RULE BASED APPROCH OF CLAUSE BOUNDARY IDENTIFICATION IN TELUGU

Ganthoti Nagaraju
Department of Linguistics and Language Technology, Central University of Kerala
gnagarajug62@gmail.com

Thennarasu S
Department of Linguistics and Language Technology, Central University of Kerala
thennarasus@gmail.com

Christopher Mala
Center for Applied Linguistics and Translation Studies, University of Hyderabad
efthachris@gmail.com

Abstract

One of the major challenges in Natural Language Processing is identifying Clauses and their Boundaries in Compu-tational Linguistics. This paper attempts to develop an Automatic Clause Bound-ary Identifier (CBI) for Telugu lan-guage. The language Telugu belongs to South-Central Dravidian language fami-ly with features of head-final, left-branching and morphologically agglutinative in nature (*Bh. Krishnamurti*, 2003). A huge amount of corpus is studied to frame the rules for identifying clause boundaries and these rules are trained to a computational algorithm and also discussed some of the issues in identifying clause boundaries. A clause boundary annotated corpus can be developed from raw text which can be used to train a machine learning algorithm which in turns helps in development of a Hybrid Clause Boundary Identification Tool for Telugu. Its implementation and evaluation are discussed in this paper.

1. Introduction

A Clause is a grammatical unit that includes, at minimum, a predicate and an explicit and implied subject and expresses a proposition (*Crystel*, 1980). In other words, a clause is defined as a group of words having a subject and a predicate. It is a well-known fact that a sentence may contain one or more clause. Simple sentences always have a single clause. Analyzing these clauses in NLP is an easy task. But when a sentence has more than one clause it becomes difficult to process. Identification of predicate and its dependent thematic elements become even more difficult. To solve this problem, identification of clause boundary is mandatory. Clause Boundary Identification is the process of dividing the given sentence into a set of clause. Correct automatic detection of major syntactic boundaries, in particular clause boundaries help in improving many other tools in NLP (*Leffa*, 1998; *Ejerhed*, 1988, *Vijay et al*., 2009; *Gadde et al*., 2010). The ***Telugu Clause boundary identifier (T-CBI)*** is an automatic tool to identify boundaries of clauses and mark their start and end points. In another words, it identifies the structure that underlies the sentence. Shallow parsed sentence are used as input to T-CBI and further parse the sentence and marks the clause with their boundaries along with their appropriate tags. This module can be used in bigger NLP systems like Machine Translation systems, Information Extraction and Information Retrieval, Search Engines, etc.

The data driven (*Puscasu*, 2004) and rule based (*Leffa*, 1998) approaches are prominent in the development of a CBI. In order to build a clause boundary identifier, using data driven approach, one needs to have a good clause boundary annotated corpus for training (*Sharma et al*, 2013). Such a corpus is not available in the Telugu language. Hence a Rule-Based approach is selected in the current study to develop an efficient CBI for Telugu. By using Morphological cues such as agreement markers (person, gender, number) and case markers/ Tense Aspect Modal

S Bandyopadhyay, D S Sharma and R Sangal. Proc. of the 14th Intl. Conference on Natural Language Processing, pages 408–416, Kolkata, India. December 2017.

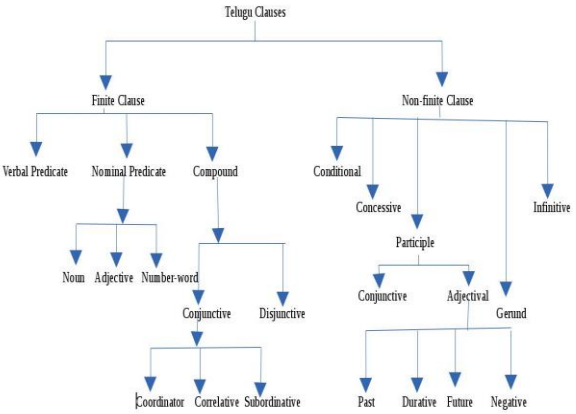
(TAM) markers to identify the start and end of the clause. Other than these, certain lexical cues are used to identify the CBI. Identified Thematic roles of the constituents are used in the T-CBI for better performance of the Rules. The development of automatic T-CBI will be used to develop the clause boundary annotated corpus for the task of clause boundary identification from raw text using machine learning process in NLP.

