

Syntactico- semantic subject (karta) in Bangla

Pragati Dhang^a, Sanjay Chatterji^a, Tanaya Mukherjee Sarkar^a,

Sudeshna Sarkar^a, , Jayashree Chakraborty^b Anupam Basu^a

^aDepartment of Computer Science and Engineering, Indian Institute of Technology, Kharagpur

^bDepartment of Humanities And Social Sciences, Indian Institute of Technology, Kharagpur

Abstract

Identification of subject (karta) in the Indian languages has been a difficult task. This is mainly because of the varied use of bibhakti markers, there being no one to one correlation between the semantics of a marker and the corresponding context in which it is used. The present paper identifies karta in Bangla sentences based on modern Bangla grammar. When the task is assigned to a computer through the dependency structure of sentences, it prefers syntactic dependencies. Other than karan and apadan, all the karakas take same case markers or suffixes or bibhakti markers. The concentration is in the identification of karta based on logical and syntactic (syntactico-semantic) point of views.

Keywords : Syntactico-semantic relation, Modern Bangla grammar and language

1. Introduction

In Bangla, subject (karta) is one of five karaks. The subject (karta) of a Bangla sentence is a noun phrase which indicates a person or a thing that performs an action, experiences something, undergoes an action or exists somewhere. Karta takes the different bibhakti markers

like ‘te’(‘তে’), ‘ke’(‘কে’), ‘e’(‘এ’), ‘aYa’(‘য়া’) and ‘ra’(‘রা’) which are also shared by other karaks. In Bangla, the karta may take a wide range of bibhakti markers or suffixes like nominative, genitive, accusative and locative and thus there is no simple rule to identify the subject (karta) based on the bibhakti markers only. In Bangla, it is difficult to analyse karta based on purely syntactic features. In several contexts, it is found that karta takes some other bibhakti markers which indicate other types of classification of karak relations. For the purpose of identification of the subject, the Paninian grammatical model [2] and some other modern grammar books [1,4] have been studied.

We have used syntactico-semantic features for identifying karta. Syntax indicates the characteristics of sentence, the order or sequence of words and agreement of words in the sentence and semantics indicates the meaningful classification of words [1]. Syntax indicates the grammatical features of a sentence, whereas semantics indicates its meaning. The term syntactico-semantic indicates both syntax and semantics. Since syntax markers do not distinguish the true picture of karta of the Bangla sentence, semantics needs to be also used to identify the karta and its role. There are instances of kartas in the same syntactic or semantic role with different bibhakti markers as well as kartas of different syntactic or semantic roles with same bibhakti marker. A few examples of each of these kartas are in order.

For all examples we have used Bangla font, itrans, gloss and its English translation respectively throughout the paper and the bold words indicate karta. The abbreviation of gloss is written as Genitive-gen, locative- loc, accusative- acc, present tense- pres, past tense- past, perfective- perf, progressive- prog, participle- ppt, future- fut, negative- neg.

Some examples of same syntactic role with different bibhakti markers:

Karta with nominative marker.

1. **রাম** ভাত খেল।

rAma bhAta khela.

Ram rice eat-past.

Ram ate rice.

Karta with locative marker.

2. **মানুষে** কথা বলে।

mAnuShe kathA bale.

Men-loc word speak.

Human beings speak.

3. **রাজায় রাজায়** যুদ্ধ করে।

rAjAY rajAY yuddha kare.

King-loc king-loc fighting do.

The king fights with a king.

In the above examples, all the kartas indicate the same syntactic role but have taken different bibhakti markers.

Some examples of same semantic role and different bibhakties

Karta with nominative marker.

4. আমি গান ভালবাসি।

Ami gAna bhAlabAsi.

I song love

I love song .

Karta with genitive marker.

5. আমার খিদে পেয়েছে।

AmAra khide peYechhe.

I-gen hunger get.

I am feeling hungry.

In the above examples, all the kartas indicate the same semantic role but have taken different bibhakti markers.

Some examples of different syntactic roles and same bibhakti

6. আমার শীত করছে।

AmAra shIta karachhe.

I-gen cold do-pres- prog.

I am feeling cold.

7. এই ভাবে **তারার** সৃষ্টি হয়।

ei bhAbe **tArAra** sriShTi haYa.

This way star-gen formation be-pres.

In this way the stars are created.

