Feature based Semantic Analyzer for Parsing Bangla Complex and Compound Sentences

Parijat Prashun Purohit, Mohammed Moshiul Hoque, Mohammad Kamrul Hassan
Chittagong University of Engineering & Technology, Bangladesh

***Abstract*—Semantic analyzer determines the semantic meaning of the words in a sentence. This paper proposes a semantic analyzer that can semantically parse the Bangla sentences. Without semantic analysis, it is very difficult to find the accurate meaning of the translated sentences in one language into their equivalent sentences in other language. To analyze the Bangla sentences semantically, this study identifies a set of features of each word categories in Bangla. Experimental results show that proposed features can be used effectively for analyzing all kinds of Bangla sentences with semantic analyzer.**

Keywords—Natural language processing, syntax analysis, semantic analysis, semantic features, anotated parse tree.

I. INTRODUCTION

Natural Language Processing (NLP) are developed to explore both general theories of human language processing tasks such as providing natural language interfaces or front ends to application system. A language-understanding program must have considerable knowledge about structure of the language including what words are and how they are combined into phrases and sentences. It is also essential to know the meaning of the words and how they contribute to the meaning of the sentence in the context within which they are being used. Semantics is the study of sentence meaning and this meaning is achieved partially by analyzing the syntactic structure (s) and the meaning of the words used in the sentences [1].

The automated creation of accurate and expressive meaning representation necessarily involves in a wide range of knowledge-sources and inference techniques. Among the source of knowledge that are typically involved are the meanings of words, the conventional meanings associated with grammatical constructions, knowledge about the structure of the discourse, common-sense knowledge about the topic at hand and knowledge about the state of affairs in which the discourse is occurring [2]. There are many ways of thinking about and representing word meanings, but one that has proved useful in the field of machine translation involves associating words with semantic features which correspond to their sense components. Associating words with semantic features is useful because some words impose semantic constraints on what other kind of words they can occur with. In this work, semantic analysis will perform by assigning the features of each word in a sentence that are based solely on knowledge gleaned from the lexicon and the grammar.

Semantic analysis plays vital role in solving the vagueness in sentence meaning. It is observed that in the sentences semantic properties of words needs to be analysed explicitly for the actual output making. In the dictionary, semantic feature of words is maintained categorically for semantic analysis. When the parser parses a sentence, the words are regained with semantic characteristics that actually establish the word meaning in a sentence. Semantic properties of words are comprised of three facets: domain, context and the task, and the semantic structures constructed from utterances must account these areas. In the knowledge domain, semantic analysis records individual words into appropriate objects, and it must create the correct structures to communicate the meaning of the individual words combined with each other.

Semantic analysis of Bangla language is a very challenging task due to its varieties of word formation and the ways spoken. Moreover, other factors contribute to the difficulty of semantic analysis, including words with multiple meanings, sentences with multiple grammatical structures, uncertainty about what a pronoun refers to, and so on. Some research scholars have already analyzed the Bangla sentences in syntactic way. However, studies on semantic analysis of Bangla sentences are rare and limited. In addition, guidelines are also inadequate for the semantic analysis of different word categories and sentences.

To introduce Bangladeshi products in the global market, it is necessary to write product instructions in various languages. In this regard, an automatic translator is the acceptable candidate. The automatic translator is capable of translating information faster than human translators that can save a lot of time and money. However, semantic analysis plays a potential role in generating the exact meaning of Bangla sentences translated into other language. Consider an example, 'goru akashe ure (গরু আকাশে উড়ে : the cow fly in the sky)'. This sentence is syntactically correct but semantically wrong because we know that the cow (goro) cannot fly. Thus, to produce the legal structure and accurate translation of a sentence semantic analysis is mandatory.

In this paper, a semantic analyzer is proposed to parse a variety of Bangla sentences semantically. To achieve this objective, a group of semantic attributes are recommended for the diverse Bangla word classes. The proposed semantic analyzer is experimented by analyzing a wide range of sentences with inconsistent word lengths, and the findings shows that the analyzer functions well to parse the Bangla sentences semantically.

