==========
Language in India www.languageinindia.com **ISSN 1930-2940 Vol. 16:1 January 2016**
==========

Dependency Framework for Marathi Parser

Yogesh Vijay Umale

==========

Abstract

This paper describes the Framework of Dependency Grammar for Marathi Parser. Dependency grammar is a grammar formalism, which is a capture direct relations between word to word in the sentence. The parser is a tools, which is automatic analysis sentence and draw a syntactic tree of sentence. The grammar formalism is mechanism to developing parser. Today in filed of computational linguistics, natural language processing and artificial intelligent have two kind of grammar formalism which is Phrase structure grammar and Dependency grammar. Both grammar formalism have their own limitation to developing a parser. In this paper I will use computational Panini grammar approach of dependency grammar. Computational Panini grammar has 37 dependency tag-set and those tag-sets are useful to annotate the Indian languages such as Hindi, Telugu and Bangla. However, I have to examine those dependency tag-set to Marathi and annotate a corpus which is useful to develop a Marathi parser. To annotate data I have use an Anno-Corp Guidelines, which develop by IIIT, Hyderabad. According to guidelines the relations are three types *karaka* relations, which is mark as k1,k2,k3,k4,k5 and k7, non-*karaka* relations which marked as r6,r6-k1,r6-k2,rt,rd,rh,ras_k*, adv, and other relations such as relative clauses.

Key words: Marathi, Parser, Dependency Framework, Corpus Annotation.

Introduction

The Parser is tools which used to analysis the sentence in term of its constituent parts. A parser aims to generate automatic syntactic trees of natural language. In filed of computational linguistic, natural language processing language and artificial intelligent have two kind grammar formalism which phrase structure grammar and dependency grammar. Those two grammar mechanism are useful to develop a Parser. Today English language have phrase structure

grammar formalism and dependency grammar formalism to develop parser and those two grammar formalism are provide good accuracy. When we apply those two grammar formalism to Indian languages, than we can see dependency grammar is provide good accuracy compared to phrase structure grammar. The reason is simple, English language have positional word order structure and most of the Indian languages have free word order structure and morphological rich."Development of a parser is a challenging task for morphological rich and free word languages such as Indian languages. Dependency grammar formalism is suitable and useful for Indian languages"(Bharati, et al, 1995).

Dependency grammar formalism have different approaches and different tag-set. Those approaches and tag-sets are may be change depend on language parameters. Indian languages have panini dependency grammar approach and tag-sets which is karaka relations (k1,k2,k3,k4,k5 and k7), non-karaka relations (r6,r6-k1,r6-k2,rt,rd,rh,ras_k*, adv,) and other relations (ccof, frgm, null etc).

Methodology

To data collection I used two Marathi grammar books and collected 500 sentences. Those 500 sentences I used as corpus. The corpus annotation I used **3A** Approach which refers to corpus Annotation, corpus Abstraction and corpus Analysis. After that I used Panini dependency approach and tag-set which developed by IIIT Hyderabad for Indian Languages such as Hindi, Telugu and Bengali. This panini dependency approach and tag-set I applied to Marathi and find out result.

Data Analysis and Interpretation

1 *karaka* Relation

The dependency grammar formalism captures the direct relation between word to word in the sentence. The case (*karaka*) shows a direct relation between nouns to verb. Marathi has six *karaka*, nominative, accusative, instrumental, dative, ablative and location. According to the dependency guidelines, I marked them as *k1, k2, k3, k4, k5* and *k7*.

1.1 Karta (dependency tag-set *k1*)

1.1.1 Nominative Subject

Most of the time the nominative form takes a syntactic and sometime it takes semantic function as *karta* (agent). The *karta* plays a major role in sentence which is doing or performing the action. Consider the following examples.

surēśa *pustaka* *vāca-tō*
suresh-nom-3msg *book-accu* *read-pres-3msg*
Suresh reads a book

Here Suresh is *karta,* Suresh performing the action *vāca-tō,* and *vāca-tō* is a transitive verb. So here verb has two arguments which is subject (*karta)* and object (*karma*). In intransitive verb does not require object. Consider the following example.

sacina *basa-lā*
sachin- nom sat-past-3msg
Sachin sat

Here the first example is transitive and the second one is intransitive verb. Both subjects are nominative with zero suffix (zero *vibhakati*). Both subjects are in agreement with verb like gender, number and parson. Here both the subject forms are marked as *k1*.