8. **আমার** খাওয়া হয়নি।

AmAra khAoYA haYani.

I-gen eating be-pres neg

I have not eaten.

In the above examples, the words in bold font have the same syntactic features but their roles are different. To identify their roles semantics must be used. Here, the words **AmAra**, **tArAra** and **AmAra** indicates feelings, undergoes an action and performing an action respectively. So, both syntax and semantics are used to identify the karta as well as the role of karta which indicates the different types of karta.

In a sentence, voice (bachya) plays an important role to identify karta. In active voice (kartribachya) karta of a sentence is identified on the basis of the doer of the action, on the basis of semantics of verbs and on the basis of the complement of the subject where both subject and its complement refers to the same entity. In passive voice (karmabachya) karta of a sentence is

identified as the subject which is the active agent of the sentence and is generally followed by postpositions like ‘dbArA’, ‘diYe’, ‘kartRRika’ etc. This subject may or may not take genitive (‘ra’) marker. In neuter voice or impersonal voice (bhabbachya), it is difficult to identify the exact karta of a clause, as agent or doer subject is silent or acclinated, it does not come in the surface. Here action is emphasized and plays an important role. According to many grammarians, neuter voice or impersonal voice (bhabbachya) is identified by the absence of object and the verb is usually intransitive. This karta takes accusative and genitive markers [1].

The paper is organized as follows; - section 2 presents Objective, section 3 defines Karta and Classification of karta and its analysis, section 4 presents Discussion and finally section 5 presents Conclusion.

2. Objective

Two processes are used to annotate the different types of karta in Bangla sentences. The first process is to identify the karta and the second one is to classify the karta based on their roles. For annotating kartas of different types, the karta of a sentence is identified and classified on the basis of syntactic and semantic features. The semantic features of verb used are the natures of the verb – volitional or non-volitional, features of obligation or necessity attached to the verb etc. These semantic features determine whether the karta will be classified as agent or experiencer or noun of proposition and so on. The features are indicated on the karta by syntactic markers like suffixes or vibhakti markers, postpositions and also morphological features like TAM etc. For identification and classification of Karta, syntactic informations as well as

semantic informations are used. In this case, syntactic informations indicate undergoing the action done by an implied agent and semantic informations indicate the type of effort, obligations in doing action.. Semantic informations include animacy of the noun phrase and the nature of the verb with which it is attached.

3. Related Work

Several works have been done in defining dependency grammar for different languages. Most of them are used in annotating data. Penn Treebank for English [8], Prague Dependency Treebank for Czech [9] etc. are some of such important efforts. Most of these prior works are inclined to syntactic properties of the constituents.

Recently some works have been done on building dependency annotation for Indian languages. Sharma et al. [5] have created a Hindi dependency Treebank named “AnnCorra” using Hindi dependency grammar. They have used Panini's grammatical model [2] which provides syntactico-semantic analysis to some extent.

There are many similarities in karta, we have tagged in Bangla and Sharma et al. [5] tagged in Hindi. We have classified karta in five categories namely, Kriya sampadak karta (doer subject), Anubhab karta (experiencer subject), Paroksha karta (passive subject), Bidheya karta or Samanadhikaran (noun of proposition) and Sadharan karta (general subject) whereas they have classified it in six categories namely, karta (doer), prayojaka karta (causer subject), prayojya karta (causee subject), madhyastha karta (mediator causer), karta samanadhikarana (noun

complement of subject) and clausal subject. In Bangla treebank, causative subject (prayojaka karta), mediator causer (madhyastha karta), causee subject (prayojo karta) are not used. The detailed study will be available in “A dependency annotation scheme for Bangla Treebank” [10].

4. Karta (Subject)

Karta is an active agent of the activity implied by the verb which indicates physical or mental exercise. It will be animate noun or personal pronoun. Karta is the experiencer or a perceiver who does not take any active effort. It indicates some mental state, emotion, desires or event. Karta is a person or thing which undergoes some action, has obligation on the action, exists somewhere, and indicates the complement of other karta. Karta may also ‘be’ or ‘become’ something. It may take any suffix. We classify karta into some subclasses.