978-1-4799-6399-7/14/$31.00 ©2014 IEEE

II. RELATED WORK

Bangla language processing is in the preliminary stage. Very few researches have been conducted on semantic analysis of Bangla sentences but a significant number of research works have been conducted on the recognition of Bangla alphabets [3, 4]. Some works has been done on syntax analysis of Bangla simple sentences using CFG's [5, 6], and CSG's [7]. Syntax analysis using CFG's for Bangla simple sentence, complex sentence and compound sentence are presented in [8, 9]. Ali et al. [10] propose a set of rules for morphological analysis to describe Bangla universal networking language. Some recent works focused on designing the machine translation system for Bangla language such as phrase-based [11], example-based [12, 13], rule based [14], and statistical approach [15].

Recently, semantic features of Bangla words with redundancy rules are presented in [16]. A new approach of Mridha et al. [17] is to solve semantic ambiguity problem of Bangla Root words using universal networking language. Basic HPSG structure was proposed in [18] for recognition of semantic correctness. Richardson and his colleagues [19] developed a general methodology for acquiring, structuring, accessing, and exploiting semantic information from natural language text. Bangla simple sentences have analyzed semantically using lexical semantics approach [20]. A recent work has focused on assigning semantic features for parsing Bangla sentence semantically. However, this work was limited to analyze the Bangla simple sentence alone, and features are designed for limited word categories. Thus, in the previous literatures, semantic analysis of complex and compound sentences are remained unexplored. In this paper, we propose a framework for semantic analyzer to parse the Bangla complex and compound sentences.

III. PROPOSED FRAMEWORK

Semantic analyzer deduce the semantic meaning of words in a sentence. In order to perform the semantic analysis of sentences, a lexicon should be implemented with several semantic features. Fig. 1 illustrates the schematic representation of our proposed framework. Following subsections provides the description of this framework in details.

A. Input Sentence

The source language sentences which to be parsed are taken as an input of the system. For semantic analysis, we have to choose complex and compound sentences of Bangla. For example, a compound sentence, "je amake porabe se amar bondhu (যে আমাকে পড়াবে সে আমার বন্ধু)" may be considered for an input.

B. Scanner

Scanner is the program module that accepts a sentence to be parsed as an unbroken string, and breaks into individual words is called token [5]. Tokens are stored in a list for further access. The token is then checked into the lexicon for validity, some words, if necessary, should be combined into groups, because, two or more words may represent a single word type.

From input sentence, scanner will generate a set of tokens. Such as, in our example we have six tokens which are *Subordinator* (je), *Pronoun* (amake), *Verb* (porabe), *Subordinator complement* (se), *Pronoun* (amar), and *Noun* (bondhu) respectively.

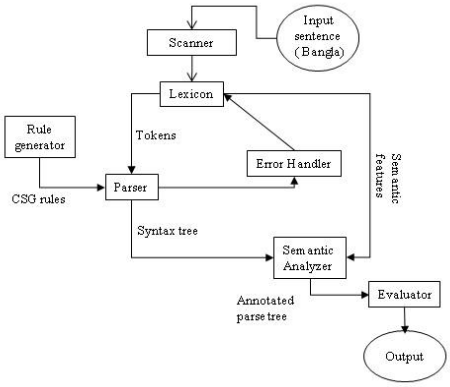

Fig. 1. Proposed framework of semantic analyzer.

C. Lexicon

A lexicon is a dictionary of words where each word contains some syntactic, semantic and pragmatic information. The semantic properties of words may include their types, number, gender, person, etc. [21]. Essentially, entries in an MT dictionary will be equivalent to collections of attributes and values (i.e. features). These features will be used for next steps. Valid tokens will also be provided from Lexicon. For the above example six valid tokens are produced: 'je', 'amake', 'porabe', 'se', 'amar', and 'bondhu' respectively. We must assign values for semantic features of each token. For these tokens, the semantic features are assigned in the following:

যে [*je*] (Subord):
[Honor(0),Number(1),Human(1),Agent(1),Alive(1)]

আমাকে [*amake*] (Pronoun):
[Person(0),Animate(1),Human(1),Number(1),Honor(0)]

পড়াবে [*porabe*] (Verb):
[Person(0),Animate(1),Human(1),Intelligent(1),Honor(1),Agent(1)]

সে [*se*] (Subcom):
[Honor(0),Number(1),Human(1),Agent(1),Alive(1)]

আমার [*amar*] (Pronoun):
[Person(0),Animate(1),Human(1),Number(1),Honor(0)]

বন্ধু [*bondhu* (Noun):
[Animate (1), Human (1), Alive (1), Intelligent (-1), Agent (1), Gender (1), Adult(-1)]

Here, amake (আমাকে) is a pronoun and first person. Thus, for first person we assign value 0. For second and third person the value is 1 and 2 respectively. One of the semantic features is animalism and humanism that values 1 in each case. As it is singular in number we assign 1. For plural number we assign 2. Finally, the token is non honorable and so we put 0 in

Honor feature. Similarly, we have put values for other two tokens. It should be mentioned if a feature for a token is not significant or confusing, we can assign (-1) to mean it.

