ADVERBIAL PHRASE FOR THE SYNTACTIC PARSER

Dr.Franklin Thambi Jose.S

Senior Lecturer

Faculty of Languages and Communication

Sultan Idris Education University, Malaysia

Dr.P.Vijaya

Assistant Professor

CAS in Linguistics

Annamalai University, India

ABSTRACT

Software plays an important role in the society especially in the field of Language. Using software became a common process even in language teaching. The scientific study of Language is known as Linguistics. In the subfield of Applied Linguistics, Computational Linguistics is important. It deals with the computer and linguistic levels. It is also said, as a branch of language studies which applies computer techniques to linguistics field. In Computational Linguistics, Natural Language Processing plays an important role. This came to exist because of the invention of Information Technology. In computational syntax, the syntactic analyser breaks a sentence into phrases and clauses and identifies the sentence with the syntactic information. Tamil language is spoken among one of the ethnic groups in Malaysia. Tamil became very popular during the period of Tamil Chola Dynasty of 11th Century. Today they form the third largest ethnic group in Malaysia after Malays and Chinese. It constitutes 8% of the total Malaysian population. There are sentence connectors in Tamil. The words like pinnar 'afterwards', aannaal 'but', atanaal 'hence', aakavee 'so', piRaku 'afterwards' are said to be sentence connector, because, such words function as a sentence connector or discourse marker at the syntactic level. In computational syntax, the first step is to provide required information regarding the head and its constituent of each phrase. This information will be incorporated to the system using programming languages. Now the system can easily analyse a given sentence with the criteria or mechanisms given to it. The major objectives of this paper are to provide simple rules to the syntactic software which is going to be developed and how this software identifies the sentence connector in a given sentence using the given rules.

Key words: linguistics, computer, software, syntax

1.1 INTRODUCTION

Computational Linguistics is a branch of linguistics, which deals with the computer and linguistic levels. It is also said, as a branch of language studies which applies computer techniques to linguistics field. In Computational Linguistics, Natural Language Processing plays an important role. This came to exist due to the invention of Information Technology.

The application of computer can be done in all the fields of Linguistics like Phonetics and Phonology, Morphology, Syntax, Lexicography, etc. This application field comes under Applied Linguistics. The subfields of Computational Linguistics are Computational Phonology, Computational Morphology, Computational Syntax and so on. In computational syntax, the first step is to provide required information regarding the head and its constituent of each phrase. This information will be incorporated to the system using programming languages. Now the system can easily analyse a given sentence with the criteria or mechanisms given to it.

Providing needful criteria or mechanisms to the computer to identify the adverbial phrase in Tamil language is the major objective of this paper.

1.2 TAMIL LANGUAGE

Tamil is one of the major languages in Dravidian Language Family. It has a rich literature background starting from 300BC. 'Tolkaappiyam' is the ancient book written in the same century. It is mainly spoken in Tamilnadu in India, Singapore, Mauritius, Fuji Island, Sri Lanka and Malaysia.

In Malaysia, Tamil language is spoken among one of the ethnic groups. Tamil became very popular during the period of Tamil Chola Dynasty of 11th Century. Today they form the third largest ethnic group in Malaysia after Malays and Chinese. It constitutes 8% of the total Malaysian population.

When we look into the grammatical categories in Tamil there are about eight. They are Noun, Verb, Adjective, Adverb, Determiner, Postposition, Particles and Intensifiers. These eight grammatical categories constitute six different phrases in Tamil language. The six phrases are:

1. Noun Phrase

2.   Verb Phrase

3.   Adjectival Phrase

4.   Adverbial Phrase

5.   Quantifier Phrase and

6.   Postpositional Phrase

1.3 TAMIL SYNTAX

Syntax is the study of arrangements of words in a sentence. The word order is somewhat strict in languages like English. But in the agglutinative languages like Tamil, the word order is not strict. Though Tamil has a standard word order pattern, said as SOV, it is not strictly followed. Both in spoken and written forms, the word order changes. Consider the following examples.

