INTELLIGENT PARSING IN NATURAL LANGUAGE PROCESSING

Sanghamitra Mohanty
Dept. Computer Science & Application
Utkal University
Bhubaneswar
ORISSA, INDIA
Email: sangham1@rediffmail.com

Rakesh Chandra Balabantaray
The ICFAI Institute of Science & Technology
Fortune Tower, Chandasekharpur
Bhubaneswar
ORISSA, INDIA
Email: rakesh_b_ray@rediffmail.com

Keywords: MT, N-GRAM and Intelligent Parser.

Abstract

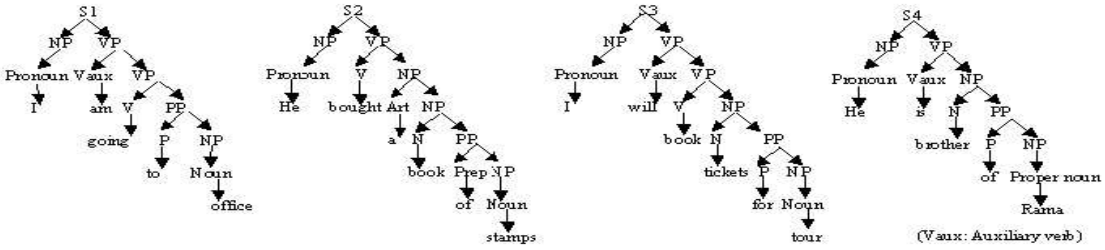
Parser does the part of speech (POS) identification in a sentence, which is required for Machine Translation (MT). An intelligent parser is a parser, which takes care of semantics along with the POS in a sentence. Use of such intelligent parser will reduce the complexity in semantics during MT apriori.

I. INTRODUCTION:

In the context of Natural Language Processing (NLP), parsing may be defined as the process of assigning structural description to sequence of words in a Natural Language(or to sequence of symbols derived from word sequences). The structural description assignment of words depends upon the grammar (a description language plus a set of structural constraints), according to which the parser attempts to analyse the sequence of symbols presented to it. Prior emphasis on semantic analysis with the help of tagged lexical resources like WordNet may enable us to achieve an intelligent parsing technique, which may be helpful for efficient Machine Translation (MT). Such type of parsers may be called as intelligent parsers. This type of parsing technique can distinguish a sentence like "The stick is laughing" for appropriate MT where as a normal parser will accept it and will produce unaccepted results during MT. We have developed an intelligent parsing algorithm, which takes care of semantic analysis of a sentence during parsing.

II. TECHNIQUES FOR INTELLIGENT PARSING:

For simple sentence like: "S1: I am going to office." can easily be parsed according to the grammar rules along with the help of WordNet. Now for sentences like: "S2: He bought a *book* of stamps." and "S3: I will *book* tickets for tour." both having the word *book* but used in different categories. In the sentence S2 *book* is noun as it is adjacent to an article (according to grammar rules). In the sentence S3 *book* is verb as it is adjacent to an auxiliary verb and there is no other main verb in it (according to grammar rule). In sentence "S4: He is brother of *Rama*.", while parsing when we encounter a word having the first character as capital and not present in the lexical resources will be considered as proper noun. In sentence "S5: *Plant* a spy in Moscow." *Plant* can be either noun or



verb like *book* but since it is the beginning of the sentence and followed by an article it will act as verb in this sentence (according to grammar rules). Now in sentences "S6: They built a *plant* to manufacture automobiles." and "S7: The growth of aquatic *plant* life in water is very good." the word *plant* is noun (according to grammar rule) but it has various senses in noun form (According to the tagged lexical resources: WordNet). Here for this word *plant* by applying Unsupervised algorithm by Yarowsky [3] based on N-gram model we are assigning the sense to the word along with parsing. So, that it will be quiet helpful for MT.

Ex: The sentences S6 and S7 both contain the word *plant* but the word plant has two senses in noun form (i.e. $sense_i$, $sense_j$). For assigning appropriate sense to the word we have to identify collocations by taking nearby words along with that word. Then depending on the values of the probabilities $Pr(sense_i / Collocation)$ and $Pr(sense_j / Collocation)$ we will assign the sense to the word.