Generally there are five categories of Bangla word (i.e., part of speech) are found in Bangla language. Table I shows some semantic features of Noun (বিশেষ্য) category.

TABLE I. SEMANTIC FEATURE OF NOUN

Words	Semantic Features							
	Animate	*Human*	*Agent*	*Honor*	*Number*	*Intelligent*	*Adult*	*Gender*
Biggani (বিজ্ঞানী)	1	1	1	-1	-1	1	1	-1
Durbrittora (দুর্বৃত্তরা)	1	1	1	0	0	-1	-1	-1
Pakhigulo (পাখিগুলো)	1	0	1	0	0	0	-1	-1
Jubok (যুবক)	1	1	1	-1	1	1	1	0

For verb (ক্রিয়া) categories, propose semantic features are listed in Table II.

TABLE II. SEMANTIC FEATURE OF VERB

Words	Semantic Features					
	Person	*Animate*	*Human*	*Intelligent*	*Honor*	*Agent*
Haschen (হাসছেন)	2	1	1	-1	1	1
Likhi (লিখি)	0	1	1	1	0	1
Hati (হাঁটি)	0	1	-1	-1	0	1
Jano (জানো)	1	1	1	1	0	1

A negative value of feature will assigned for a word that is not significant or relevant to it. We have assigned the features for Pronoun, Adjective, and Adverb in similar way. Some examples are presented in Tables III, IV, and V respectively.

TABLE III. SEMANTIC FEATURE OF PRONOUN

Words	Semantic Features				
	Person	*Animate*	*Human*	*Number*	*Honor*
Tomader (তোমাদের)	1	1	1	0	0
Unara (উনারা)	2	1	1	0	1
Amake (আমাকে)	0	1	1	1	-1

TABLE IV. SEMANTIC FEATURE OF ADJECTIVE

Words	Semantic Features		
	Animate	*Human*	*Gender*
Buddhimoti (বুদ্ধিমতী)	1	1	0
Soktiman (শক্তিমান)	1	1	1
Khusi (খুশি)	-1	1	-1
Sohoj (সহজ)	-1	1	-1

TABLE V. SEMANTIC FEATURE OF ADVERB

Words	Semantic Features					
	Emphasis	*Animate*	*Human*	*Intelligent*	*Honor*	*Tense*
Besh (বেশ)	1	-1	-1	-1	-1	-1
Agamidin (আগামীদিন)	0	0	0	-1	-1	2
Ekhon (এখন)	0	0	1	-1	-1	0
Onek (অনেক)	1	-1	1	1	-1	-1
Sorbotro (সর্বত্র)	1	-1	-1	-1	-1	-1

Table VI and Table VII illustrates the semantic features for various markers that are used in conjunction with the complex and compound sentence as described in [23]. These markers indicate whether it is complex or compound sentenced and divide the entire sentence into two or more clauses for analysis [7]. Most of the cases, these markers are found as a pair in the complex sentence.

TABLE VI. SEMANTIC FEATURES OF COMPLEX MARKERS

Pair of Markers		Honor	Number	Human	Agent	Alive
Jodi (যদি)	Tahole (তাহলে)	-1	-1	-1	-1	-1
Jini (যিনি)	Tini (তিনি)	1	1	1	1	1
Jokhn (যখন)	Tokhn (তখন)	0	-1	0	0	0
Ja (যা)	Ta (তা)	-1	1	0	0	1
Jara (যারা)	Tara (তারা)	0	0	1	1	1
Je(যে)	Se(সে)	0	1	1	1	1
Jekhane (যেখানে)	Sekhane (সেখানে)	0	-1	0	0	0

TABLE VII. SEMANTIC FEATURE OF CONJUNCTION MARKERS

Markers	Negative	Optional	Honor	Contradictory	Sequence
Ebong (এবং)	0	0	-1	0	0
Othoba (অথবা)	0	1	-1	0	0
Kintu (কিন্তু)	1	-1	-1	1	0
O(ও)	0	0	-1	0	0
Boroncho (বরঞ্চ)	1	0	-1	1	0
Tai (তাই)	0	0	-1	0	1

