# DEVELOPING LINKS OF COMPOUND SENTENCES FOR PARSING THROUGH MARATHI LINK GRAMMAR PARSER

Vaishali B. Patil<sup>1</sup> and B. V. Pawar<sup>2</sup>

<sup>1</sup>Institute of Management Research and Development,Shirpur, Maharashtra 425405, India <sup>2</sup>School of Computer Sciences, North Maharashtra University,Jalgaon, Maharashtra 425001, India

#### ABSTRACT

Marathi is a verb-final language with a relatively free word order. Complex Sentences is one of the major types of sentences which are used commonly in any language. This paper explores the study of complex sentence structure of Marathi language. The paper proposes various links of complex sentence clauses and modelling of the complex sentences using proposed links in the Link Grammar Framework for parsing purpose.

#### **KEYWORDS**

Marathi Complex Sentences, Link Grammar, Marathi Link Grammar Parser

## **1. INTRODUCTION**

Link Grammar is a formal grammatical system defined on the basis of natural language property which states that if arcs are drawn connecting each pair of words that relate to each other, then the arcs will not cross [16]. This property is called as planarity. A parsing system has been developed to capture many phenomenon of English grammar by providing roughly seven hundred definitions that includes the word of the language and their linking requirements and an algorithm [6] for parsing sentences according to the given grammar.

A given sentence is accepted by a system if the linking requirements of all the words in a sentence are satisfied (connectivity property), none of the links between the words cross each other (planarity property) and there exists at most one link between any pair of words (exclusion property).

Parsed output is very fundamental requirement for natural language processing (NLP) applications like Information retrieval, Information extraction, Question Answering, etc. especially in Machine translation [17]. Indian languages are resource deficient languages as it does have very limited electronically managed tools like morphological analyzer, part of Speech tagger, parser etc. Marathi language is also not an exception however since last decade there are numerous efforts has been witnessed among this we have gone through [3, 4, 5, 12, 13, 14, 15]. Our proposed Marathi link Grammar parser is one attempt to develop such tools which will be helpful in various applications wherever it suits better. Following figure will give quick glimpse of our proposed system.

DOI: 10.5121/ijnlc.2014.3601



Figure 1 Block Diagram of Proposed Marathi Link Grammar Parser

Our proposed Marathi link grammar parser is rule based parsing system which contains link database, the handcrafted rules and an algorithm to get parsed output if one exists. So far by studying Marathi noun phrases, verb phrases and subject/object to verb agreement we have proposed 31 links [8, 9, 10]. Based on computational Panini grammar [1] we proposed Karaka links [11] which defines the relation between nominal words with verb of a sentence summarized in table1. Karka relations are the relations of nominal that participate in the action specified by the particular verb mentioned in the sentence. Links between any pair of words gives the functional association between that pair of words. For eg consider the sentence "Ram aamba khato (राम आंबा खातो : Ram eats mango)" by our proposed system links between words will be established between verb→karta and verb→Karma as sentence consists it. Hence Ka\_karta link will be established on khato (खातो: eats) and Raam (राम : proper Noun) and Ka\_karma link will be established on khato (खातो: eats) and aamba (आंबा: Mango) word pairs.

Karaka	Link	Functionality	
Karta	Ka_Karta	Verb to Subject	
Karma	Ka_Karma	Verb to Object	
Karan	Ka_karna	Verb to Instrument of the Activity	
Adhikarna	Ka_Adhikarna	Verb to time and place of the activity	
Aapadan	Ka_Aapadan	Verb to word which gives separation meaning	
Sampradan	Ka_Sampradan	Verb to word which gives donation meaning	

Table 1:	Karaka	and	its	links
----------	--------	-----	-----	-------

The task of our system is building links by judging each individual word's role in the sentence. A system gets complete linkage if it satisfies all the rules laid as per link grammar framework i.e. Planarity, Connectivity and Exclusion.

# 2. COMPOUND SENTENCES IN MARATHI

In Marathi language, coordination is of two type sentence coordination and constituent coordination [2][7]. There are three major coordinators namely Conjunctives, Disjunctive and Adversative.

# 2.1. Sentence Coordination

Any number of sentences can be coordinated with "aani" (आणि : and) which is always placed before the last conjunct. In a sequence of more than two sentences, all preceding sentences before the last are simply juxtaposed as given in following example:

Ex 1: babu aala aani lili ghari geli (बाबू आला आणि लिली घरी गेली : Babu left and Lili came home) Ex 2:babu aala, lili ghari geli aani lagech minila phone kela.