1.1.2 Ergative Subject

Ergative subject occurs with *ne* or *ni* postposition in Marathi. In this contraction ergative subject does not show agreement feature with verb. Consider the following example.

surēśa-nē *cēṇḍū* *phēka-lā*
Suresh-erg ball-3msg throw-past-3msg
Suresh throws the ball

Here the ergative subject construction takes a *ne* case marker but the agreement feature show with karma which is cēṇḍū, here this relation we marked as k1.

1.1.3 Dative Subject

The dative subject in Marathi takes *_lā* case marker and does shows agreement with verb, see the following example,

Surēśa-lā āmbā kha-llā pāhijē
suresh-dat mango-acc-3msg eat-impl.3msg should
Suresh should eat a mango

In this construction syntactic subject is *āmbā* because verb has agreement with *āmbā* but semantically *surēśa-lā* is subject so we marked as *k1*

1.1.4 Subject in Passive Construction

Subject in passive construction show by *kadun* and *dvara* case marker, in this construction *kadun* and *dvare* postposition block agreement feature with verb, consider the following example,

surēśa-kaḍuna/dvārē *āmbā* *khā-llā gēlā*
suresh- by mango-msg eat-ptcp-pass-past gone
Mango was eating by Suresh

Here *surēśa* is subject but that subject does not agree with verb, so we can mark as k1.

1.2 karma (dependency tag-set *k2*)

1.2.1 Accusative

The accusative (Karaka) object in Marathi takes *_0, _sa* and *_lā* case marker

surēśa ***pustaka*** *vāca-tō*
suresh-nom-3msg book-acc read-pres-3msg
suresh reads book

pōlisa ***cōra-lā/-sa*** *māra-tō*
Policeman-nom-3msg thief-acc beat-pres-3sm

The policeman beats the thief

Above both examples are shown relation with verb as object because they takes case masker *_0, _sa* and *_lā* as well as those construction does not show a agreement Patten with verb so we can marked them as *k2*.

1.2.2 Object in Passive Sentence

In passive construction object (*karma-karaka*) control agreement with verb and takes *_0, _sa* and *_lā* case marker consider the following example,

pōlisāṅ-kaḍūna ***cōra*** *pakaḍalā gēlā*
policeman-by thieves-acc-3mpl catch –past-3mpl go-pass-past-3mpl
The thieves were caught by policeman

pōlisāṅ-kaḍūna ***cōra/sa/lā/nāṁ*** *pakaḍalē gēlē*
policeman-by thieves-acc catch-past- 3nsg go-pass-past-nsg
The thieves were caught by the policeman

When the passive construction occurs in the sentence then we marked object as *k2*.

1.3 karaNa (Instrument) (dependency tag-set *k3*)

Instrument (*karaka*) case marker takes a *_ne* postposition. The instrument *_ne* case marker express function as instrument with verb,consider the following example,

surēśa-nē ***cāku-nē*** *āmbā kāpa-lā*
suresh-erg knife-inst mango-3msg cut-past-3msg
Suresh cut mango with a knife

Above example shows instrument relation with verb so that relation we can mark as *k3*.

1.4 sanprdan (Recipient/Beneficiary) (dependency tag-set *k4*)

Recipient (Karaka) case marker expressed recipient or beneficial meaning of the verb. In term of syntactic category we can called as indirect object but in dependency tag-set we called them recipient karaka. Consider the following example,

Surēśa-nē ***sacina-lā*** *pustaka dilē*
suresh-erg sachin-dat pustak gave-past-3msg
Suresh gave book to sachin

tyā-nē ***dēśā-sāṭhī*** *jīva dilā*
he-ag country-for life give-3-msg
He gave (his) life for his country

The above construction *–lā -sāṭhī* both are the case marker as well as postposition. In this construction we mark them *k4*.