D. Rule Generator

The most common way to represent grammar is a set of production rules which says how the parts of speech can put together to make grammatical sentences. CSG are the methods of describing language. A set of CFG and CSG rules [7, 8] will also be used in our framework. Tokens are matched with rules to form a parse tree in parser. In our example, we will need a set of CFG rules that will generate "Subord P V Subcom P N" to be matched with tokens. So, the set of rules that will be provided to parser are:

S → CS; CS → IC DC; IC → Subord SS1; SS1 → NP VP; NP → P; VP → V; DC → Subcom SS2; SS2 → NP NP; NP → P NP → N;

Subord → je (যে); P → amake (আমাকে); V → porabe (পড়াবে); Subcom → se (সে); P → amar (আমার); N → bondhu (বন্ধু)

E. Parser

The function of the parser is to take an input string or sentence and produce a parse tree according to CFG rules. Output tokens of the scanner will match with the appropriate grammatical rules. If the right hand symbols of a rule are matched with a token then the token is assigned with the appropriate word category. Error handler will check the possible errors while parsing and resolve these if necessary. The output of the parser is generally a parse tree or can be represented in the list form of data structure [22]. For our example, we can represent the parse tree as the following list form.

S(CS(IC(Subord, je)(SS1(NP(P, amake))(VP(V, porabe)))) (DC (Subcom, se)(SS2(NP(P, amar))(NP(N, bondhu)))))

F. Semantic Analyzer

For MT engine, the semantic attributes are necessary for transfer as well as generation phases. Semantic attributes must be transferred into target language in such a form so that the generation of target output is semantically correct. The words and their semantic attributes of different categories are stored in a lexicon and they are retrieved while parsing. Thus, the parse tree goes to be assigned with semantic features. After generating the parse tree, the analyzer incorporates all the features of word according to category that are assigned in the lexicon. As a result, the analyzer generates an annotated parse tree for the input sentence. Leaves of an annotated parse tree will contain specific semantic features. A sentence whether correct or not, does not depend on all semantic features of the word. We compared such features that are significant to check a sentence's correctness. The features of a token that were stored in lexicon will now be added in parse tree to form the annotated parse tree. Fig. 2 depicts the annotated parse tree for our example.

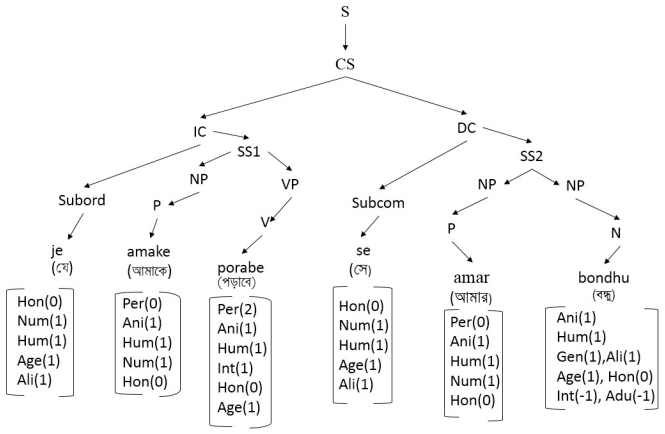

Fig. 2. Annotated parse tree for the complex sentence, "je amake porabe se amar bondhu (যে আমাকে পড়াবে সে আমার বন্ধু)".

G. Evaluator

The evaluator verifies the correctness of the input sentence on the basis of semantic features. All tokens with appropriate features are stored in the lexicon. After generating the annotated parse tree relevant feature values will map. If the feature values are matched, then the sentence will be assigned semantically correct. Otherwise it will be incorrect. Fig. 3 illustrates the mapping of features for the complex sentence, Jadi tumi boi poro tahole tumi valo korbe (যদি তুমি বই পড়ো তাহলে তুমি ভালো করবে).

