LANGUAGE IN INDIA

Strength for Today and Bright Hope for Tomorrow

Volume 14:1 January 2014
ISSN 1930-2940

Parsing Noun Inflections: Tamil

Dr. P. T. Kingston and Dr. Samikkanu Jabamoney a/l Ishak Samuel

==========

Abstract

"Parsing is an important process of Natural Language Processing (NLP) and Computational Linguistics which is used to understand the noun inflections of a natural language (NL) sentences confined to the grammar. Parser is a computational system which processes input sentence according to the productions of the grammar, and builds one or more constituent structures which conform to the grammar"[1].Grammar based structural parsing provides solutions to some extent but it is very tedious for larger vocabulary corpus. This paper focuses noun inflections parsing of natural language text in Tamil language.

Features of Tamil Language

"Grammar of Tamil language is agglutinative in nature. Suffixes are used to mark noun class, number and case. Tamil words consist of a lexical root to which one or more affixes are attached. Most of the Tamil affixes are suffixes which can be derivational or inflectional. The length and extent of agglutination is longer in Tamil resulting in long words with large number of suffixes"[2].

In Tamil, nouns are classified into rational and irrational forms. Human comes under the rational form whereas all other nouns are classified as irrational. Rational nouns and pronouns belong to one of the three classes: masculine singular, feminine singular and rational plural. Irrational nouns belong to one of two classes: irrational singular and irrational plural. "Suffixes are used to perform the functions of cases or post positions. Tamil verbs are also inflected through the use of suffixes. The suffix of the verb will indicate person, number, mood, tense and voice. Tamil is consistently head-final language"[3]. The verb comes at the end of the clause with a typical word order of Subject Object Verb (SOV).

However, Tamil language allows word order to be changed making it a relatively word order free language. Other Tamil language features are plural or honorific noun, frequent echo words, and null subject feature i.e. not all sentences have subject, verb and object.

Parsing Noun Inflections

Parsing is important in Linguistics and Natural Language Processing to understand the noun inflections of a natural language grammar. "Parser is a computational system which processes input sentence according to the productions of the grammar, and builds one or more constituent structures called parse trees which conform to the grammar. Parsing natural language text is challenging because of the problems like ambiguity and inefficiency"[4]. A parser permits a grammar to be evaluated against a

potentially large collection of test sentences, helping the linguist to identify shortcomings in their analysis.

Noun and verb are the primary grammatical categories. Suffixes of noun and verb are the secondary grammatical categories. There are two kinds of suffixes:

1. Derivational suffixes
2. Inflectional suffixes.

Derivational suffixes change the noun with verb vice versa

1. (a) *Atu + al* *ATal*
dance + ing *dancing*

(b) *OTTu + nar* *OTTunar*
Drive + er *Driver*

ATu, *OTTu* are the verbs. Suffixes *–al*, *-nar* change the verbs with noun.

2. (a) *nI* *ennaik* *KAtali*
you me love
you love me

Suffix *-ai* changes the noun *KAtal* into verb. *–al*, *-nar*, *-I* are the Derivational Suffixes.

(b) *nI* *en* *KAtali*
you my lover
you are my lover

Here suffix *-i* does not change the noun *KAtal* in to verb. So this suffix *-i* is an Inflectional suffix. We should not parse the suffixes for machine translation. *ATal, OTTunar, KAtal* should appear in dictionary as they are.

We do parse the tense marker and pronominal termination from the inflected forms of verbs.

3 (a) *paTi - tt - An*
read – past- he
He read

(b) *ezut – in - AL*
Write - past – she
She wrote

Suffixes *–tt-An, -in-AL* leave from the verbs *paTi-, ezut-,* similarly pluralarker, increment, case marker and ditties come under parsing when nouns are taken into account. All these come under Inflectional suffixes. Examples are given from 4 to 7 to indicate these types under noun inflectional suffix.

