Modeling Complex Sentences for parsing through Marathi Link Grammar Parser

Vaishali. B. Patil[1] and B. V. Pawar[2]
[1] Institute of Management Research and Development,
Shirpur, Maharashtra 425405, India

[2] 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 modeling 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 [5]. 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 [8] 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, Machine translation, Question Answering, etc. 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 [2, 3 , 4, 6, 7, 11, 12]. 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.

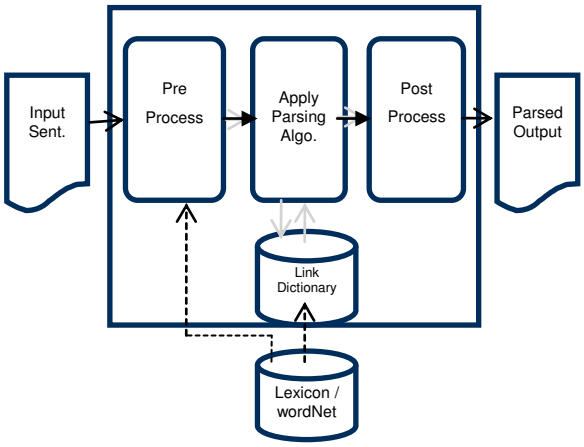

Figure 1 Block Diagram of Proposed Marathi Link Grammar Parser

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 [13, 14, 15]. Similarly we proposed 22 links for compound sentences [17]. Based on computational Panini grammar [1] we proposed Karaka links [16] 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.

Table 1: Karaka and its Links

Karaka	*Link*	*Functionality*
Karta	Ka_Karta	Verb to Subject
Karma	Ka_ Karma	Verb to Object
Karan	Ka_ Karan	Verb to instrument of the activity
Adhikaran	Ka_ Adhikaran	Verb to time and place of the activity
Sampradan	Ka_ Sampradan	Verb to word which gives donation meaning
Aapadan	Ka_ Aapadan	Verb to word which gives separation meaning

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. Complex Sentences in Marathi

In Marathi language complex sentences are either of the complement or the correlative type. In both the types there is certain interdependence between the main and the dependent clause [9, 10].

A complement clause is embedded under a main clause and may be finite, non finite or small clause. Marathi complement system is complex. The Principal Complementizer is "ki" (कि). "ki"(कि) precedes the complement clause and in main clause words such as "asa/he/hi goshta/ (asa) mhanun" (असं /हे /हि गोष्ट/ (असं) म्हणून: so/this/this story/ saying so) are included. There exist many variations of complement structure.

A correlative structure consists of a pair of clauses containing relative and correlative elements in mutual relationship. The relative clause is considered subordinate to the correlative. It usually precedes the correlative though other orders are also found. Each clause carries its own relative marker J and correlative marker T. Relative and correlative markers handled in our system are "Ji-Ti" (जी-ती), "Jar-Tar" (जर-तर), "Jevha-Tevha" (जेव्हा-तेव्हा), "Jyane-Tyane" (ज्याने-त्याने), "Jo-To" (जो-तो), "Jya-Tya" (ज्या-त्या), "Jine-Tine" (जिने-तिने), "Jase-Tase" (जसे-तसे) etc.

3. Modeling Complex Sentences for Marathi LG Parsing

Marathi complex sentences can usually be expressed in more than one way. The linking scheme for Marathi complex sentences is developed so that linking of all types of structure is consistent.

The biggest challenge dealing with complex sentences is crossing of the links. That is planarity rule. We observed that, in general planarity cannot be maintained for Marathi complex sentences. For eg. following complex sentence violet the planarity rule if system builds links in its usual manner.

Sentence – Ji mulgi ghari geli Ti dha aahe (जी मुलगी घरी गेली ती ढ आहे : The girl who went home is stupid)

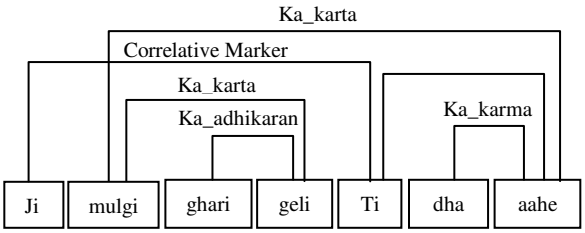

Figure 2 Crossing of the Links

The crossing of the links occurs because of the correlative structure. In above example since mulgi (मुलगी : girl) is subject of the verb phrase "dha aahe" (ढ आहे : stupid is), ka_karta link is also required in it and so crosses the correlative marker "Ti" (ती).

To avoid such crossing of links complex sentences can be parsed in two levels: the first level giving the clausal links and the second level giving the internal clause links. That is splitting the parse structure in two levels the upper level deals with relative-correlative marker and chunks of clauses and lower level deals with the words within the clause. New links are proposed to have valid and functional linkage between the words of complex sentences.