(बाबू आला, लिली घरी गेली आणि लगेच मिनीला फोन केला : babu left, Lili came home and immediately phoned Mini)

Sentence coordination is used to express various semantic distinctions such as contrast, contingence, sequential events and even casual connections.

# 2.2 Constituent/word level Coordination

Various parts of speech can be coordinated at constituent level. Nouns of all categories may be coordinated. Pronouns, adjectives, adverbs and active and passive verbs can also be coordinated. While coordinating within a sentence part of speech follows certain agreement rules on the conjoining category. Following are few examples on constituent level coordination,

Ex 3: Noun (Subject) Coordination lili sudha aani mini gharat hotya. (लिली सधा आणि मिनी घरात होत्या : Lili, Sudha and Mini were in the house) Ex 4: Noun (object) Coordination liliNe aambe keli aani peru khalle (लीलीने आंबे केळी आणि पेरू खाल्ले : Lili ate mangoes, bananas and guavas ) Ex 5: pronoun coordination mi aani tu udya baget jau (मी आणि तू उद्या बागेत जाऊ : I and you will go in the garden tomorrow) Ex 6: Adjective Coordination lili jara bavali aani vedi aahe (लिली जरा बावळी आणि वेडी आहे : Lili is a little bit disorderly and crazy) Ex 7: Adverb Coordination lili halu halu aani mand swarat bolate (लिली हळू हळू आणि मंद स्वरात बोलते : Lili speaks slowly and in a low voice) Ex 8: verb coordination chor kholit shirala aani lagech pakadala gela (चोर खोलीत शिरला आणि लगेच पकडला गेला : Thief entered the room and was immediately caught)

## 2.3 Conjunctive Coordination

The basic conjunctive coordinator is "aani" (आणि : and) with alternates such as wa (व : and, ankhi (आणखी : and), aankhin (आणखीन : and), aanik (आणिक : and), an (अन : and). The first alternate i.e. wa (व) is a perso-Arabic borrowing. It is used mostly in literary styles however; its use is increasing in Modern Marathi. The rest are used in conversational speech. All examples mentioned in section 2.1 and 2.2 are confined to conjunctive coordinator "aani" (आणि).

## **2.4 Disjunctive structures**

There are three disjunctives, kinva (किंवा :or), ka/ki (का/कि : gives meaning of or) and athava (अथवा : or) all expressing the sense of 'or'. The first, kinva (किंवा : or) is prevalent. The second, ka/ki (का/कि : gives meaning of or) is used in interrogatives and in subordinate clauses expressing the sense of 'whether'. The last is confined to the formal language. In both sentence and constituent coordination kinva (किंवा : or) is placed immediately before the last sentence or constituent as the case may be. It may also appear before each sentence or sentential constituent. It is never placed in the beginning of the first sentence or first sentence constituent. Although kinva (किंवा : or) allows a juxtaposed sequence like aani (आणि: and), unlike aani (आणि: and) it may however not be totally absent from the sequence. The last placement of kinva (किंवा: or) is obligatory. Following is one example,

Ex 9: lili ghari geli asel kinva baget basali asel.

(लिली घरी गेली असेल किंवा बागेत बसली असेल : Lili may have gone home or she may be sitting in the garden)

### 2.5 Adversative Structures

The three adversative coordinators pan (पण: but), parantu (परंतु: but) and tathapi (तथापि: but) expressing the sense of 'but' are semantically identical except in their usage. The last one is used mostly in formal contexts. The first two are nearly exchangeable. Adversative conjunctions encode a contrast with various semantic implications, for example

Ex 10: lili hushar aahe pan abhyas karat nahi (लिली हशार आहे पण अभ्यास करत नाही: Lili is intelligent but does not study )

# **3. DEVELOPING LINKS FOR MARATHI COMPOUND SENTENCES**

We have adopted two level linking schemes specifically considering complex sentences and compound sentences. The challenge in dealing such sentences is crossing of the links. Crossing of the links occurs due to violating planarity rule which states that links drawn between two words shall not cross any other link connecting any pair of words. Planarity cannot always be preserved in free word order languages. Considering Marathi compound sentences, we observed that coordination either sentential coordination or constituent coordination is used majorly.