3.1. Classification of Karta

Considering the syntactic and semantic informations, the following roles of karta are important. Based on these roles karta is classified into five finer divisions-

- a. Doer/agent subject /kriyasampadak karta tagged as k1d
- b. Passive subject/paroksha karta as k1p
- c. Experiencer subject/Anubhab karta as k1e
- d. General subject/ sadharan karta as k1m
- e. Noun of proposition subject/vidheya/Samanadhikaran karta as k1s

(a) Doer/agent subject /kriyasampadak karta(k1d)

Doer/agent subject /kriyasampadak karta is a person or an animate object being actively engaged or participated in the action. This karta also make the work to do by someone or by something. Here, the verb is actively indicating some physical or mental exercise of the doer. The doer karta usually takes nominative or zero markers, but in some cases it may also take ‘e’, ‘Ya’, ‘te’, ‘ete’, vibhakti markers. It is tagged as k1d.

9. ছেলেটি ফুটবল খেলছে।

ChheleTi phuTabala khelachhe.

Boy-class football play-pres-prog.

The boy is playing football.

10. মিনা গান গাইছিল।

minA gAna gAichhila.

Mina song sing-past-prog.

Mina was singing a song.

11. পায়রাতে গম খেয়েছে।

pAYrAte gama kheYechhe.

Pigeon-loc wheat eatpres-perf.

Pigeon has eaten wheat.

12. মায়ে **ঝিয়ে** ঝগড়া করে।

mAYe **jhiYe** jhaga.DA kare.

Mother-loc daughter-loc quarrel do.

Mother quarrels with her daughter.

13. বাড়ির খবর জানান আমার **পিসি**।

bA.Dira khabara jAnAna AmAra **psi**.

House-gen information gives I-gen aunty.

My aunty gives the information of our house.

14. **সীতা** আয়া দ্বারা বাচ্চকে খাওয়াচ্ছে।

sItA AYA dbArA bAchchAke khAoYAchchhe .

Sita nurse by child-acc feed-pres-prog.

Sita is feeding to child by the nurse.

(b) Passive subject/paroksha karta (k1p)

In passive construction, doer of the action is called passive subject/paroksha karta. In other way, when the verb indicates an action in passive construction then its subject is defined as passive subject/paroksha karta. It is generally followed by postpositions ‘diYe’, ‘dbArA’, ‘kartRRika’, ‘hate’, ‘haite’ etc. Sometimes these postpositions are not present in the surface structure, but are implied. It is tagged as k1p. It is the logical subject of the sentence of passive

construction and must be animated. It may also take nominative and genitive marker. The verb is used here as a conjunct verb and past participle form of the verb with affix-‘A’ and the auxiliary verb ‘yA’ (‘go’) and ‘haYa’ (‘is’) etc.

15. সীমার দ্বারা এই রান্না হল ।

sImAra dbArA ei rAnnA hala .

Sima-gen by this cooking be-past.

This dish was cooked by Sima.

16. দেবদাস লেখক শরৎচন্দ্রের দ্বারা রচিত হয়।

debadAsa lekhaka **sharata chandrera** dbArA rachita haYa.

Debdas writer Saratchandra-gen by written be-pres

Debdas is written by the writer Saratchandra.

If the agent of the active voice is present, it is followed by the postposition ‘দ্বারা’, ‘দ্বিা’, ‘দিয়ে’, ‘কর্তৃক’, ‘হতে’, ‘হইতে’ and ‘কে’ (‘dbArA’, ‘diYA’, diYe’, ‘kartRRika’, ‘hate’, ‘haite’ and ‘ke’ respectively) in the passive sentence.

In some cases passive constructions take place with the auxiliary verbs ‘পড়া’ (‘pa.DA’ ‘fall’), ‘চলা’ (‘chalA’ ‘go’), ‘হয়’ (‘haYa’ ‘happen’) etc. instead of ‘yA’.

17. **ৰামবাবুৰ** দ্বাৰা সাপটা মাৰা গেছে।

rAmabAbura dbArA sApaTA mArA gechhe.

Rambabu-gen by snake-class kill go-prs-perf.

The snake was killed by Rambabu.

18. **জনতাৰ** দ্বাৰা চোৱটি মাৰা পড়ে।

janatAra dbArA choraTi mArA pa.De.

mob-gen by thief-class kill fall

The thief gets killed by the mob.

19. ইংৰেজ **বাহিনীৰ** দ্বাৰা মাৰাঠীৰা পৰাজিত হল।

i.nreja **bAhinIra** dbArA mArAthIrA parAjita hala.