2. Review of Earlier Researches

Clause Boundary Identification started with *Eva Ejerhed's* Clause Identification System (*Ejerhed*, 1988) for text to speech system. *Leffa* (1998) has developed a rule-based system to identify clauses from English to Portuguese machine translation system. *Papageorgiou* developed a rule-based clause boundary system as a pre-processing tool for bilingual alignment parallel text (*Papageorgiou*, 1997). *Tomohiro Ohno et al.* (2006) built CBI for Japanese to implement Spoken Monologue System. The dependencies within a clause are identified by dividing a sentence into clauses and executing stochastic dependency parsing for each clause. Later, the dependencies over clause boundaries are identified stochastically, and the dependency structure of the entire sentence is thus completed. This method executes dependency parsing in two stages: at the clause level and at the sentence level. According to their evaluation, the recall of the system is 95.7% and the precision is 96.9%. *Phani Gadde et al.* (2010) have attempted to improve data driven dependency parsing using clausal information. They have used Stagel parser of *Husain et al*. (2009), to provide the clause boundary information that is then incorporated as features during the actual parsing process. They experimented with different combinations of the information provided in the data such as Vibhakti and TAM fields. *Daraksha Parveen et al.* (2011) have built a CBI for Urdu using Conditional Random Field (CRF) as the classification method and clause markers. A hybrid approach is proposed to use both techniques i.e. rule based and machine learning to build an identifier for different clause boundaries of Urdu language. *Lakshmi, S. et al.* (2012) have built a clause boundary identification system for *Malayalam* sentences using the CRF. Here, the clause boundaries are identified using grammatical features. *Sobha L. et al.* (2013) have built Malayalam CBI using CRF. They have developed a corpus with tagging of different type of clauses as well as the start and end of the clause. They selected approximately 6415 tourism and 385 health corpus sentences from the Web and the training set consisted of 5000 sentences from both the domains. Testing of the system was done with 401 unseen sentences from the corpus of tourism. They achieved a precision and recall on different types of clauses of about 70% and 80% respectively. *Aniruddha Ghosh et al.* (2013) have built CBI for the Bengali language. They used a syntactic rule based model with CRF, a machine learning technique. They have achieved 73% and 78% of precision and recall respectively. *Rahul Sharma et al.* (2013) have attempted to build a clause boundary to Hindi Treebank data. They have used the dependency attachments and dependency annotated relations to mark clauses. They chose 16,000 sentences and conducted an exercise on 238 clauses and got the result of 94% of accuracy in the clause boundary identification.

3. Description of Rule Format.

Classification of Telugu Clauses



3.1. Finite Clauses

Finite Clause contains subject and finite verb, the finite verb is in agreement with the subject. Finite clauses are classified as Verbal Predicate, Nominal Predicate and Compound. Nominal predicate in sub-categorized to Noun, Adjectival and Number-words and Compound Clauses are sub-categorized to Conjunctive and Disjunctive Clauses. Conjunctive clauses are further sub-categorized into co-ordinate, correlate and subordinate clauses.