1.5 aapadan (Source) (dependency tag-set *k5*)

The source *karaka* expresses a meaning of separation and point of departure with verb. Source (*karaka*) case marker takes *-kaḍhuna –hūna***,** see the following example,

malā ***surēśa-kaḍhuna*** *bātamī kāḍha-lī*
I-dat suresh from newfindout get-psat-3fsg
I got new from suresh

surēśa ***mumbaī- hūna*** *ālā*
Suresh-nom Mumbai-from come-past-3msg
Suresh came from Mumbai

The above examples, *-kaḍhuna* and *–hūna* case markers provide us a meaning of separation and departure so here we mark them as k5.

1.5.1 Source of Material

In this construction verb denoting source of material meaning in the sentence, now see the following example,

kapaṛē ***kāpasā-pāsūna*** *bana-tāta*
cloth-nom-3pl cotton-from make-hab-be-presp-3pl
Cloth are made from cotton

In the above sentence *kāpasā –pāsūna* is the natural source and it gives the source indication by postposition *-pāsūna.* In this construction we mark this relation as *k5*.

1.6 adhikarana (Location of Time) (dependency tag-set *k7t*)

The time location is express by tense like yesterday, tomorrow, now etc. a postposition like *-lā* and *-ta* also express a meaning of location. Consider the following example,

mī ***kāla*** *mumbaī-hūna* *ālō*
I-1msg yesterday Mumbai-abl come-past-1msg
Yesterday, I came from Mumbai

Here time gives a meaning of location, so here we can mark this relation as *k7t*.

1.6.1 Location of space (dependency tag-set *k7p*)

Space location expressed by locative suffix of *–ī* and*–ta* and postposition of *madhyē*. Consider the following example:

tō āja ***gharī/gharāta*** *nāhī*
he today home-loc-at/home-loc-in neg-3sg
he is not at home/in the house today

tyā-nē ***rastāta/madhyē*** *gāḍī thāmbavalī*
he-ag street-in the middle of car-3sgf stop-past3sgf
he stopped the car in the middle of the street

This construction we can mark as *k7p*.

1.6.2 Location of elsewhere (dependency tag-set *K7*)

The location is expressed the mental place and take same locative suffixes *–ī* and*–ta* which is follow by noun of location, consider the following example,

mājhyā ***manā-ta*** *rāga āhē*
my mind-in a nger is
I am anger in mind

mājhē māna mumbaī-ta āhē
my mind Mumbai-in is
I am mentally in mumabi

Here *–ī* and*–ta* give a meaning of location, so here we can mark it as *k7*.

B.2 Non-karaka relation

The non-karaka relations depend on the noun. The non-karaka relations capture the direct relation between noun to noun in the sentence. They do not show direct relation with verb.

2.1 shashti (Genitive /possessive) (dependency tag-set r6)

The genitive or possessive relation which holds between two nouns has to be marked as r6 consider the following example:

mulā-cē *nāka*
boy-of nose
Nose of boy

līlā-cī *bahina*
lilaa- of sister
Sister of Lila

Here the postposition ***-cē*** and ***–cī*** provide a meaning of genitive as well as possession. Here we can mark this relation as r6.

2.2 genitive/possessive relations with conjunct verb (dependency tag-set *r6-k1, r6-k2*)

A conjunct verb is composed of noun or adjective followed by verbalize. Sometime the argument (*karta or karma*) come with genitive case. Whenever the argument of conjunct verb is in genitive case it will have a dependency relation with the noun of conjunct verb. The class of conjunct verb (a noun+verb sequence which functions as a single verb unit) is very large in Marathi. Consider the following example:

kāla ***mandira-cē*** *udaghāṭana jhālē*
yesterday temple-of inauguration happed
yesterday the temple got inaugurated

mī *rōja* *rātri* ***parīcī*** *pratīkṣā kara-tō*
I-1msg everyday night-loc angle-poss waiting do-1msg
I wait of angle everyday night

In this above construction we can mark dependency relations as *r6-k1* and *r6-k2*.

