONAL MACHINE TRANSLATION SYSTEM

YIN YIN WIN

University of Computer Studies, Mandalay, Myanmar
E-mail: yinyinwin.mdy@gmail.com

Abstract- This paper presents English-Myanmar bidirectional text to text machine translation. Language is very important to communicate with each other in the world. One obstacle of the communication is language problem. Language translation system is easy to learn and understand for Myanmar users and foreigners. This system is implemented for English-Myanmar translation based on rule based machine translation (RBMT) approach. The objectives of the system are to translate Myanmar to English and vice versa, to support Myanmar users and foreigners during conversation with each other, and to provide an efficient language translator. The system applies tree to tree transformation for rule based translation using Synchronous Context Free Grammar (SCFG) rules. It generates a target sentence through target parse tree. This system can be used in language communication for Myanmar and English.

Keywords- RBMT, Tree to Tree Transformation

I. INTRODUCTION

The term Machine Translation (MT) is a standard name for computerized systems responsible for the production of translations from one natural language into another with or without human assistance. It is a sub-field of computational linguistics that investigates the use of computer software to translate text or speech from one natural language to another. MT systems can be designed either specifically for two particular languages, called a bilingual system, or for more than a single pair of languages, called a multilingual system. A bilingual system may be either unidirectional, from one Source Language (SL) into one Target Language (TL), or may be bidirectional. Multilingual systems are usually designed to be bidirectional, but most bilingual systems are unidirectional. However, this system serves as bidirectional MT system between Myanmar and English language. Researchers have proposed different paradigms for machine translation from Myanmar to English or from English to Myanmar. However, English-Myanmar bidirectional machine translation System is not proposed yet. Therefore this system is proposed in this paper. There are three approaches in machine translation system: statistical machine translation, example-based machine translation and rule-based machine translation. This System based on rule-based approach using Synchronous Context Free Grammar rules.

The reason of choosing rule-based approach does not require more memory storage consumption because no need to save massive parallel corpus. Moreover, by using synchronous context free grammar rules, it can be fast processing speed during translation time. The remaining parts of the paper are organized as follows: the works concerning machine translation system are presented in section 2, theory background of machine translation is discussed in section 3, overall proposed system of bidirectional translation system is described in section 4 and section 5 concludes the paper.

II. RELATED WORKS

Language is very important part of the communication. There are many different languages spoken in this world among which English is the global language. The most of the information is available in English. Mr.Uday C. Patkar et.al introduced mechanism which converts multi sentences, question sentences of English to Sanskrit text to speech conversion. They stated that the model consists of array of translation rules to translate from source to target sentence, which is the frame of Rule based Machine Translation System.

Fai Wong et.al described the application of MT based on Constraint Synchronous Grammar (CSG) in devices with limited re-sources. This paper presented the application of Constraint Synchronous Grammar (CSG) formalism to MT for handheld devices.

Shibli Syeed Ashrafi et.al proposed a bi-lingual MT system for Bangla translation of an English simple assertive sentence employing structural analysis using grammatical rule-based approach in the form of context-free grammars (CFGs).

Khaled Shaalan et.al described the development of a novel English-Arabic bi-directional rule-based transfer MT tool in the agriculture domain.

T. T. Zin et al presented Myanmar phrases translation model with morphological analysis. The system was based on statistical approach. In statistical machine

translation, large amount of information was needed to guide the translation process. When small amount of training data was available, morphological analysis was needed especially for morphology rich language. Bayes rule was also used to reformulate the translation probability of phrase pairs. Experiment results showed that proposed system could improve translation quality by applying morphological analysis on Myanmar language.

III. MACHINE TRANSLATION TECHNIQUES

Machine Translation refers to the use of computers to automate some of the tasks or the entire task of translating between human languages. The major machine translation techniques are:
1) Statistical Machine Translation (SMT),
2) Example Based Machine Translation (EBMT) and
3) Rule Based Machine Translation (RBMT).