3.1.1. Verbal Predicate clause

In verbal predicate clause, a finite verb occurs in the end of a sentence. It distinguishes for tense such as past, present and future. Each tense distinguishes positive and negative clauses. Examples for verbal predicate clauses are given below:

Past Tense

i) Positive

nēnu āme-nu tiṭṭ ā-nu
I she ACC scold PST 1.SG.
'I scolded her'

ii). Negative

nēnu āme-nu tiṭṭ a-lēdu
I she ACC scold INF- PST. NEG
'I did not scold her'

3.1.2. Nominal Predicate clause

Noun/adjective/number word phrase in its predicate position agrees in gender, person and number with the subject. In the case of first or second person singular or in first person plural in Telugu, they inflect with number and with person. The pronominal suffixes -ni/-nu, -mi/-mu and -vi/-vu are manifested on the predicate for first person singular, first person plural and second person singular respectively.

i). Noun phrase in Predicate position:

nēnu oka abbāyi- ni
I a boy 1.SG.
'I am a boy'

This is an example for nominal predicate sentence. The subject *nēnu* agrees with the nominal predicate. The agreement is manifested with the marker *-ni*

ii). Adjective phrase in Predicate position:

nīvu/nuvvu poḍugu vāḍi-vi
you tall poss. 2.SG.
'You are a tall boy/guy'

The subject *nīvu/nuvvu* agrees with the adjectival predicate. The agreement is manifested with the marker *-vi*.

iii). Number word phrase in Predicate position:

mēmu muggu- ra-mu
we three 3.SG.H. 1.PL.
'We are three persons'

This is an example for number word predicate sentence. The subject *mēmu* agrees with the number-word predicate. The agreement is manifested with the marker *-mu.* No Nominal Predicate Agreement for 2.PL and 3.SG/PL.M/F/N

3.1.3 Compound Clause

A sentence containing two or more coordinate independent clauses, usually joined by one or more conjunction markers, but no dependent clause, as in the below sentence.

e.g.: [The lightning flashed]$_{MC}$ and [the rain fell.]$_{MC}$

3.1.3a Conjunctive Clause

Some compound sentences are joined by a conjunction. Some of the conjunction marker in Telugu are: ani, mariyu, leka, (k*ā*ni), k*ā*ka, Ena, kanuka, endukante.

Examples of compounds in sentences include:
m*ā* *ā*yana panilo nimaGFulE unn*ā*ru aMdukani nenu baj*ā*ruku velYl*ā*nu.

Other compound sentences are joined with a semicolon. If a semicolon is used, it may or may not have a conjunctive adverb.

i). Coordinative Clause

This is done using one of two or more clauses of equal status in a sentence, especially when joined by a coordinating conjunction. Here *kāni* is the coordinative marker.

mēmu vacc-ā- -mu kāni [vāḍu rā-lēdu
we come PST. 1.PL. but he come.INF PST.NEG
'We came but he did not come.'

ii) Correlative Clause

A correlative conjunction is a paired conjunction that links balanced words, phrases, and clauses. The elements connected by correlative conjunctions are usually parallel - that is, similar in length and grammatical form. Each element is called a *conjoin*. Some of the Correlatives in Telugu are: *leka/kāka/gāka, kādu/kani and "-e kāka/-e gāka/-V kūda.*"

evaḍu vaccāḍ-ō vāḍu nā tammu-ḍu
who come RTM he ACC brother 3.SG.
'The one who came is my brother'

Here we are using morphological cue
X-*ō Y*
ō-relative-correlative marker, X and Y are clauses.

iii) Subordinate Clause

A subordinating conjunction is a word that connects a main clause to a subordinate clause. A main clause is an independent clause that can stand alone by itself as a sentence. In other words, a main clause does not need any additional information to operate as a sentence.
Here we use lexical cues:
X- COMP *Y*
COMP -complementizer, X and Y are clauses.

ani	*annā*
anēdi	*anna*
atlu	*aMte*
annatlu	

MC1[nēnu vacc- ā- -nu]MC1 ani MC2[vāḍi-to cepp- ā- nu]MC2
I come PST 1SG. COMP he ASS. say PST. 1.SG
'I told him that I came'