4 (a) *MATu - KaL*
Bullockcart – plural
Bullockcart

(b) *mANavar - KaL*
Student - plural
Students

5 (a) *vITT – iR - Ku*
Home – increment- Dative case
To home

(b) *nila - tt - ai*

Land – increment – accusctive case

Land

6 (a) *nATT - ai*

Country – accussative case

Country

(b) *nATT - uKKu*

Country - Dative case

To country

7 (a) kizaKK - um

East - also/ and and east

(b) *nI - tAn -E*

You only - definitely

You are definitely

In these examples inflectional suffixes are *–Kal, -iR, -tt-, -ai, -uKKu, -tAn-, -E* these four kinds of suffixes are noun inflections.

While there is more than one suffix occurring after noun, plural marker, increment, case marker and clitics come in that order respectively.

8 (a) *mana - tt - iR - ku*

Mind - increment 1 – increment 2 – dative case to mind

(b) *KizaKK - il – E – Y – um – tAn – E*

East – pcatiracok – definitely – guide – and – also – definitely in the east also.

More than two increments come together. Up to four clitics come together.

Noun	Plural marker	Increment – 1	Increment - 2	Case marker	Clitics - 1	Clitics - 2	Clitics - 3	Clitics – 4

So, we hold it for the structure of noun inflections.

We do restrict the meanings and the number of the noun inflections. We can avoid the suffixes *–til, -man, -kol* etc. As these suffixes are not in modern Tamil. The case marker *-in* occurs with various meaning in old Tamil.

9 (a) *KANunar inmai – y – in cettanaL (aKam .16 – 8)*
Viewer / lover without – cause died – she

(b) *yAnaiyum malaiy – in tOnRum (puram 42 – 2*)

Elephant mountail – like look
Elephant look likes mountain

(c) *malai – y – in izintu mAkkaTal nOkki (puram 42 – 19)*
Mountain – from big sea towards

(d) *yAnai – y – in vantu ninranai (KuRunt . 161 – 7)*
Elephant – on came stand – you
You came and stood on elephant

(e) *Kuruamp – in - aruntum (puRam – 384)*
Sugar came – of flower
SUCK the flower of sugar cane

Reason, causative, comparative, adjectives, locative and genitive and the various meaning of case marker. *–in.* Now we neglect all the case meanings except genitive meaning. In modern Tamil suffix *-in* comes with this meaning.

We relax the rule one meaning for one suffix. String *ai* functions as noun, case and increment.

10 (a) *enn / ai / KKu Ur iHtu anmai Ayinum*
My lover – to Town it is not it
If the town is not

(b) enn / ai nil / anmin tevvir

In these old Tamil sentences noun *ai appear* with the meanings of lover, king, husband, leader, God etc. In modern Tamil, increment *ai* appears.

11 (a) *ippOt – ai – Kku*
Now – increment – dative case

(b) *nERR - ai - KKu*
Yesterday – increment – dative case

ai is a case marker in the long history of Tamil language. The noun ai comes as a seperarte word. Increment ai comes between noun and case. We consider the three types of ai forms for machine translation.

We restricted the various meanings of (CM) case marker *–in*. But we do not restrict the increment *–in* (example No: 13) and CPM conditional participle marker *–in* (example No: 12). Three important differences are between them.

12 *mazai peyy – in Kulam niraiyum*
Rain rains – CPM tank will fill

If it is raining tank will fill

13. *coll – in celvan*
Word – possessive case sin son of word
Son of word

The differences are: 1. CPM comes after verb. CM comes after noun 2. CPM appears in complex sentence. CM appears in adjectival phrase 3. CPM modified by verb or predicate. CM modified by a noun. This is easy way to identify them. Here syntactical levels also help as for parsing the inflectional suffixes.

Case marker comes after noun only. Tense marker comes after verb only. Plural marker and clitics come after noun and verb. While more than one suffixes coming after noun. We consider the suffixes as a single string. Firstly we do separate the whole suffixes from the noun.

14. *Kutirai – KKA kompu muLaittirukkiratu*
Horse – for horn growing
Is horn growing for horse?

We treat the *–KKA* as a one string at the time of separating it from the noun Kutirai, horse. Then we park the string *-KKA* into case marker *-KKu* - and question marker

–A. This method gives an easy way in processing of the parsing. We do make many tables to explain the combination of inflectional suffixes.