3.1 Statistical Machine Translation

The statistical machine translation (SMT) is a machine translation paradigm where translations are generated on the basis of statistical models whose parameters are derived from the analysis of bilingual text corpora. The SMT is a corpus based approach, where a massive parallel corpus is required for training the SMT systems.

The SMT systems are built based on two probabilistic models: language model and translation model. The advantage of SMT system is that linguistic knowledge is not required for building them. The difficulty in SMT system is creating massive parallel corpus.

3.2 Example Based Machine Translation

The example based machine translation (EBMT) is the corpus based approach without any statistical models. The example based systems are trained with the parallel corpus of example sentences, similar to SMT systems. The example based systems generally don't learn from the corpus.

They store the parallel corpus and uses matching algorithms to search and retrieve the sentences. The translation memories (TM) are built to aid the human translators by serving as an assisting tool for translation. The advantage of translation memories are easy to implement and linguistic knowledge is not required.

3.3 Rule Based Machine Translation

The rule based machine translation system translates the source text into target text by a set of linguistic rules. Three techniques of machine translation – Direct, Interlingua and Transfer based are applicable to rule based machine translation system. The rule based machine translation system is developed by hand coded rules for translation. The system requires good linguistic knowledge to write the rules and a bilingual dictionary is also needed. Other MT systems like SMT and EBMT requires huge parallel corpus for training. The rule based systems are highly suited for translation of English-Myanmar Languages because the bilingual dictionary could be collected easily compared to parallel corpus and the rules could also be written well with the help of linguists.

The rule based system which has been developed follows the transfer based approach of reordering rules. The drawback of rule based system is that the system is confined with the rules and the rules will evolve with the language over time.

3.3.1 Direct Approach

Direct approach involves in four stages to translate any language to other language. Morphological analysis can be done i.e., identified the tense for the verb then Identify the constituents and Reorder the constituents based on target. Replace the source words to target with the help of dictionary. But direct approach in not a minimal structure and semantic analysis also won't produce a long term solution for MT.

3.3.2 Transfer Approach

Transfer model involves three stages: analysis, transfer and generate. Analyze the source sentence, transfer the structure of source sentence to the structure of target sentence finally translate the word, number, gender in the target words. But in this approach n generating components, n analysis components and n(n-1) transfer components are needed for n language translation, it will increase memory and working principle.

English structural order is SVO (Subject-Verb-Object) whereas Myanmar structural order is SOV. So the system needs to use rules and then parse tree yields the structure of a sentence. On the basis of the structural differences between the source and target language, a transfer approach is used by tree to tree transformation.

3.3.3 Interlingua Approach

The Interlingua approach considers MT as a two stage process: Extracting the meaning of a source language sentence in a language-independent form, and, Generating a target language sentence from the meaning. In this approach, burden on the analysis and generation components increases.

Have to choose between various possible parses for a sentence, identify the universal concepts that the sentence refers to, and understand the relations between various concepts expressed in the sentence.

IV. ENGLISH-MYANMAR BIDIRECTIONAL TRANSLATION SYSTEM

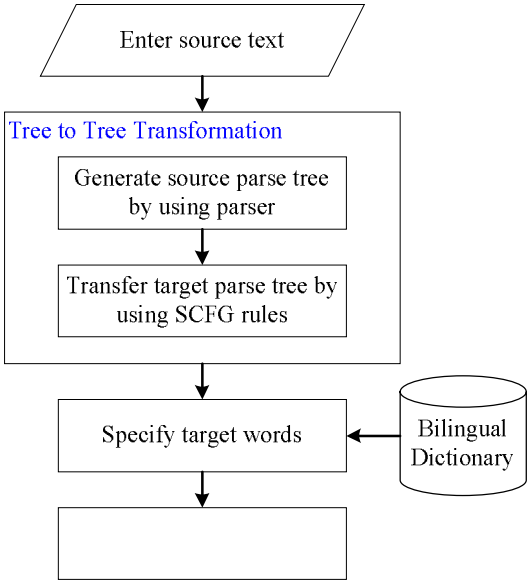


Fig. 1. System Architecture

In this paper, English-Myanmar Bidirectional Translation System is implemented. System architecture of this system is shown in fig.1. This System carries out according to the following steps.