3.1.3b. Disjunctive Clause

A coordinate construction is the one that uses a *disjunctive conjunction* to indicate a contrast. The items on either side of the disjunctive conjunction are called disjunct.
Here we use morphological cues.
X- *ō Y-ō Z*
ō -Disjunctive marker, X, Y and Z are clauses

vast- ā- ḍ- ō rā- ḍ- ō nāku tēliyadu
come FUT. 3.SG.M DISJ come. FUT. NEG-3.SG.M DISJ I-DAT do not know
`I do not know either he comes or not'

3.2. Non-finite Clauses

Non-finite clause are formed with non-finite verb and verb does not marked with gender, number and person suffixes in agreement with grammatical subject of the sentence, but they form by adding appropriate tense-mode suffix to a verb stem. And they are always depended and embedded.

3.2.1. Conditional clause

In this conditional clause we use morphological cue -te `positive conditional marker'
-rākapōte/ rākuMṭe `negative conditional marker'
SC[nuvvu addamu nu jāraviḍis-tē]SC MC[adi pagilipō- tuM- di]MC
you mirror ACC. drop COND that break FUT 3.SG.N.
'If you drop the mirror, it will break'

This is an example for complex sentence with conditional clause. As we noted here, there are two clauses, one of them is subordinate clause or dependent clause and the other one is super ordinate clause or main clause. The verb is present in both the sentences. The subject *nuvvu* in SC ends with conditional marker -*t*ē and *adi* in MC agrees the agreement and is manifested with the marker -*di* respectively

3.2.2. Concessive clause

In this concessive clause we use morphological cues. *-inā* `positive concessive marker' *-rākapōyinā/ -rākunnā* `negative concessive marker

SC[jōhn veḷl- inā]SC MC[mary veḷla-du]MC
John go CONCM Mary go 3.SG.
'Though John goes, Mary don't go'

This is an example for complex sentence with concessive clause. As we noted here, there are two clauses, one of them is subordinate clause or dependent clause and the other one is super ordinate clause or main clause. The verb is present in both the sentences. The subject *Jōhn* in SC ends with concessive marker *-inā* and *Mary* in MC agrees the agreement and is manifested with the marker *-du* respectively.

3.2.3. Participle Clause

3.2.3a. Conjunctive Participle

We use morphological cues in this conjunctive participle clause. *-i* `positive conjunctive participle marker' *-aka/ akuMḍā* `negative conjunctive participle marker'

MC[SC[ataḍu iḍli tin- i]SC kāṅī coffee trāga lēdu]MC
he idli eat CP coffee drink NEG
'Having eaten idli, he didn't take coffee'

This is an example for complex sentence with verbal participle clause and also forward control. As we noted here, there are to clauses, one of them is subordinate clause or dependent clause and the other one is super ordinate clause or main clause. The verb is present in both the sentences. The subject *ataḍu* in SC ends with verbal participle marker-*i* and in MC agrees the agreement and is manifested with the negative marker *lēdu* respectively.

sujāta snānaM ceyy- aka vāraM rōjulay- iM-di
sujatha bath PV NCP week days PST 3.SG.
'Sujatha has not bathed from the last one week'

According to *Chekuri ramarao,* in his book *Telugu Vakyam,* he explained in this sentence (48) also that it is a temporal expression, even though *'vāraM rōjulayiMdi'* is plural marker it takes only singular agreement marker *-di.* In this case we should not use yesterday, day before yesterday, etc. for this sentence also. Here we have negative participle marker *-aka* is there in the sentence.

3.2.3b. Adjective Participle

We use morphological cues in this adjectival predicate. *-ina* `positive past adjectival participle marker' *-tunna* `positive durative adjectival participle marker' *-ē* `positive future adjectival participle marker' *-ani* `Negative adjectival participle marker'

Adjectival participle form of verbs can be pronominalized.

i) Past Adjectival Participle:

SC[nēnu vāḍi- ki icc- ina pustakaM]SC
I he-DAT give PP book
'I gave a book to him'