15. KizaakkEyumtAnE mazai peykiRatu
East – and – also – definitely markers rain raining
It is raining in the east also

In this example four suffixes, *E – y – um – tAn – E* come after noun *Kizakku*, east. The places consider the initial, medieval and final positions of the inflectional suffixes after noun.

If rest of the first letter of noun and suffix are in the same form, the computer parses the word wrongly.

15. *mATu – KaL*
    - Plural

16. *makaL*
    Daughter

In the example 16 – *KaL* is the part of the word.

17. *nila – tt – ai*
    Land – increment – accusative case

18. (a) *ezu - tt – ai*
    Raise – increment – accusative case

    (b) *ezutt - ai*
    Letter – accussative case

ezu raise is a verb. Case marker ai should not come after verb. Example 18 a is wrong. *Ezutt* is the alternate form for the noun *ezuttu*..

Selectional restriction rule functions between alternate forms of case markers and alternate forms of nouns.

19 (a) *muLL - uKKu*

Thorn - dative case

(b) *ill – ukku*

Home – dative case

muLL ill are the alternate forms of nouns *muL, il*. *uKKu* is one of the alternate form of dative case *Ku*. The noun pattern CVCC – and VCC – select – the uKKu form.

There are four kinds of alternate forms to noun.

I) a noun stands without change

II) One of the alternate forms stands for noun

III) A noun stands with guide (y, v)

IV) A noun stands with germination of its final letter.

For example there are three alternate forms for the noun *kATu,* forerst as

20. (a) *kATu – KaL*

Forest – plural marker

(b) *KATT – ai*

Forest – accusative case

(c) *KAT - um*

Forest – and

Here KATu, KATT, KAT are the three alternate words.

21. (a) *manu – V- in – ai*

Petition – glide – increment – accusative case

(b) *mETai – y – il*

Dais – glide – locative case

V, y is the glides. The nouns stand with glides. In examples 19 final letter of the noun germinates.

There are some words common for noun and verb. Similar there are common suffixes between inflectional suffixes of noun and verb. While these words coming together we cannot parse them,

22 (a) *pATTu icaiKKum nEram etu*
Song singing time which

Which is the time for singing a song?

(b) *icaiKKum - muKKiyam taraVENTum*
Music – also important give
Give important to music also

(c) *paravai kAlaiyil pATTuicaiKKum*
Bird morning-locative case song sings
Bird sings song in the morning

In the above examples computer cannot – parse the word *icaiKKum*.

Some strings are common to noun and post positional phrase

23. (a) *KaTal - ai*
Sea – accusative case

(b) *KaTalai*
Ground nut

Conclusion

Machine cannot identify the deep cases. In these contexts, manual editing is possible.

References

1. Selvam. M. et al., (2009) "Structural Parsing of Natural Language Text in Tamil Using Noun Inflection Structure using Hybrid Language Model" International Journal of Computer Processing of Languages, Volume 22, Issue 02n03, June & September 2009.

2. Selvam. M et al., (September 2009) "Structural Parsing of Natural Language Text in Tamil Using Noun Inflection Structure using Hybrid Language Model" International Journal of Computer Processing of Languages, Volume 22, Issue 02n03, June & September 2009.

3. http://en.wikipedia.org/wiki/Tamil_language

4. Selvam. M. et al., "Lexicalized and Statistical Parsing of Natural Language Text in Tamil using Hybrid Language, Models http:// www.researchgate.net/publication/239752471_Lexicalizedand_Statistical_Parsing_of_Natural_Language_Text_In_Tamil_Using_Hybrid_Language_Models.

==========

Dr. P. T. Kingston
Senior Lecturer
Fakulti Bahasa dan Komunikasi
University Pendidikan Sulthan Idris
35900 Tanjong Malim
Perak Darul Ridzuan, Malayasia
fkingston@gmail.com

Dr. Samikkanu Jabamoney a/l Ishak Samuel
Senior Lecturer
Fakulti Bahasa dan Komunikasi
University Pendidikan Sulthan Idris
35900 Tanjong Malim
Perak Darul Ridzuan, Malayasia
samjabarose@yahoo.com.my