III. ALGORITHM FOR INTELLIGENT PARSING

ALGORITHM (INTEL PARSE) //Intelligent Parsing

1. Start
2. Do parsing.

3. For i =1 to number of words in a sentence
4. Start assigning POS to the words // i.e. to $word_i$
5. If the word having first character as capital and not present in WordNet assume it as Proper noun.
6. If the word having more than one category resolve it according to grammar rules (as already discussed).
   6.1. After assigning category if the word has more than one sense apply algorithm based on N- gram model [3] to assign appropriate sense.
7. End for
8. Until end of sentences in the text.
9. End

IV. RESULTS AND DISCUSSION

In this section we will compare our technique with the traditional parsing techniques based on the following examples. The results are shown in Table-I.

S8: I am going to *bank* for money.
S9: I have a manufacturing *plant*.
S10: I have a big *tank* for water.

TABLE-I

Traditional Parsing	Our Parsing
For S8 only parsing structure will be extracted as discussed earlier. (see figure 1) For S9 only parsing structure will be extracted as discussed earlier. (see figure 2) For S10 only parsing structure will be extracted as discussed earlier.	Here along with parsing structure the sense of *bank* in this context is extracted. (see figure1) Here along with parsing structure the sense of *plant* in this context is extracted. (see figure2) Here along with parsing structure the sense of *tank* in this context is extracted.

Traditional Parsing **Intelligent Parsing**

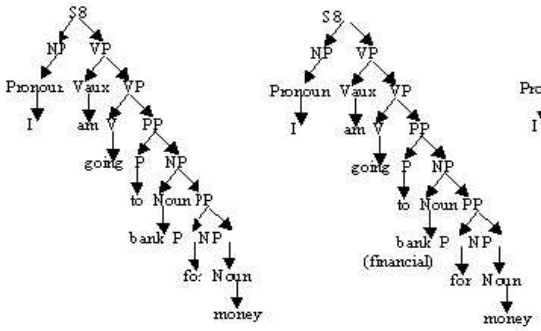


Figure 1 : Shows the comparison between the traditional parsing and intelligent parsing for the sentence "I am going to *bank* for money".

Traditional Parsing **Intelligent Parsing**

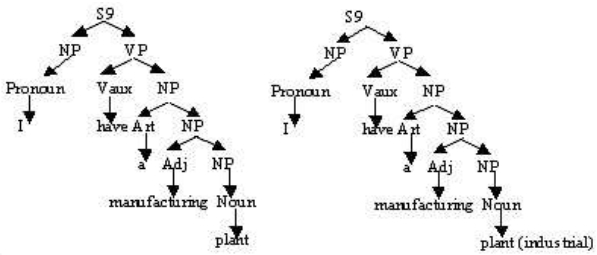


Figure 2 : Shows the comparison between the traditional parsing and intelligent parsing for the sentence "I have a manufacturing *plant*".

This can also be extended to other examples, which will produce effective outputs for MT.

V. CONCLUSION:

Parsing algorithms are only concentrating on the syntactic structure that a sentence has according to the grammar. These syntactic structures are not good enough for the MT. Our intelligent parsing is dealing with semantics along with the traditional syntactic structure, which will make the MT task easier. The strength of this algorithm is that being sensitive to a wide range of language details is capable of providing better outputs.

REFERENCES

1. Yarowsky D.(1993), "One Sense Per Collocation", Proceedings of ARPA Human Language Technology Workshop, Princeton.
2. Yarowsky D.(1994), " Decision list for Lexical ambiguity resolution : Application to Accent restoration in Spanish & French", Proceedings of the 32nd Annual Meeting of the ACL, Las Cruce, NM.
3. Yarowsky D.(1995), "Unsupervised word sense disambiguation rivaling supervised methods", Proceedings of the 33rd Annual meeting of the ACL,PP 189-196.