1. *raaNi pazattaic caappiTTaaL*

'Rani ate the fruit'

2. *pazattai raaNi caappiTTaaL*

'Rani ate the fruit'

3. *raaNi caappiTTaaL pazattai*

'Rani ate the fruit'

4. *pazattaic caappiTTaaL raaNi*

'Rani ate the fruit'

In the above sentences (1, 2, 3 & 4), the word order is SOV, OSV, SVO and OVS respectively. Though the word order of these sentences is changed, the meaning remains unchanged. Out of these four sentences, 3 & 4 are not very common. They are mostly used in spoken and sometime in written when the user wants to stress the object or subject.

The grammarians like Arden (1942), Lehmann (1989) and Kothandaraman (1997) have described that Tamil is a free word ordered language. According to Thambi Jose.F.S (2007) 'Tamil is not free word ordered language, but it is free phrase order language'. Strictly speaking, words within a phrase (except the head) and the phrases in a sentence are freely movable. Consider the following examples.

1. *raaman nalla pazattai caappiTTaan*

'Rama ate the good fruit'

2. *raaman pazattai nalla caappiTTaan*

'Rama good ate the fruit'

In sentence (1), raaman 'Rama' and nalla pazattai 'good fruit' are two noun phrases and caappiTTaan 'ate' is the verb phrase. If, the order of the phrase is changed, then it is an acceptable sentence.
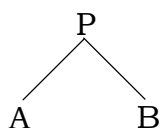
3. *nalla pazattai raaman caappiTTaan*

'Rama ate the good fruit'

If the word order is changed, then it becomes an unacceptable sentence, as given in the sentence (2). We may call the language as a free phrase order language, instead of free-word order language.

Within a phrase, the components of the head can change the order, but the head remains in the same position (i.e. the right most position of the phrase).

1.4 PHRASES

Phrasal category is the combination of lexical categories, that is, the lexical categories such as nouns, postpositions, adjective, adverbs, quantifiers, etc. combine with other syntactic elements to form phrasal categories. When they combine to from the phrasal category, they occur as head constituent in the final position of the phrasal element and their co-elements always precede them. This can be illustrated by a free diagram.
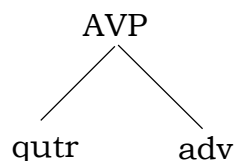
```
        P
       / \
      /   \
     A     B
```

Where 'P' is for phrase (any phrase), 'B' is for any lexical category and 'A' is for the components of the lexical category. The components will be other than the given lexical category.

1.5 ADVERBIAL PHRASE (AVP)

If an adverb occurs as the head constituent of a phrase, then that phrase is considered as an adverbial phrase.

'Adverb can optionally be modified by a quantifier to its left side' (Lehmann, 1989; 165).

```
              AVP
             /    \
          qutr     adv
```

For example:

avar mikavum arumaiyaaka(p) peecinaar

    'He talked very nicely'
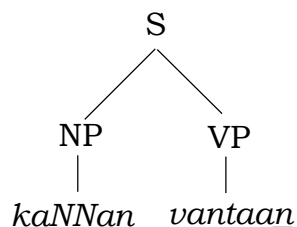
1.6 SYNTACTIC PARSER

In computational syntax, the syntactic parser or analyser breaks a sentence into phrases and clauses and identifies the sentence with the syntactic information.

'The sentence level analysis gives the information about the sentence, like whether it is simple, compound or complex' (Ganesan, 2005; 80).

Grammatical tagging is often seen as the first stage of a more comprehensive syntactic annotation, which assigns a phrase marker or labeled bracketing, to each sentence of a corpus' (Leech, 1993; 277).

The output of the morphological analysis in addition to other (syntax-level) information becomes the input for the analysis. Further, the input, should have word level tagged information such as avan_pro_<PRO> 'he', where 'pro' is pronoun and the angular brackets, '<>' indicate the grammatical category of a word.

The output of the syntactic analyzer is tagged by square brackets '[ ]' for phrase and sentences. It can also represented by the following diagram:

```
                S
              /    \
           NP       VP
           |        |
        kaNNan   vantaan
```

    'Kannan came'

It shows as *kaNNan*<NP_nom> *vantaan*<FV_vp>

Where 'NP' is Noun Phrase, 'nom' is nominative case, 'FV' is finite verb and 'VP' is verb phrase. To identify each and every phrase or a sentence, criteria or mechanisms should be framed and given to the system using programming language. All the phrases are named as modules for the system. When the system identifies <NN> module (noun module), it moves to the noun file and checks for the information related to the given noun and labels it as NP, noun phrase.