English army-gen by marathis defeated be-past.

The Marathis had been defeated by the English army.

(c) **Experiencer /Anubhav karta(k1e)**

The subject of mental verbs or the verbs which express mental state, emotion, attitude, experience, feelings, perception, or event is identified as Anubhav karta, and marked as k1e. If this subject guesses something, does consider something and regards something or persons then

the subject is also identified as Anubhav karta. K1e takes genitive ('ra') marker and nominative marker. Generally, the marker is dependent on the verb with which it is associated.

The subjects of some conjunct verbs having 'haoYA', 'pAoYA', etc are examples of anubhav karta. Here verbs like 'mane haoYA', 'bodha haoYA', 'khide pAoYA', 'ghuma pAoYA', etc are used as the verb of anubhav karta. 'pAoYA' is used here only as non-volitional verb or stative verb which indicates feeling, not as a sense of receiving.

When the verb 'karA'('do') combining with noun is used as conjunct verb, it acts as mental works of the subject. So the verbs like 'chintA karA', 'mane karA', 'rAga karA'," are not mental verbs because in these verbs karta is doing the action volitionally or working mentally.

Some verbs express something which does not indicate mental works of the subject, as for example, 'bisbAsa karA', 'anubhaba karA', 'AshA karA', 'sanmAna karA', etc. Here, the nature of the verbs indicates this class of the subject.

20. **আমার** মনে হয় সে এই কাজটা পারবে।

AmAra mane haYa se ei kAjaTA pArabe.

I-gen mind-loc be-pres he this work-class can-fut.

I think he can do this work.

21. **আমার** ঘুম পেয়েছে।

AmAra ghuma peYechhe.

I-gen sleeping got-pres-perf

I am feeling sleepy.

22. **আমার** রামকে বুদ্ধিমান মনে হয়।

AmAra rAmake buddhimAna mane haYa.

I-gen Ram-acc intelligent mind –loc be-pres.

I consider Ram as intelligent.

23. **সীমা** অমলকে ভালবাসে।

simA amalake bhAlabAse.

Sima Amal-acc love-loc.

Sima loves Amal.

24. **রাম** খুব আনন্দ পেল।

rAma khuba Ananda pela.

Ram very happy got-past.

Ram was very happy.

25. **তিনি** মনে খুব কষ্ট পেয়েছেন।

tini mane khuba kaShTa peYechhena.

He mind-loc very sorrow got-pres-perf.

He was afflicted with sorrow.

(d) General subject/sadharan karta (k1m)

The karta which undergoes some action, exists somewhere or have obligation on the action is referred to as general subject/ sadharan karta. It is tagged as k1m. In these following sentences, copula or be verb is used. Sometimes this verb is dropped or does not surface in the sentence and in this case <NULL> verb is inserted as in the following examples. This verb indicates state of being and it is used as the words ‘Achhe’ , ‘haYa’, ‘haYechhe’, ‘hai’, ‘chhila’ etc. It can be thing, event, incident, etc which exists.

26. আমাদের **স্কুল** খুব ভাল। <NULL>

AmAdera **skula** khuba bhAla <NULL>.

Our school very good

Our school is very good.

27. সেখানে অনেক **দোকান** আছে।

sekhAne aneka **dokAna** Achhe.

There many shops be-pres.

There are many shops.

28. এই বাড়িটি খুব পুরানো। <NULL>

Ei **bA.DiTi** khuba purAno <NULL>.

This house-class very old.

This is heritage house.

29. খেয়াপারাপার নিষেধ ছিল।

kheYApArApAra niShedha chhila.

Boating prohibition be-past.

Boating was prohibited

(e) **Noun of proposition/ Samanadhikaran karta (k1s)**

The complements of the karta are called samanadhikaran. The subject and its complement are the same entity which indicates the complement of each other. The complements may be on the basis of meaning, person, objects or things, events, place, incident, etc.

30. নরেন ভাল ছেলে।

Narena bhAla **chhele** <haYa>.

Nren good boy <haYa>.

Naren is a good boy.

31. দেবদা ভাল গায়ক ছিলেন।

debadA bhAla **gAYaka** chhilena.

Debda good singer be-past.

Debda was a good singer.

33. এটা ঐ দিনেরই ঘটনা ছিল।

eTA oi dinera;i **ghaTanA** chhila.