Step1: Accepting input Source sentence (Myanmar/English)
Step2: Parsing input Source sentence to generate parse tree by using parser
Step3: Transferring Source parse tree to Target parse tree according to SCFG rules
Step4: Specifying Target word for each Source word by using Bilingual Dictionary
Step5: Generating output Target sentence

When user inputs a sentence (Myanmar or English) in step 1, the system parses input sentence using Stanford parser as shown in fig.2 and fig.4(by step 2). It transfers tree to tree grammar by using SCFG rules in step 3. Tree to tree transformation is explained in section 4.1 and SCFG rules are explained in section 4.2. During transformation, the system is considered to omit article (a, an, the) for English and (ယောက်, ခု, လုံး, etc) for Myanmar. After transformation, transferred parse tree is shown in fig.3 and fig.5. In step 4, it looks up bilingual dictionary to specify target words. Finally it generates the target sentence.

4.1 Tree to Tree Transformation

It changes syntactic structure of source text with respect to the target text. The syntactic information of source sentence from parser is checked for the matching CFG rules. If the syntactic pattern of source sentence matches with source rule, then the corresponding target rule is taken. The source tree structure of parser is modified with respect to the target rule.

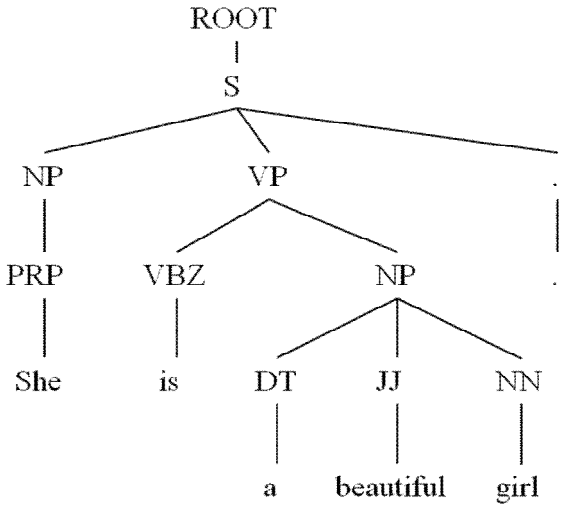


Fig. 2. An example of a CFG with a parse tree of the sentence "She is a beautiful girl."

4.2 Content-Free Grammar

The language defined by a CFG (context-free grammar) is the set of strings derivable from the start symbol S (for Sentence). The four categories of Context-free grammars are Set of non terminal symbols- grammatical categories, Set of terminal symbols- words, Set of Productions- (unordered) (rewriting) rules and Distinguished symbol- start symbol.

Some of CFG rules generated by using Stanford parser are described in the followings and SCFG rules are presented by using this CFG rules in the followings.

ROOT→S | FRAG | SQ | SBARQ | NP
S→VP | NP VP | NP ADVP VP | S VP
NP→NP PP | DT NN | NN | PRP | NN CC NN | JJ NN | DT JJ | DT | NP NP | NNP POS | NNP | CDNN | PRP$ JJ NN
FRAG→WHNP | WHADJP | ADVP NP | ADJP | PP, NP | NP
SQ→MD NP VP | VP | VBP NP VP | VBZ ADVP | NP VP

SBARQ→ WHNP SQ | WHADVP SQ
VP→VBP NP | VB | VB NP | VB PRT | VBP | TO VP | MD VP | VB ADJP NB S | MD RB VP | VBD ADJP | MD RB VP | VB RB VP | VB NP PP | NNP NP | VBP ADJP, S