In this example, the analyzer first maps the features of complex markers ('Jadi' and 'tahole'). If all the features contains same value (either either 0 or 1) then the analyzer check the semantic correctness of sub sentences. Here two modules map features of two sub sentences: 'tumi boi poro' and 'tumi valo korbe'. In each module corresponding features are retrieved first with values. If the value is don't care (-1) then we avoid the mapping. In module1, Animate has -1 value. Therefore, no mapping is needed. If two tokens have common features and they have either 0 or 1 then the analyzer maps those pair-by-pair. Matching pair are shown by an arrow. All the mapping must have green arrow to be semantically correct. If there is a mismatch in value therefore the sentence can be regarded semantically incorrect. In this example, there is no

such mismatch. So output from two modules finally verifies the sentence semantically correct.

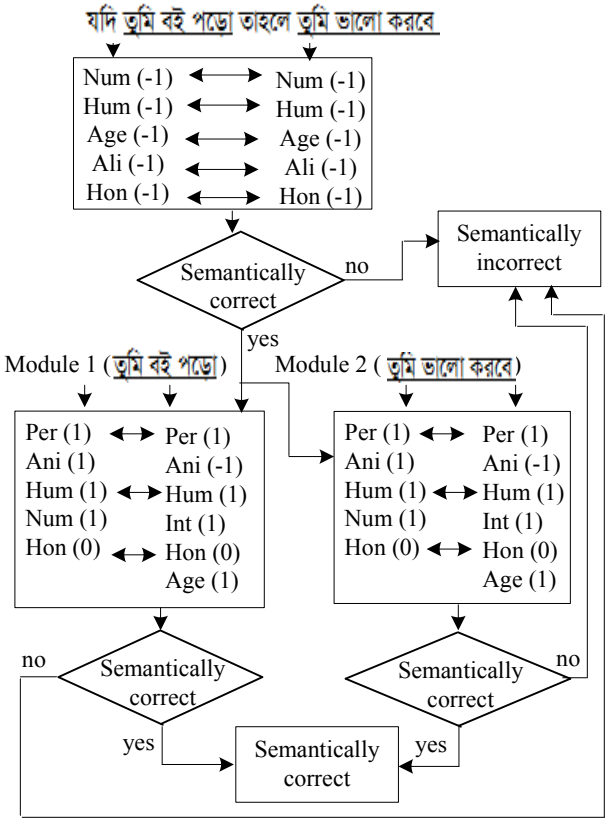

Fig. 3. Features mapping process diagram for the complex sentence, jadi tumi boi poro tahole tumi valo korbe (যদি তুমি বই পড়ো তাহলে তুমি ভালো করবে).

H. Output

The semantic analyzer generates the parsing output of a given input sentence with appropriate decisions (i.e., semantically correct or not). It will display a tag, such as 'semantically correct' or 'semantically incorrect' depending on the correct or wrong sentences. In addition to that it also provides the information about tokens, grammatical rule used; parse tree information and semantic features involved in the parsed sentence.

IV. EXPERIMENTAL RESULTS

A. Output of Semantic Analyzer

A complex or compound sentence is combination of simple sentences. Two simple sentences are connected with a conjunction (□□□□□). If two sentences are individually correct then overall evaluation of sentence is also correct.

A sentence consisting of one principal clause and one or more sub-ordinate clauses is a complex sentence. For complex sentence two clauses are required to be verified. If principal clause and subordinate clause are semantically correct then it is correct. A snapshot of the result of analyzer for a complex sentence are shown in Fig. 4.

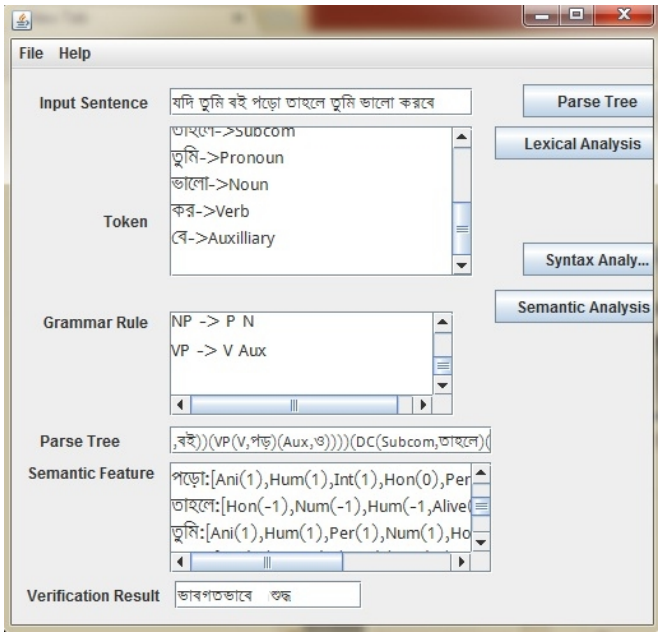

Fig. 4. Semantic analysis of Bangla complex sentence, jadi tumi boi poro tahole tumi valo korbe (যদি তুমি বই পড়ো তাহলে তুমি ভালো করবে).