This is an example for accusative nominalization in complex sentences with adjectival participle clause. As we noted here, the subject *nēnu* in SC ends with past participle marker-*ina* and the accusative *pustakaM* is nominalized as shown in the example. The actual sentence is *'nēnu vāḍiki pustakaM iccānu',* before doing nominalization *pustakaM* is in the accusative position.

nēnu vāḍi- ki ivv- ani pustakaM
I he DAT give NPP book
'I didn't give a book to him'

This is an example for accusative nominalization in complex sentences with adjectival participle clause. As we noted here, the subject *nēnu* in SC ends with negative past participle marker *-ani* and the accusative *pustakaM* is nominalized as shown in the example. The actual sentence is *' nēnu vāḍiki pustakaM ivva lēdu',* before doing nominalization *pustakaM* is in the accusative position

ii) Durative Adjectival Participle

vāḍi- ki pustakaM is- tunna nēnu
he DAT book give DP i
'I am giving a book to him'

This is an example for nominative nominalization in complex sentences with adjectival participle clause. As we noted here, the subject *vāḍu* in SC ends with durative participle marker *-tunna* and the nominative *nēnu* is nominalized as shown in the example. The actual sentence is *'nēnu vāḍiki pustakaM istunnānu',* before doing nominalization *nēnu* is in the nominative position.

vāḍi- ki pustakaM ivv- ani nēnu
*h*e DAT book give NDP i
'I am not giving a book to him'

This is an example for nominative nominalization in complex sentences with adjectival participle clause. As we noted here, the subject *vāḍu* in SC ends with negative durative participle marker *-ani* and the nominative *nēnu* is nominalized as shown in the example. The actual sentence is *'nēnu vāḍiki pustakaM ivvaḍaM lēdu',* before doing nominalization *nēnu* is in the nominative position

iii) Future Adjectival Participle:

vāḍi- ki pustakaM icc- ē nēnu
he DAT book give FP I
'I will give a book to him'

This is an example for nominative nominalization in complex sentences with adjectival participle clause. As we noted here, the subject *vāḍu* in SC ends with future participle marker *-ē* and the nominative *nēnu* is nominalized. The actual sentence is *'nēnu vAdiki pustakaM iswānu',* before doing nominalization *nēnu* is in the nominative position.

vāḍi- ki pustakaM ivv-ani nēnu
he DAT book give NFP i
'I will give a book to him'

This is an example for nominative nominalization in complex sentences with adjectival participle clause. As we noted here, the subject *vāḍu* in SC ends with negative future participle marker *-ani* and the nominative *nēnu* is nominalized as shown in the example. The actual sentence is *'nēnu vAdiki pustakaM ivvanu',* before doing nominalization *nēnu* is in the nominative position.

3.2.4. Gerundival Clause

In gerundival clause also we use morpholgical cues: *-aḍaM* `gerundival marker'

SC[vāḍi- ki cepp- aḍaM]SC MC[nā- ku isṭaM lēdu]MC
he DAT book GEND me DAT like NEG
'I don't like telling him'.

This is an example for gerundival clause. As we noted here, the subject *vāḍu* in SC ends with gerundival marker *-aḍaM.*

3.2.5. Infinitival Clause

In infinitival clause we use morphological cues. *-a* `infinitive marker', *-a_ gānē* 'as soon as', *-a_baṭṭi* 'because'

āme cepp- a- baṭṭi nēnu cēs- ā- nu
she say INF because I do PST 1.SG
'Because she told, I did'.