1.7 CRITERIA FOR <AV> MODULE

When the parser encounters an adverb <AV>, then it is directed to <AV> module.

(i)   If an adverb <AV> is followed by another adverb <AV>, treat both of them as an adverbial phrase.

1. *tinamum* <AV: (adv)> *koñcaneeram* <AV: (adv)> *paTi* <FV: (vb)>.

'Daily study for sometime'

Since, an adverb *tinamum* 'daily' is followed by another adverb *koñcaneeram* 'some time', [*tinamum* <AV: (adv)> *koñcaneeram* <AV: (adv)>]_{AVP} 'Daily for sometime' is an adverbial phrase.

(ii)   If an adverb <AV> is followed by a finite verb <FV>, mark both of them as a verb phrase and treat the finite verb as head of the verb phrase.

2. *citraa* <NN: (ppn)> *nanRaaka* <AV: (adv)> *paaTuvaaL* <FV: (vb, fu, 3sf)>.

'Chitra will sing well'

In this sentence, [*nanRaaka*<AV: (adv)> *paaTuvaaL* <FV: (vb, fu, 3sf)>]$_{VP}$'s.o. will sing well' is verb phrase, where the finite verb *paaTuvaaL* 'will sing' is the head of the phrase.

(iii) If an adverb <AV> is followed by a non-finite verb <NV>, marked for _vpm_ or _inf_ or _con_, mark both of them as a verb phrase.

3.   *raataa* <NN: (ppn)> *azakaaka* <AV: (adv)> *paaTi* <NV: (vb, pst, vpm)> *paricu* <NN: (ian)> *vangkinaaL* <FV: (vb, pst, 3sf)>.

‌   'Radha sang beautifully and got the prize'

In this sentence *azakaaka* 'beautifully' is an adverb and followed by a nonfinite verb *paaTi* 'sang' and hence, this two constitute a verb phrase [*azakaaka* <AV: (adv)> *paaTi* <NV: (vb, pst, vpm)>]$_{VP\_vp}$ 'sang beautifully'.

1.8 CONCLUSION

The syntactic parser identifies the adverbial phrase of the given text with these three criteria mentioned above. When the same is tested on a corpus, there may be few more rules needed to improve the parsers. Further, the functions of various phrases in a sentence need to be identified, so that they can be used in different applications for understanding or generating various types of sentences.

REFERENCES

1.   Arden, A. H. 1942. *A progressive grammar of the Tamil language*. 4th ed. Madras, India: Christian Literature Society.
2.   Burrow, Thomas. 1968. *Collected papers on Dravidian linguistics*. Annamalainagar, India: Annamalai Univ. Press.
3.   Caldwell, Robert. 1913. *A comparative grammar of the Dravidian or South-Indian family of languages*. 3d ed. rev. Edited by J. L. Wyatt and T. Ramakrishna Pillai. Madras, India: Univ. of Madras Press.
4.   Charles Meyer. 2002. *English Corpus Linguistics: An Introduction*. Cambridge: Cambridge University Press.
5.   Grishman, Ralph. 1986. *Computational linguistics: An introduction*. Cambridge, UK: Cambridge Univ. Press.
6.   Lehmann, Thomas. 1989. *A grammar of modern Tamil*. Pondicherry, India: Pondicherry Institute of Linguistics and Culture.
7.   Mary Dalrymple. 2001. *Lexical Functional Grammar*, volume 34 of *Syntax and Semantics*. New York: Academic Press.
8.   Nitin Indurkhya and Fred Damerau, 2010. editors. *Handbook of Natural Language Processing*. CRC Press, Taylor and Francis Group, 2 edition.
9.   Paramasivam, K. 1979. *Effectivity and causativity in Tamil*. Trivandrum, India: Dravidian Linguistics Association.
10.   Patrick Blackburn and Johan Bos. 2005. *Representation and Inference for Natural Language: A First Course in Computational Semantics*. Stanford, Ca: CSLI Publications.
11.   Thambi Jose, Franklin.S. 2007. *Syntactic Analysis of Tamil: A Computational Approach*, Annamalainagar, India: Annamalai University (unpublished PhD Dissertation).