This figure indicates the semantic features for the input sentence, "jadi tumi boi poro tahole tumi valo korbe (যদি তুমি বই পড়ো তাহলে তুমি ভালো করবে)" as follows:

যদি (jadi): [Hon(-1),Num(-1),Hum(-1,Alive(-1,Agent(-1)]

তুমি (tumi): [Ani(1),Hum(1),Per(1),Num(1),Hon(0)]

বই (boi): [Ani(0),Hum(0),Ali(0),Int(1),Age(-1,Gen(-1), Adu(-1),Hon(-1)]

পড়ো (poro): [Ani(1),Hum(1),Int(1),Hon(0),Per(1),Age(1)]

তাহলে (tahole): [Hon(-1),Num(-1),Hum(-1,Alive(-1,Agent(-1)]

তুমি (tumi): [Ani(1),Hum(1),Per(1),Num(1),Hon(0)]

ভালো (valo):[Ani(-1),Hum(-1),Ali(0),Int(-1),Age(0,Gen(1), Adu(-1),Hon(-1)]

করবে (korbe): [Ani(1),Hum(1),Int(-1),Hon(0),Per(2),Age(1)]

The sentence is correct syntactically as well as semantically. If we consider the markers "jodi (যদি)" and "tahole (তাহলে)" both of these complex markers have simillar semantic features. Then we check for the semantic correctness of individual clauses. The two clauses are individually correct in semantic manner. So output is thereby semantically correct (ভাবগতভাবে শুদ্ধ). A sentence having more than one principal clauses linked by one or more corordinating conjunctions prededed by a comma is called compound sentence. Conjunction are used in Bangla compound sentences are ebong (এবং), othoba (অথবা), kintu (কিন্তু), o (ও), and so on.

An analysis of the compound sentence, durbrittora hamla kore ebong pulishra greftar kore (দুর্বৃত্তরা হামলা করে এবং পুলিশরা গ্রেফতার করে) is shown in Fig. 5.

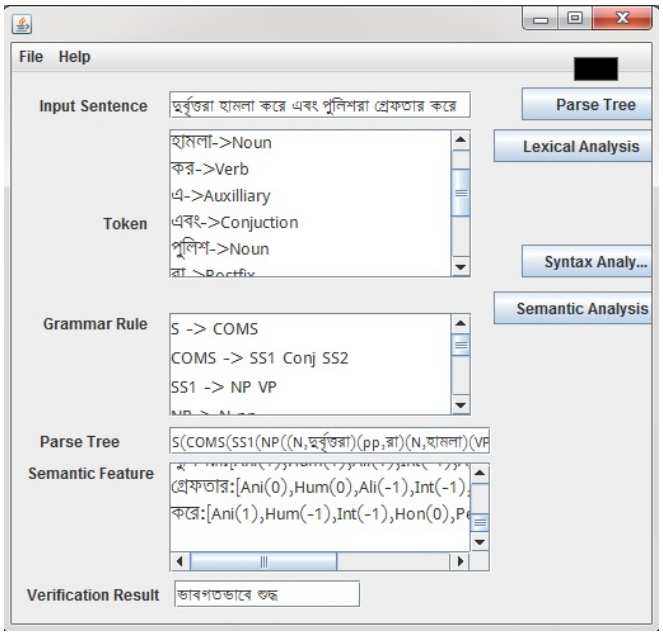

Fig. 5. Semantic analysis of Bangla compound sentence, durbrittora hamla kore ebong pulishra greftar kore (দুর্বৃত্তরা হামলা করে এবং পুলিশরা গ্রেফতার করে).