ADVP→RB | ADVP NP
WHNP→WDT | WP | WP NP | WP NNS
WHADJP→WRB JJ PP | WRB
ADJP→RP | RB JJ | RB | JJ | RB RB
PP→IN PP | IN S | IN NP | IN | TO NP | IN WHNP

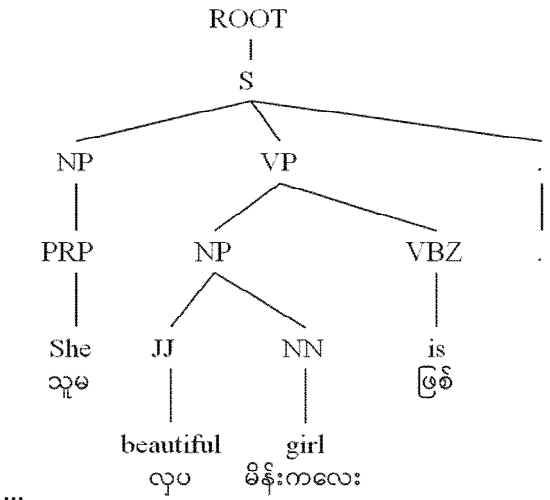


Fig. 3. An example of a CFG with a parse tree of the sentence "She is a beautiful girl" after transformation

Input: She is a beautiful girl.
Output: သူမသည်လှပသောမိန်းကလေးဖြစ်သည်

The following CFG rules are generated by using Myanmar3 parser.
S→NP_NOM<&>NP<&> VEND |
NP_NOM<&>VEND | VEND |
NP_NOM<&>NP_COMPLEMENT<&>VERB_STATIVE | NP_OBJ<&>NP_NOM <&> VEND
NP →N | NP_ADJN | PRON_PERSON
NP_NOM→N | NP <&> PREP_NOM | NP
ADJ_P→N<&>ADJ | ADJ
VEND→ADV<&> VEND | V | V<&>PREP_VERB
NP_COMPLEMENT→NP
NP_OBJ→NP
NP_OBJ→NP<&>PREP_OBJ
…

4.2.1 Synchronous Context Free Grammar

Synchronous context free grammar is a kind of context free grammar that generates pair of strings. Myanmar CFG rules and English CFG rules pairs are shown as SCFG rules in table 1 and table 2.

Table. 1.Example SCFG rules for English parse tree

Phrases	E-CFG	M-CFG
NP	NP PP	PP NP
NP	DT NNS	NNS
NP	NNP	NNP
NP	PRP$ NN	PRP$ NN
VP	VBZ VP	VP
VP	VB NP PP	PP VB
VP	VBG PP	PP VBG
VP	VBZ NP	NP VBZ
VP	VBD PRT NP	NP PRT VBD
VP	VBP NP PRT PP	PP VBP
PP	IN NP	NP IN
PP	TO NP	NP TO

Table. 2.Example SCFG rules for Myanmar parse tree

Phrases	M-CFG	E-CFG
S	NP_NOM NP_ARRIVAL VEND	NP_NOM VEND NP_ARRIVAL
S	NP_NOM NP_COMPLEMENT VERB_STATIVE	NP_NOM VERB_STATIVE NP_COMPLEMENT
NP_NOM	NP PREP_NOM	NP
NP_OBJ	NP PREP_OBJ	NP
NP_ARRIVAL	NOUN_ARTIFACT PREP_ARRIVAL	PREP_ARRIVAL NOUN_ARTIFACT
NP_DEPATURE	NOUN_ARTIFACT PREP_DEPATURE	PREP_DEPATURE NOUN_ARTIFACT