This-class that day-gen incident be-past.

It was an incidence of that day only.

34. শিক্ষকের নাম পবন রায়।

shikShakera nAma **Pabana rAY**.

Teacher-gen name Paban Roy

The name of the teacher is Paban Roy

5. Discussion

While analyzing karta according to the categories mentioned above, we realized that there are a number of cases which do not fit properly or neatly in any of these categories. They

give rise to confusion or ambiguity. In the following section, we will discuss these problem cases and how they have been resolved.

A noun phrase ending with ‘ra’ marker and followed by a conjunct verb poses a problem regarding identifying the subject.

35. প্রচন্ড শক্তির উদ্ভব হয়।

prachanDa **shaktira** udbhaba haYa.

heavy energy-gen formation be-pres

Heavy energy is formed.

36. এই ভাবে তারার সৃষ্টি হয়।

ei bhAbe **tArAra** sriShTi haYa.

this way star-gen creation be-pres.

In this way the stars are created.

37. হঠাৎ . আমার চৈতন্যোদয় হল ।

haTAt **AmAra** cha;itaNyodaYa hala

suddenly i-gen realization be-past

Suddenly I became realized.

38. এ ধরনের উচ্চারণের উদ্ভব ঘটেছে পাঠের জন্য।

e dharanera **uchchAraNera** udbhaba ghaTechhe pAThera janya.

this type-gen pronunciation-gen formation happen-pres-perf lesson-gen for.

This type of pronunciation is formed for the lesson.

In these examples, we identify the noun with 'ra' marker as karta. The reason is that in all these examples the verb relates to some happening or the occurrence of some state and the person or entity undergoes this happening.

When, the noun or pronoun with 'ra'marker is followed by another noun and there is a sense of possession between these two nouns, the relation between them is that of 'sambandha'. So the noun or pronoun with 'ra'marker is not considered the subject but the head noun of this noun phrase is identified as the subject. In the following examples, the words 'AmAra', 'AmAdera', 'Narendra nAthera' and 'tomAra' are not related with the verb (action). So, the head words '**pena**', '**kaleja**', '**smritishakati**' and '**kA.Ndha**' will be treated as subject.

39. আমার পেন আছে।

AmAra **pena** Achhe.

i-gen pen be-pres .

I have a pen.

40. আমাদের **কলেজ** খুব ভাল।

AmAdera **kaleja** khuba bhAla.

Our college very good.

Our college is very good.

41. নরেন্দ্রনাথের **স্মৃতিশক্তি** ছিল বিস্ময়কর।

NarendranAthera **smritishakti** chhila bismaYakara.

Narendranath –gen memory be- past wonderful.

The memory of Narendranath was wonderful.

42. তোমার **কাঁধ** চওড়া ছিল।

tomAra **kA.Ndha** chao.DA chhila.

You-gen shoulder broad be-past.

Your shoulder was broad.

The semantics of the verb ‘pAoYA’ as main verb and conjunct verb has two folds-

- i. animate subject receiving something concrete
- ii. inanimate subject attaining some state

43. **সে** রাজকোষ থেকে ভরণপোষণ পায়।

se rAjkoSha theke bharaNposhaNa pAYa.

he treasury from help get-pres.

He gets help from treasury.

44. **রোগ** হ্রাসবৃদ্ধি পায়।

roga hrAsabRRiddhi pAYa.

disease increased and decreased get.-pres

Disease is increased and decreased.

45. **দাম** বৃদ্ধি পায়।

dAma bRRiddhi pAYa.

cost increase get-pres .

The cost is increased.

Here the words **se**, **roga** and **dAma** are all sadharan karta(k1m).

When the verb ‘pAoYA’ is used as conjunct verb and the main verb relates to the act of experience, feelings, emotions etc., in this context the subject should be an animate subject and the subject takes ‘ra’ and also zero or nominative marker.

46. **তোমার** খিদে পেয়েছে।

tomAra khide peYechhe.

You-gen hunger get-pres, prog.

You are hungry.

47. সে কষ্ট পেয়েছে।

se kaShTa peYechhe.

he hurt get –pres,perf

He feels hurt.

Here the words **tomAra** and **se** are experiencer karta(k1e).

When an event is caused by an animate subject which comes in the position of grammatical subject of the sentence, in such cases the subject is the logical subject and it takes “ra” marker. In the following sentences, the word ‘dbArA’ is not in the surface structure, but it is implied [1].