In two level linking the compound sentences are split in two levels. The upper level deals with coordinators encountered in a sentence to be parsed and lower level deals with the inner clause placed before any coordinators. We have proposed various link types for compound sentences based on their functionality. These functional link names helps in identifying the roles it is playing linking between two distinct sentences or constituents. Following diagram illustrates two level linking scheme,



Figure 2 Two level linking in compound sentence

The links proposed to connect sentential and constituent structures are summarized in a table below followed by brief description of each identified structure with an illustration,

Sr No	Link Name	Functionality of link		
1	SeCC	Sentence to Conjunctive Coordinator		
2	CCSe	Conjunctive Coordinator to Sentence		
3	SeDC	Sentence to Disjunctive Coordinator		
4	DCSe	Disjunctive Coordinator to Sentence		
5	SeAC	Sentence to Adversative Coordinator		
6	ACSe	Adversative Coordinator to Sentence		
7	SCC	Subject to Conjunctive Coordinator		
8	CCS	Conjunctive Coordinator to Subject		
9	OCC	Object to Conjunctive Coordinator		
10	CCO	Conjunctive Coordinator to Object		
11	AjCC	Adjective to Conjunctive Coordinator		
12	ССАј	Conjunctive Coordinator to Adjective		
13	AvCC	Adverb to Conjunctive Coordinator		
14	CCAv	Conjunctive Coordinator to Adverb		
15	SDC	Subject to Disjunctive Coordinator		
16	DCS	Disjunctive Coordinator to Subject		
17	ODC	Object to Disjunctive Coordinator		
18	DCO	Disjunctive Coordinator to Object		
19	AjDC	Adjective to Disjunctive Coordinator		
20	DCAj	Disjunctive Coordinator to Adjective		
21	AvDC	Adverb to Disjunctive Coordinator		
22	DCAv	Disjunctive Coordinator to Adverb		

Table 2: Proposed	Links for	r Complex	Sentence	structures

## 3.1. CO1

The structure identified in following figure is two sentences coordinated with conjunctive coordinator "aani" (आणि). In this structure links proposed to connect sentence 1 to conjunctive coordinator is SeCC and from conjunctive coordinator to sentence 2 it is CCSe.



Figure 3 Compound sentence structures 1

## 3.2. CO2

In this structure of compound sentence two separate sentences are connected with Disjunctive Coordinator "kinva" (किंवा). The links proposed to connect sentence 1 to disjunctive coordinator is SeDC and DCSe link is proposed to connect disjunctive coordinator to sentence 2.



Figure 4 Compound Sentence structure 2

## 3.3 CO3

The following figure shows the compound sentence structure 3 in which sentence 1 is connected with sentence 2 with an Adversative coordinator "pan"(पण). Links proposed to connect it are sentence 1 to adversative connector is SeAC and adversative connector to sentence 2 is ACSe.



Figure 5 Compound sentence structure 3

#### 3.4 CO4, CO5, CO6 and CO7

In Constituent coordination subject, object, adjective and adverb can be connected together in a sentence. In this category of coordination such constituent or word level coordination is connected with conjunctive coordinator. Following figure shows these structures,



Figure 6 Compound sentence structure 4: subject coordination with Conjunctive



Figure 7 Compound sentence structure 5; Object coordination with Conjunctive



Figure 8 Compound sentence structure 6; Adjective coordination with Conjunctive



Figure 9 Compound sentence structure 7; Adverb coordination with Conjunctive

### 3.5 CO8, CO9, CO10 and CO11

As discussed in above section constituent coordination subject, object, adjective and adverb builds links with disjunctive coordinator too, which is illustrated in following figure



Figure 10: Compound sentence structure 8; Subject coordination with Disjunctive



Figure 11: Compound sentence structure 9; Object coordination with Disjunctive



Figure 12: Compound sentence structure 10; Adjective coordination with Disjunctive



Figure 13: Compound sentence structure 11; Adverb coordination with Disjunctive

We have modelled compound sentences in the form of possible valid linkage and proposed various links to connect the sentential and constituent structures in appropriate way. Our system identifies 11 such compound sentence structures.

## **4.** CONCLUSIONS

In this paper we have proposed links for compound sentence structure of Marathi language in Link Grammar framework. By studying compound sentence structure of Marathi language, links were developed to build connection on the sentential and constituent level. Total 22 new links are proposed. More such structures will be studied and links will be developed by using this framework.

# **5. ACKNOWLEDGEMENT**

The authors are thankful to the University Grants Commission, New Delhi, India forSupporting this research under the Special Assistance Programme (SAP) at the level of DRS-I (No: F.3-52/2011(SAP-II).