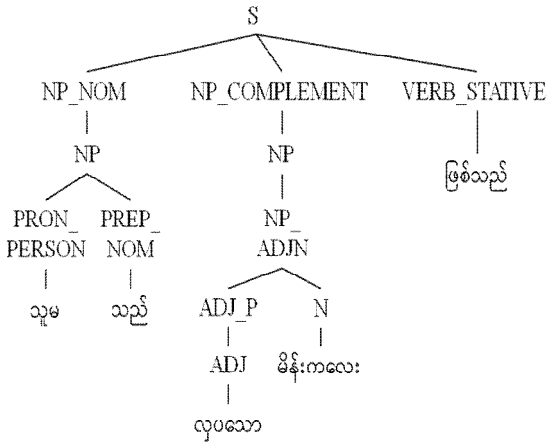


Fig. 4. Example Myanmar sentence before transformation

Input: သူမသည်လှပသောမိန်းကလေးဖြစ်သည်
Output: She is a beautiful girl.

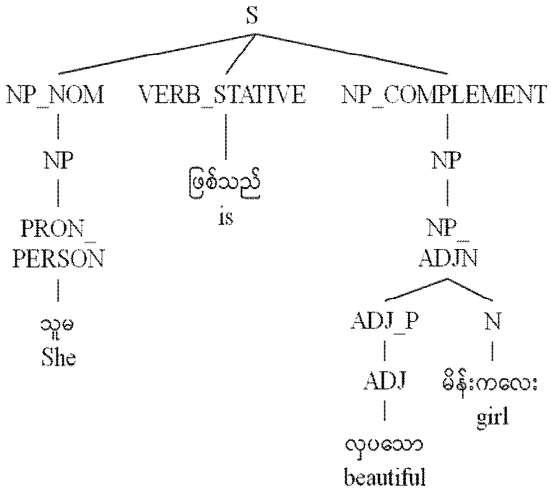


Fig. 5. Example Myanmar sentence after transformation

CONCLUSION

Myanmar-English translation system implements bidirectional translation system. Bilingual dictionary is used for translation of source words. Translating from one language to another, direct translation

cannot perform because of structural difference of source and target language. So CFG rules cannot be the same. For translation, SCFG rules are generated via CFG rules of source and target language structure. Tree to tree transformation approach is applied by using SCFG rules. This system can translate short sentences only. This system cannot be to translate for ambiguous words until now. This system will reduce delay time during translation. This system can support for Myanmar-English language communication. It can get acceptable result and can be applied in real world.

REFERENCES

[1] P. J. Antony, "Machine Translation Approaches and Survey for Indian Languages" The Association for Computational Linguistics and Chinese Language Processing Vol. 18, No. 1, March 2013, pp. 47-78.

[2] S.R.Priyanga, AP, A.AzhaguSindhu, AP, "Rule Based Statistical Hybrid Machine Translation", Internnational Journal of Science and Modern Engineering (IJISME) ISSN: 2319-6386, Volume-1, Issue-5, April 2013.

[3] Khaled Shaalan et.al, "An English-Arabic Bi-directional Machine Translation Tool in the Agriculture Domain", IFIP International Federation for Information Processing 2010

[4] Mr.Uday C. Patkar et.al, "Transformation of Multiple English Text Sentences To Vocal Sanskrit Using Rule Based Technique", International Journal of Computers and Distributed Systems, Vol. No.2, Issue 1, December 2012

[5] Fai Wong et.al, "Handheld Machine Translation System Based on Constraint Synchronous Grammar", Machine Translation Summit XIII, Sep. 2011, p. 439-446.

[6] Shibli Syeed Ashrafi et.al, "English to BanglaMachine Translation System Using Context-Free Grammars", IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 3, No 2, May 2013.

[7] T. T. Zin et al, "Myanmar Phrases Translation Model with Morphological Analysis for Statistical Myanmar to English Translation System", International Journal of Computer Applications, Volume 28, No 1, 2011, pp 13-19

[8] http://en.wikipedia.org/wiki/Statistical_machine_translation.

★★★