48. তোমার খাওয়া হয়েছে?

tomAra khAoYA haYechhe?

You-gen eating be-pres,perf.

Have you eaten?

49. তার ভাত খাওয়া হয়েছে।

tAra bhAta khAoYA haYechhe.

he-gen rice eating be-pres,perf.

He has eaten rice.

Here the words **tomAra** and **tAra** are passive karta (k1p).

In the following sentence, ‘kichhu jinisa’ is considered as object (karma) and not as subject (karta). Here the inherent meaning of the verb ‘chokhe pa.Dlo’ indicates an animate subject. Though the subject does not surface but its occurrence is indicated by the semantics of the verb.

50. রথের মেলায় কিছু জিনিস চোখে পড়ল।

rathera melAYa kichhu jinisa chokhe pa.Dala.

rath-gen fair-loc few thing eye-loc fall-past.

In the fair of Rathajatra a few things came to our sight.

In the following sentence, the word ‘**meYedera**’ is identified as subject because of the form of the verb ‘dekhA’ that is the participle form of this verb. Therefore, the grammatical subject of the verb ‘**meYedera**’ is actually the logical subject here.

51. মেয়েদের দেখা নেই কেন?

meYedera dekhA nei kena?

Girls-gen see-ppt be-pres-neg why

Why the girls are not present?

6. Conclusion

It can be concluded that use of semantics is required to identify and classify the karta. It is necessary to identify the logical subject and distinguish from the grammatical subject. Semantics is also used to understand the nature of the verb whether it comes as a main verb or as conjunct verb. Through semantics the nature of the verb indicates the role of the karta. Semantics also indicates whether the karta is animate or inanimate. We have used syntactico-semantic version of the grammar for identifying the karta. Both syntax and semantics are used to identify the roles of karta in a Bangla sentence. This notion of grammar may be used in the other dependency relation. This syntactico-semantic relation may be used in a Treebank which may be further used in a statistical parser. The categorization of verb which indicates the semantics or nature of verb is needed. With the help of this idea semantic annotation can be developed. The creation of such resources is expected to improve Bangla NLP system greatly.

References

- [1] Chatterji, Suniti Kumar. “BHASHA-PRAKASH BANGLA VYAKARAN” [A Grammar of the Bangla Language] Rupa, New Delhi, pp 126, 236-239, 363, 2003.
- [2] Bharati, A., Chaitanya, V., Sangal, R. “Natural language processesing: A paninian perspective” (1999).
- [3] Gangopadhyay, Malaya. “The noun phrase in Bengali: assignment of role and the kāraka theory”, Publisher: Motilal Banarsidass, Foreign Language Study, 1990.
- [4] Chakraborty, Bamandeb, “UCHCHATARA BANGLA BYAKARAN” (Higher Bangla Grammar) Akshaya Malancha, Kolkata, 2007.

- [5] Sharma, D.M., Sangal, R., Bai,L., Begam, R., “AnnCorra : TreeBanks for Indian Languages”, Language Technologies Research Center, IIIT, Hyderabad, India
- [6] de Marneffe, M., Manning, C.D. “Stanford typed dependencies manual”, (2008)
- [7] Chatterji, S., Sarkar, T.M., Sarkar, S. and Chakraborty, J. “Karak Relations in Bengali” Proceedings of 31st All-India Conference of Linguists (AICL 2009), Hyderabad, India, pp 33-36, December, 2009.
- [8] Marcus, Mitchell P., Marcinkiewicz, Mary Ann and Santorini, Beatrice. “Building a Large Annotated Corpus of English: The Penn Treebank”,Comput. Linguist., MIT Press, June 1993,19(2), pages: 313-330, Cambridge, MA, USA
- [9] J. Hajic. “Building a Syntactically Annotated Corpus: The Prague Dependency Treebank”, In Issues of Valency and Meaning, pp. 106-132, Karolinum, Praha 1998
- [10] Chatterji, Sanjay, Mukherjee Sarkar, Tanaya, Dhang, Pragati, Deb, Samhita, Sarkar, Sudeshna, Chakraborty, Jayshree, Basu, Anupam, “A dependency annotation scheme for Bangla Treebank”, Language Resources and Evaluation, Volume 48 Issue 3, pp 443-477, September 2014