This compound sentence consists of two simple sentences and these are checked for their correctness. To determine the sentence as compound we required the features of the token 'ebong (এবং)'. When the features were matched the compound sentence is subdivided into two portions. Each will verify for semantic correctness. Semantic features of all the words in the example sentence are illustrates in the following:

দুর্বৃত্তরা (durbrittora): [Ani(1),Hum(1),Ali(1),Int(-1),Age(1,Gen(-1), Adu(1), Hon(0),Age(1)]

হামলা (hamla): [Ani(0),Hum(0),Ali(0),Int(-1),Age(0,Gen(-1),Adu(-1),Hon(0)]

করে (kore): [Ani(1),Hum(-1),Int(-1),Hon(0),Per(2),Age(1)]

এবং (ebong): [Neg(0),Opt(0),Hon(-1,Con(0,Seq(0)]

পুলিশরা (pulishra): [Ani(1),Hum(1),Ali(1),Int(-1),Age(1,Gen(-1),Adu(1), Hon(0),Age(1)]

গ্রেফতার (greftar): [Ani(0),Hum(0),Ali(-1),Int(-1),Age(-1,Gen(-1),Adu(-1), Hon(0)]

করে (kore): [Ani(1),Hum(-1),Int(-1),Hon(0),Per(2),Age(1)]

B. Performance Analysis

Our proposed system can analyze simple, complex and compound sentences of Bangla. We have tested the system for 1120 sentences in total with several of sentence length. Among them a total of 550 were complex sentences and rest of them was compound sentences. Table III summarizes the accuracy measures for different length of sentences.

TABLE VIII. PERFORMANCE EVALUATION OF PROPOSED SYSTEM

Sentence Type	No. of input sentences	Word Lengths	No. of correctly parsed sentences	Error	Accuracy
Complex	200	5	200	0.00%	100%
	100	6	100	0.00%	100%
	150	7	147	2%	98%
	80	8	67	16.25%	83.75%
	20	9	14	30%	70%
Compound	100	5	100	0.00%	100%
	110	6	110	0.00%	100%
	250	7	249	0.4%	99.6%
	80	8	70	12.5%	87.5%
	30	9	20	33.33%	66.67%

Here, 'accuracy' refers to the ratio between the total number of sentences that are correctly parsed and total number sentences that are inputted in the framework. The 'sentence length' refers to the total number of words in a given sentence. Sentences are collected from different Bangla books and newspapers. The input sentence length varies from 5 to 9. The result of analysis of accuracy for different sentence length is illustrated in Fig.5. This result reveals that the accuracy of the system varies sharply with increasing word length for both types of sentences.

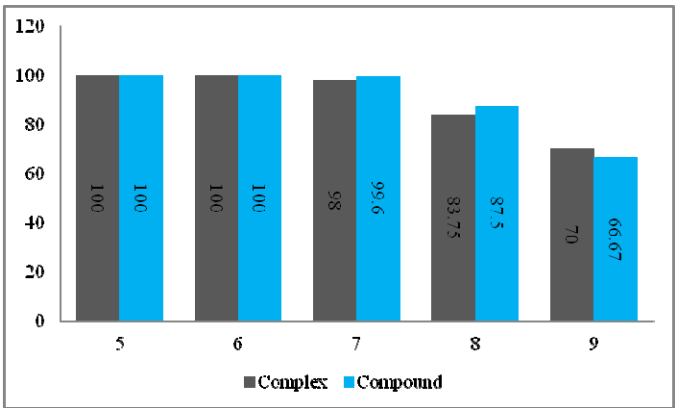

Fig. 6. Analysis of accuracy vs. sentence length.

V. CONCLUSION

In any language semantic analysis can play a significant role in machine translation and language understanding problems. Semantic properties of words of a source language must be considered to generate a meaningfully correct machine translation. For proper machine translation from Bangla to English or vice versa, semantic features of Bangla words should be structured. Our proposed framework can analyze the complex and compound sentences with a set of semantic features. Experimental result shows that the performance of the system is quite good. Semantic analyzer potentially serves in language to combine the meaning of words and phrases. Future extension can be possible to include more features to analyze Bangla idioms and phrases, complex phrases, and punctuation symbols.

References

[1] D. W. Patterson, Introduction to Artificial Intelligence and Expert System, Printice Hall, India, 2002.

[2] D. Jurafsky, and J. H. Martin, Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition, Prentice Hall, Englewood Cliffs, New Jersey, 2006.

[3] M. K. I. Molla, and K. M. Talukder, Bangla number extraction and recognition from document image, Proc. of 10th Int. Conf. on Computer and Information Technology (ICCIT), pp. 512-517, 2007.