This is an example for infinitival clause. As we noted here, the subject *āmē* in SC ends with infinitival marker *-baṭṭi.*

Rule for Clause Identification

Rules for clause boundary identification are described in the below:

Clause type	Clause start	Clause end	Information		
			Chunk	Morph-tam	2nd column
Conditional clause		'-te'(past)	VGNF	'-te'(+ve) '-akapōte'(-ve) '-akuMte'(-ve)	
		'-tunMte'(durative)	VGNF	'-tunMte'(+ve) '-akuMte'(-ve)	
Concessive clause		'-inā' (past)	VGNF	'-inā'(+ve) '-kapōyinā'(-ve) '-akunnā-(-ve)	
		'-tunnā'(durative)	VGNF	'-tunnā'(+ve) '-akunnā-(-ve)	
Verbal participle		'-i'(past)	VGNF	'-i'(+ve) '-aka'(-ve)	
		'-tū'(durative)	VGNF	'-tū'(+ve) '-akuM?ā(-ve)	
Adjectival participle					
1.past		'-ina'	VGNF	'-ina'(+ve) '-ani'(-ve)	
2.durative		'-(t)unna'	VGNF	'(t)unna'(+ve) '-ani'(-ve)	
3.future		'-ē'	VGNF	'-ē'(+ve) '-ani'(-ve)	
Gerundival cluse					
		'-a?aM_ki' '-a?aM_valana/valla'	VGNF VGNF	'-a?aM_ki' '-a?aM_valana/valla'	
		'-a?aM_ce'	VGNF	'-a?aM_ce'	
		'-a?aM_to'	VGNF	'-a?aM_to'	
		'-a?aM_e'	VGNF	'-a?aM_e'	
		'-a?aM_gani'	VGNF	'-a?aM_gani'	
		'-a?aM_0_e' '-a?aM_to_e'	VGNF VGNF	'-a?aM_0_e' '-a?aM_to_e'	
		'-a?aM_nu'	VGNF	'-a?aM_nu'	
		'a?aM_gala_a'	VGNF	'a?aM_gala_a'	
		'a?aM_kūdā'	VGNF	'a?aM_kūdā'	
		'-a?aM_ki'	VGNF	'-a?aM_ki'	
		'-a?aM_kosaM_ainā'	VGNF	'-a?aM_kosaM_ainā'	
Tense		'-ā'(Past)	VGNF	'-ā'(+ve) 'a_lēdu'(-ve)	
		'-tunn'(durative)	VGNF	'-tunn'(+ve) '-a?aM_lēdu'(-ve)	
		'-tā'(Future)	VGNF	'-tā'(+ve) '-a'(-ve)	
Mood / Mode		'-0'(Imperative)	VGNF	'-0'(+ve)	
		'-a_vaddu'(Prohibitivee)	VGNF	'-a_vaddu'(+ve)	
		'-dā'(Hortative)	VGNF	'-āmu'(+ve)	
		'-āli'(Obligative)	VGNF	'-āli'(+ve)	
		'-gala'(Capabilitative)	VGNF	'-gala'(+ve) a_lenu	
Non-finitive		'-an'	VGNF	'-an'	
Auxiliaries					
Negative past		'a_le'	VGNF	'a_le'	
Passive		'-a_ba?u'	VGNF	'-a_ba?u'	
Inceptive		'-a_bo'	VGNF	'-a_bo'	
Inchoative		'-a_nāBiMcu'	VGNF	'-a_nāBiMcu'	
Optative		'-a_gōru'	VGNF	'-a_gōru'	
Permissive		'-a_ivvu'	VGNF	'-a_ivvu'	
Causative		'-a_manu' '-0_komanu'	VGNF	'-a_manu' '-0_komanu'	
Abilitative		'-a_tagu'	VGNF	'-a_tagu'	
Probability		'-a_vaccu'	VGNF	'-a_vaccu'	
Forcive		'-a_valayu' '-a_budDavvu '	VGNF	'-a_valayu' '-a_budDavvu '	
Negative permissive		'-a_kū?adu'	VGNF	'-a_kū?adu'	
Reflective		'-ko'	VGNF	'-ko'	
Reciprocal		'-ko'	VGNF	'-ko'	
Complitive		'-i_pō'	VGNF	'-i_pō'	
Discourse markers:					
And			CCP		'mariyu'
But			CCP/BLK		'kāni'
Or			CCP		'lēka'
Because			CCP		'eMdukanaga/eMdukaMte
So			CCP		'aMduvalla' aMduvalana
If					'okavē?a'
Then			VGNF/CCP		'ayitē'
Even though					'ayinappatiki'
Yet			NP		'iMkā'
Because of that					'dānivalana'
Perhaps			VM		'bahuś ā'
Not only this					'idikā'kuM?ā'
Also			BLK		'kū?a'
Lastly					'civaragā'
Next			NP		'taruvāta'
Meanwhile					'madyalo'