Sentence – Ji mulgi ghari geli Ti dha aahe (जी मुलगी घरी गेली ती ढ आहे: The girl who went home is stupid)

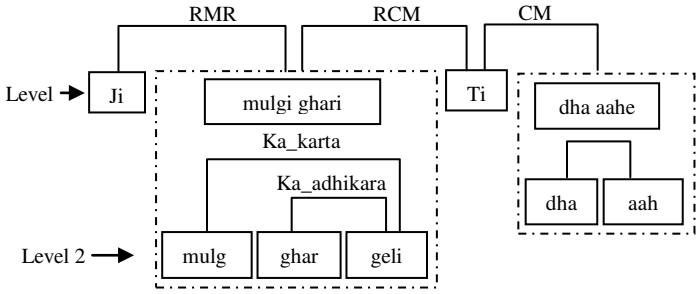

Figure 3 Two Level Linkage Parsing

The links proposed as shown in above figure are RMR which connects relative marker to relative clause, link RCM connects relative clause to correlative marker and link CMC connects the correlative marker to correlative clause.

4. Modeling Complex Sentences for Marathi LG Parsing

Possible complex sentence structures were studied and modeled for Marathi link grammar parsing system. The links proposed to connect clauses, header, Complementizer etc are summarized in a table below, followed by brief description of the modeled complex structure and proposed links in it.

Table 2: Proposed Links for Complex Sentence Structures

Sr No	*Link Name*	*Functionality of link*
1	HM	Header to Main Clause
2	HC	Header to Complementizer
3	MCO	Main Clause to Complementizer
4	COC	Complementizer to Complement clause
5	CH	Complement Clause to Header
6	SH	Subject to header
7	CAM	Complement Clause to “Asa Mhanun”
8	OC	Object to Complement Clause
9	SM	Subject to Main Clause
10	RMR	Relative Marker to Relative Clause
11	RMCM	Relative marker to Correlative Marker
12	RCM	Relative Clause to Correlative marker
13	CMC	Correlative Marker to Correlative clause
14	CMRM	Correlative marker to Relative Marker
15	CRM	Correlative clause to Relative Marker
16	CMS	Correlative Clause to Subject
17	RC	Relative Clause to Correlative Clause
18	HS	Header to Subject
19	SC	Subject to Correlative clause
20	ADM	Adverbial Cause to Main Cause
21	MCP	Main clause to Conjunctive Particle
22	CPA	Conjunctive Particle to Adverbial Clause

4.1 CX1:

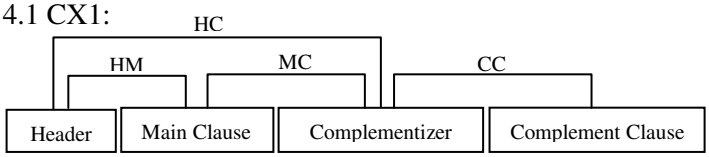

Figure 4: Complex Sentence Structure 1

Links proposed to connect complement type complex structure are HM which connects Header "hi" (हि) to main clause, MC connects main clause to Complementizer "ki" (कि), CC which connects Complementizer to complement clause. Eg – Hi goshta vichitra aahe Ki liliNe lagna kela (हि गोष्टं विचित्र आहे कि लीलीने लग्न केलं : The story that Lili got married is strange)

4.2 CX2:

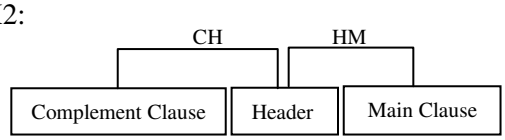

Figure 5: Complex Sentence Structure 2

In this structure the Complementizer is absent; this is the variation of complement clause. In such structure link CH is used to connect complement clause to header.

Eg – LiliNe lagna kela Hi goshta vichitra aahe.(लीलीने लग्न केलं हि गोष्टं विचित्र आहे: Variation of , The story that Lili got married is strange)

4.3 CX3:

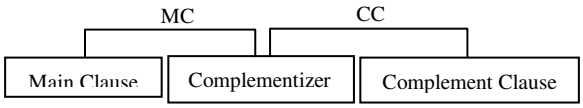

Figure 6: Complex Sentence Structure 3

This is another variation of complement clause, here header is absent and it is still grammatical. Link MC is used to connect main clause to Complementizer.

Eg LiliLa mahit aahe Ki mini ithe nahi. (लिलीला माहित आहे कि मिनी इथे नाही: Lili knows that Mini is not here)

4.4 CX4:

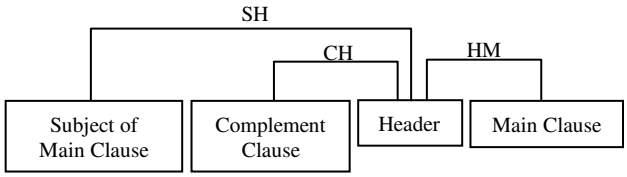

Figure 7: Complex Sentence Structure 4

In this structure, subject of main clause is separated from main clause and positions before complement clause without header. Link SH is proposed to connect subject with header of main clause.