[4] M. M. Hoque, M. R. Karim, M. G. Hossian. M. S. Arefin and M. M. U. Hasan, "Bangla numeral recognition engine (BNRE)", Proc. of Int. Conf. on Electrical and Computer Engineering (ICECE), pp. 644-647, 2008.

[5] M. M. Murshed, "Parsing of bengali natural language sentences", In Proc. International Conference on Computer and Information Technology, ICCIT'98, Dhaka, Bangladesh, pp. 185-189, vol. 1, 1998.

[6] M. R. Selim and M. Z. Ikbal, "Syntax analysis of phrases and different types of sentences in Bangla", In Proc. International Conference on Computer and Information Technology, ICCIT '99, Sylhet, Bangladesh, pp. 175-186, vol. 2, 1999.

[7] M. M. Hoque and M. M. Ali, "Context-sensitive phrase structure rules for structural representation of Bangla natural language sentences", In Proc. Int. Conf. on Computer and Information Technology (ICCIT), pp. 615-620, 2004.

[8] M. M. Hoque and M. M. Ali, "A parsing methodology for Bangla natural language sentences", In Proc. International Conference on Computer and Information Technology, ICCIT'03, Dhaka, Bangladesh, pp. 277-282, vol. 2, 2003.

[9] L. Mehedy, S. M. Arefin and M. Kaykobad, "Bangla syntax analysis: A comprehensive approach", In Proc. International Conference on Computer and Information Technology, ICCIT'03, Dhaka, Bangladesh, pp. 287-293, vol. 5, 2003.

[10] M. N. Y. Ali, M. Z. H. Sarker, G. F. Ahmed, J. K. Das, "Rules for morphological analysis of Bangla verbs for universal networking language," International Conference on Asian Language Processing (IALP), pp.31-34, 2010.

[11] M. Rabbani, K. M. R. Alam, M. Islam, "A new verb based approach for English to Bangla machine translation", Proc. International Conference on Informatics, Electronics & Vision (ICIEV), pp. 1-6, 2014.

[12] S. K. Naskar and S. Bandyopadhyay, "A phrasal EBMT system for translating English to Bengali, Proc. of the Workshop on Language, Artificial Intelligence and Computer Science for Natural Language Processing Applications (LAICS-NLP), 2006.

[13] D. Saha and S. Bandyopadhyay, "A semantics-based English-Bengali EBMT system for translating news headlines", *MT SUMMIT X,* Phuket, Thailand 2005.

[14] J. F. Islam and M. M. Mia, Adapting Rule Based Machine Translation from English to Bangla: An Easy Solution for Translation, Lap Lambert Academic Publishing, 2012.

[15] M. Z. Islam, English to Bangla Phrase-Based Statistical Machine Translation, Master Thesis, Department of Computational Linguistics, Saarland University, Germany, 2009.

[16] M. M. Hoque and M. M. Ali, "Semantic features and redundancy rules for analyzing Bangla sentences", In Proc. International Conference on Computer and Information Technology (ICCIT'05), Dhaka, Bangladesh, pp. 1198-1201, vol. 4, 2005.

[17] M. F. Mridha, A. K. Saha, J. K. Das, "New approach of solving semantic ambiguity problem of Bangla root words using universal networking language (UNL)", Proc. International Conference on Informatics, Electronics & Vision (ICIEV), pp. 1-6, 2014.

[18] M. A . Islam, K. M. A. Hasan, M. M. Rahman, "Basic HPSG structure for Bangla grammar", International Conference on Computer and Information Technology (ICCIT), pp.185-189, 2012.

[19] S. D. Richardson, W. B. Dolan, and L. Vanderwende, "MindNet: acquiring and structuring semantic information from text" , Proc. *COLING-ACL*, pp. 1098-1102, 1998.

[20] M. M. Hoque, M. J. Rahman and P. K. Dhar, "Lexical semantics: A new approach to analyze the Bangla sentence with semantic features", Proc. International Conference on Information and Communication Technology (ICCIT), Dhaka, Bangladesh, pp. 87-91, vol. 1, 2007.

[21] A. J. Thomson and A. V. Martinet, A Practical English Grammar, Green View Publishers, Dhaka, Bangladesh, 2001.

[22] A. Trujillo, Translation Engines: Techniques for Machine Translation, Springer-verlag, London, 1992.

[23] H. Azad, Bakkyatotto(□□□□□□□□□□□□), Dhaka University, Dhaka, Second Edition, 1994.