4. Testing and Evaluation

We have taken 5000 sentence from tourism and health domain corpus to test the rule set. This corpus includes all type of sentence like Relative Participle, Conditional, Main clauses and etc.

S. No.	Clause Type	No. of Sent
1	Conditional	907
2	Concessive	280
3	Verbal participle	1017
4	Adjectival participle	674
5	Gerundival	436
6	Non-finite	869
7	Relative	817
	Total	5000

These sentences were first processed with shallow parser from which we extract Morphological, POS and Chunk information later this is given as input to simple parser which assigns the karaka roles (thematic relationships) which indeed helps the T-CBI perform well. Evaluation of the Rules is given below.

Clause Type	No. of Sentence	Correct Identifcation	% of recognition
Conditional	907	857	
Concessive	280	184	
Verbal participle	1017	802	
Adjectival participle	674	490	
Gerundival	436	421	
Non-finite	869	532	
Relative	817	593	
Total	5000	3879	77.58

Out of 5000 sentence T-CBI was able to identify 3879 sentence correctly which means the performance of the T-CBI can be scaled to 77.58% of accuracy. Using this we can create huge amount of corpus of CBI pre-annotated.

5. Conclusion

Thus, this paper has attempted to show how implicit clausal information captured in a shallow parsed text can be extracted and are used to develop CBI for Telugu. The paper has also discussed some of the issues in identifying clause boundaries using the above said approach. A plan has been implemented with these rules in a computational algorithm for future exercises. Once the rules are implemented it would be easy to scale the performance of the rules designed. A clause boundary annotated corpus can be developed from raw text which can be used to train a machine learning algorithm which in turns helps in development of a Hybrid Clause Boundary Identification Tool for Telugu.

Reference

Bharati, S. Husain, B. Ambati, S. Jain, D. Sharma and R. Sangal. 2008. `Two Semantic features make all the difference in Parsing accuracy'. Proc. of ICON-08.

Bharati, Vineet Chaitanya, Rajeev Sangal. *Natural Language Processing A Paninian Perspective.* Prentice Hall of India (1995).

Ghosh, A. Das, and S. Bandyopadhyay, “Clause Identification and Classification in Bengali,” in *Proceedings of the 1st Workshop on South and Southeast Asian Natural language Processing* (WSSANLP, 23rd International Conference on Computational Linguistics (COLING), Beijing, August 2010, pp. 17-25.

E. Ejerhed, “Finding Clauses in Unrestricted Text by Finitary and Stochastic Methods,” in *Proceedings of the 2nd Conference on Applied Natural Language Processing*, Austin Texas, 1988, pp. 219-227.

Gadde, Phani, et al. 2010. "Improving data driven dependency parsing using clausal information." *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*. Association for Computational Linguistics.

Lakshmi S, Vijay Sundar Ram R and Sobha Lalitha Dev. 2012. Clause

Boundary Identification for Malayalam Using CRF. In *Proceedings of the Workshop on Machine Translation and Parsing in Indian Languages* (MTPIL).24th International Conference on Computational Linguistic

Sharma, Rahul, et al. 2013. "*Automatic Clause Boundary Annotation in the Hindi Treebank.*" WSSANLP-2013: 83.

"*A rule based approach for automatic Claus boundary detection and classification in Hindi"* by Rahul Sharma and Soma Paul [Proceedings of the Conference the 5th Workshop on South and Southeast Asian NLP WSSANLP - 2014].