Eg : LiliLa mini ithe nahi asa vatat(लिलीला मिनी इथे नाही असं वाटतं :Lili believes / thinks that Mini is not here)

4.5 CX5:

This is the correlative structure, which is explained in Figure 3.

4.6 CX6:

There are other variations exists like deletion of relative marker, which gives following structure

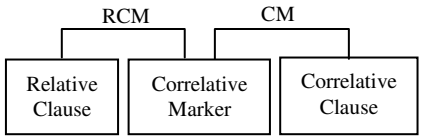

Figure 8: Complex Sentence Structure 8

For eg – ghari geli ti mulgi dha aahe (घरी गेली ती मुलगी ढ आहे : variation of, The girl who went home is stupid)

4.7 CX7:

Another variation to this structure is,

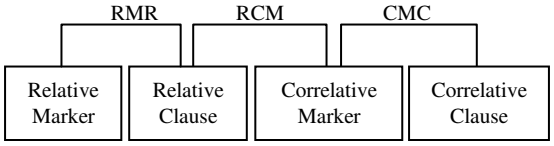

Figure 8: Complex Sentence Structure 9

Eg – Ti mulgi dha aahe Ji mulgi ghari geli (ती मुलगी ढ आहे जी घरी गेली: variation of, The girl who went home is stupid)

Based on this structures or types, it is observed that in correlative clause structure four patterns exists,

1. Full Correlatives – In this relative and correlative markers as well as clauses exists.
2. Gap Relatives – In such structures there is deletion of relative marker and noun common to both clauses.
3. Free Relatives – These structures are headless relatives
4. Multiple headed relatives – In multiple headed relative clauses several Noun Phrases are simultaneously relativized.

We have modeled complex sentences in the form of possible valid linkage and proposed various links to connect the clauses in appropriate way. Our system identifies 20 such complex sentence structures.

5. Conclusions

In this paper we have explained how we have modeled complex structure of Marathi language in Link Grammar framework. By studying complex sentence structure of Marathi language links were developed to build connection between the clauses. Total 22 new links are proposed. The paper also proposes two level linkage representations to resolve crossing of the links problem. This work helps to parse complex sentences through our proposed Marathi Link Grammar Parser. Sentence structure can be modeled in similar way to other Indian languages as well.

6. Acknowledgement

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

References

[1] A. Bharati, V. Chaitnya and R. Sangal. Natural Language Processing: A Paninian Perspective, New Delhi: Prentice-Hall of India, 1995.

[2] B. V. Pawar. 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, 2001, pp 141-154.

[3] B. V. Pawar. Comparision 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, 2004, pp 133-140.

[4] B. V. Pawar and N. S. Chaudhari. Marathi language Grammar Parsing using Tomita's approach, Indian Linguistics, Journal of Linguistic society of India, Vol. 61, No.1-4, 2000, pp 69-96.

[5] D. D. K. Sleator and D. Temperley. Parsing English with a Link Grammar, Technical Report CMU-CS-91-196, 1991.

[6] H. Gune, M. Bapat, M. Khapara and P. Bhattacharya. 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, 2010, pp 347-355.

[7] H. B. Patil, A. S. Patil and B. V. Pawar. Part-of-Speech Tagger for Marathi Language using Limited Training Corpora, International Journal of Computer Applications, 0975 – 8887, 2014, pp 33-37.

[8] J. Lafferty, D. Grinberg and D. Sleator. A Robust Parsing Algorithm for Link Grammars, Technical Report CMU-CS-95-125, 1995.

[9] R. V. Dhongade and K. Wali. Marathi, Amsterdam/Philadelphia: John Benjamins Publishing Company, 2009.

[10] R. Pandharipande. MARATHI, London: Rutledge Publication, 1997.

[11] S. R. Kolhe and B. V. Pawar. A Connectionist Approach for Learning Regular Grammars, Journal of Computer Society of India, Vol.37, No. 3, 2007, pp. 79-86

[12] S. R. Kolhe and B. V. Pawar. 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, 2010, pp 24-31.

[13] V. B. Patil and B. V. Pawar. 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.

[14] V. B. Patil and B. V. Pawar. 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.

[15] V. B. Patil and B. V. Pawar. 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.

[16] V. B. Patil and B. V. Pawar. 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.

[17] V. B. Patil and B. V. Pawar. Developing Links of Compound Sentences for Parsing through Marathi link Grammar Parser, International Journal on Natural Language Computing (IJNLC) Vol.3, No.5/6, December 2014, pp 1-9.

Vaishali. B. Patil

She 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

He 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.