## REFERENCES

- [1] A. Bharati, V. Chaitnya & R. Sangal (1995) Natural Language Processing: A Paninian Perspective, New Delhi: Prentice-Hall of India.
- [2] R. V. Dhongade & K. Wali (2009) Marathi, Amsterdam/Philadelphia: John Benjamins Publishing Company.
- [3] H. Gune, M. Bapat, M. Khapara & P. Bhattacharya, (2010), "Verbs are where all the action lies: experiences of shallow parsing of a morphologically rich language", In Proceedings of the 23rd International Conference on Computational Linguistics: Posters (COLING '10). Association for Computational Linguistics, Stroudsburg, PA, USA, pp 347-355.
- [4] S. R. Kolhe & B. V. Pawar (2007) "A Connectionist Approach for Learning Regular Grammars", Journal of Computer Society of IndiaVol.37 Issue No. 3, PP. 79-86
- [5] S. R. Kolhe & B. V. Pawar (2010) "Learning Subset of Natural Language (Marathi) Grammar Using Neural Networks", International Journal of Computer Engineering and Computer Applications, ISSN 0964-4983, Vol. 02, No.3, pp 24-31.
- [6] J. Lafferty, D. Grinberg & D. Sleator (1995) "A Robust Parsing Algorithm for Link Grammars", Technical Report CMU-CS-95-125.
- [7] R. Pandharipande (1997) MARATHI, London: Rutledge Publication.
- [8] V. B. Patil & B. V. Pawar (2011) "Developing Verb Phrase links for Marathi link grammar parser", ICGST International journal on Artificial Intelligence and Machine learning, ISSN 1687-4846, Vol. 11, No.2, 2011, pp 33-38.
- [9] V. B. Patil & B. V. Pawar (2011) "Developing Subject/Object links with Verb for Marathi Link Grammar Parser", In the proceedings of the National Conference on Advances in Computing (NCAC-2011), ISBN 978-81-910591-2-0, 2011, pp 285-288.
- [10] V. B. Patil & B. V. Pawar (2012) "Developing links for Marathi noun phrase morphology for proposed Link Grammar parser", Karpagam Journal of Computer Science, Vol. 6, No. 6, ISSN 0973-292, 2012, pp 291-299.
- [11] V. B. Patil & B. V. Pawar (2013) "Influence of Karaka Relation in the framework of Marathi Link Grammar Parser", in National Conference on Advances in Computing (NCAC'13), ISBN 978-81-910591-7-5, 2013, pp 255-258.
- [12] H. B. Patil, A. S. Patil & B. V. Pawar (2014) "Part-of-Speech Tagger for Marathi Language using Limited Training Corpora", International Journal of Computer Applications, 0975 – 8887, pp 33-37.

- [13] B. V. Pawar (2001) "LA Grammar Formalism and Parsing of Simple Marathi Sentences using LA algorithm", Indian Linguistics, Journal of Linguistic society of India, Vol. 62, No.1-4, pp 141-154.
- [14] B. V. Pawar (2004) "Comparison of results of Tomita's algorithm and LA algorithm for parsing Marathi sentences", Indian Linguistics, Journal of Linguistic society of India, Vol. 65, No1-4, pp 133-140.
- [15] B. V. Pawar & N. S. Chaudhari (2000) "Marathi language Grammar Parsing using Tomita's approach", Indian Linguistics, Journal of Linguistic society of India, Vol. 61, No.1-4, pp 69-96.
- [16] D. D. K. Sleator & D. Temperley (1991) "Parsing English with a Link Grammar", Technical Report CMU-CS-91-196.
- [17] G.V.Garje & G.K.Kharate (2013) "Survey of Machine Translation Systems in India", International Journal on Natural Language Computing, Vol. 2, No. 4, pp 47-67.

#### Authors

Vaishali B. Patil has completed Bachelor of Computer Science in 2000 and Master of Computer Application (MCA) in 2003 from North Maharashtra University, India. She is currently pursuing her Ph.D in Computer Science at School of Computer Sciences, North Maharashtra University, India under the supervision of Dr B. V. Pawar and has total 11 years of teaching experience in PG course at Institute of Management Research and



Development, Shirpur (MS), India. She is a member of Computer Society of India. Her research interests include Natural Language Processing, Information Retrieval and Intelligent Tutoring Systems.

**Dr B. V. Pawar** Received his B. E. (Production) degree in 1986 from VJTI, Mumbai University, Mumbai, (India) and M.Sc.(Computer Science) degree in 1988 from Department of Computer Science, Mumbai University, Mumbai, (India). He received his Ph.D. degree in Computer Science in 2000 at North Maharashtra University, Jalgaon (India). He is having 26 years teaching experience. Presently he is working as Professor and Director, School of



Computer Science at North Maharashtra University, Jalgaon (India). He is member of various professional bodies like CSI & LSI. He has been recognized as a Ph.D. guide for the subjects Computer Science, Information Technology & Computer Engineering by various Universities in the state of Maharashtra (India). Till date he has guided 05 students towards their Ph.D. degree. His research areas include pattern recognition, neural networks, Natural Language Processing, Web Technologies & Information Retrieval. His work has been published in various international and national journals and conferences.