2.3 Adverbs only manner (dependency tag-set *adv*)

Adverbs of manner are expressed which are placed immediately preceding the verb. Adverbs of manner are marked as adv. Consider the following example:

surēśa ***bharābhara*** *cālatō*
suresh fast walk-pres-3msg
suresh walks fast

In this construction adverb, we would mark it as adv.

2.4 Purpose (dependency tag-set *rt*)

The purpose is expressed by dative case marker *–lā* and postposition *-sāṭhī* use in sentence. Consider the following example:

tō amērikē-ta ***śikaṇyā-sāṭhī/lā*** *gēlā*
he America-loc study-dat go-past-3msg
He went to America to study.

tō ***kuṭumbā-sāṭhī*** *kaṣṭa karatī*
he family-for- hard work do-pres-3msg
He works hard for the sake of (his) family.

In above examples *–lā* and *sāṭhī* we would mark dependency relation as *rt*.

2.5 Direction (dependency tag-set *rd*)

The label rd stands for relation direction. In Marathi postposition *-kaḍē* express a meaning of direction. Consider the following example:

surēśa ***gāva-kaḍē*** *jāṭa hōtā*
suresh village-towards go-prog be-past-3msg
Suresh was going towards his village

The participant indicating 'direction' of the activity has marked as '*rd*'.

6 Reason (dependency tag-set *rh*)

The reason or cause of activity is to be marked as *rh*. Consider the following example:
Surēśa-nē ***mōhana-muḷē*** *pustaka vikata ghē-ta-lē*
suresh-erg mohan of because book bought- past-3msg
Suresh bought book because of Mohan

In this construction *-muḷē* postposition provides a meaning of reason or cause, so here we can mark this dependency relation as *rh*.

7 Associative (dependency tag-set *ras_k)**

Where two participants perform the same action but syntactically one is expressed as primary and other as semantically associated. So, we would marked the *ras_k** consider the following example,

surēśa āpalyā ***vaḍilā barōbara*** *gārī gēlā*

suresh own father with home went-past-3msg

suresh went to home with his father

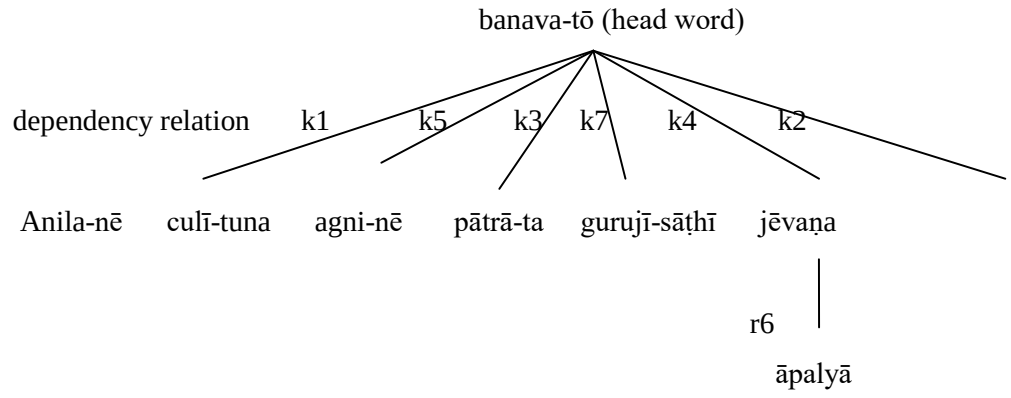
In the above the example *barōbara*show has the meaning of associative, so here we would mark this relation as *ras_k**.

Tree of Dependency Framework for Marathi

anila-nē culī-tuna agni-nē pātrā-ta āpalyā gurujī-sāṭhī jēvaṇa banava-tō

anile-nom. from furnace-abl. by the fire-inst. in a vessel-loc. for his master-dat. food-3msg cooks-3msg

Anil cooks food in a vessel by the fire from the furnace for his master



3 Other Relations

In other relations, dependency is captured as direct relation between clauses to clause. Marathi has two types of clause; one is sentential clause and other is participle clause. In this paper I have explain only sentential clause.

3.1 Pre-nominal relative clause (Dependency tag-set nmod_relc)

In this construction relative clause occur with left of head noun and it take a relative pronoun *Jō* as a demonstrative marker *tō* along with noun, consider the following example,

Jō māṇūsa *yēthē śikavatō* ***tō*** *(Ø)mājhyā bhā'ū āhē*
rel man here teach-pres-3-sm cor (man) I-poss-3-msg brother is
The man who teaches here is my brother

Here this dependency relation we would mark as nmod_relc.

3.2 Pronominal Relative Clauses

In this construction the relative clause come to the right of head noun and relative pronoun in such case behaves like a full-fledge pronoun consider the following example,

jō māṇūsa *yēthē śikavatō* ***tō māṇūsa*** *mājhyā bhā'ū āhē*
cor man rel here teach-pres-3sm I-poss-3sm brother is
The man who teaches here is my brother

Above construction is pre-nominal and *Jō* is modifying of main clause with *tō*. *tō* itself refer to *Ø* (*māṇūsa)* which came with relative subordinate clause and clause along with the relative pronoun ***tō***. Here we can mark as *nomd_relc.*

Here *jō māṇūsa* which is a subordinate clause refers to main clause, which is *tō māṇūsa***.**

C. Conclusion

The above dependency tag-set provides us linguistic information such as syntactic and semantic. Above analysis method also provides us dependency relation in terms of word to word relations in sentences. Today in computational linguistics, we need this kind knowledge for annotate a language corpus and depending on annotated corpus we would develop a Parser.

==========

References

Bharathi, Chitanya, and Sangal,R. 1995. Natural language processing: a paninian perspective. New Delhi: Prentice-hall of India.

Damale, M.K. 1911. *Shastriya Marathi vyakaraN* . Pune: Deshmukha and Company.

Fillmore, C.J. 1985. *The case for case.* In E.bach and R.T. Harms (ed). Universal linguistics theory. New York: Holt Rinehart and Winston.

Navalkar, G.P. 2001. *The student Marathi grammar.* New Delhi: Asian education Service.

Nivre, J. 2013. *Dependency grammar and dependency parsing*. http://stp.lingfil.uu.se/~sara/kurser/5LN455-2013/lectures/5LN455-2013-12-11.pdf.

Pandharipande, R. 1997. *Marathi*. Ladoan and New York: Rutledge.

Uma, Maheshwar R. and Kulkarani, A. 2007. *Natural language and computing*. PGDCAIL. vol.411. CDE: University of Hyderabad.

Uma Maheshwar Rao G., K. Rajya Rama, A. Srinivas. 2012. *Dative case in telugu: a parsing perspective.Proceedings of the Workshop on Machine Translation and Parsing in Indian Languages.* Mumbai: *(MTPIL-2012), pages 123–132,*COLING 2012.

Valanbe, M.R. 2012 *Sugam Marathi vyakaraN*. Nitin. Pune: Nitin Prakashan.

Wali, K. 1997. *Marathi : a study in comparative south asian languages*. Delhi: Indian institute of language studies.

==========

Appendix

Set of dependency labels:

S.No	Labels	Description(Relations)	Gloss/Additional
1	k1	karta	doer/agent/subject
2	k2	karma	object/patient
3	K3	karana	instrument
4	k4	sampradana	recipient
6	k5	apadana	source

7	k7t	kAlAdhikarana	location in time
8	k7p	deshadhikarana	location in space
9	k7	vishayadhikarana	location elsewhere
11	r6	shashthi	genitive/possessive
12	r6-k1, r6-k2		karta or karma of a conjunct verb (complex predicate)
13	r6v	kA	relation between a noun and a verb
14	adv	kriyAvisheSaNa	adverbs - ONLY 'manner adverbs' have to be taken here
15	Sent-adv		Sentential Adverbs
16	rd	relation prati	direction
17	rh	hetu	reason
18	ras-k*	upapada_sahakArakatwa	associative
19	nmod__relc, jjmod__relc, rbmod__relc		relative clauses, jo-vo constructions

Yogesh Vijay Umale
Ph.D. (12HAPH06)
Center for Applied Linguistics and Translation Studies
School of Humanities
University of Hyderabad
Hyderabad 500046
Telangana
India umaleyogesh1@gmail.com