

## Parsing for Indian Languages: A Literature Survey

Manisha Prajapati<sup>1\*</sup>, Archit Yajnik<sup>2</sup>

<sup>1</sup>Gujarat Technological University, Ahmadabad, India

<sup>2</sup>Sikkim Manipal University, Sikkim, India

\*Corresponding Author: purvansi263@gmail.com

Available online at: [www.ijcseonline.org](http://www.ijcseonline.org)

Accepted: 14/Aug/2018, Published: 31/Aug/2018

**Abstract**— Syntactic parsing is an important undertaking which is required for NLP applications including machine interpretation. It is a testing assignment to build up a subjective parser for morphologically rich and agglutinative dialects. Syntactic investigation is utilized to comprehend the linguistic structure of a characteristic dialect sentence. It yields all the linguistic data of each word and its constituent. Likewise, issues identified with it assist us with understanding the dialect in a more point by point way. This writing study is preparation to comprehend the distinctive parser advancement for Indian dialects and different methodologies that are utilized to grow such apparatuses and procedures. This paper gives a study of research papers from surely understood diaries and meetings.

**Keywords**— Morphological examination, Syntactic Parsing, NLP

### I. INTRODUCTION

Syntactic investigation is the way toward examining and deciding the linguistic structure of a sentence concerning a given formal language structure. Syntactic Parsing of a characteristic dialect sentence is thought to be an essential middle of the road arrange for semantic investigation that can impact numerous pipelined uses of Natural Language Processing, for example, data extraction, word sense disambiguation and so forth.

The investigation of structure of sentence is called linguistic structure. It endeavours to portray the syntactic request in a specific dialect in term of guidelines which in detail clarify the basic structure and a transformational procedure. Linguistic structure gives guidelines to assemble words to shape parts of sentences and to assemble these segments to frame important sentences. In view of the significant uncertainty display in the human dialect, whose utilization is to pass on various semantics, it is much hard to plan the highlights for common dialect preparing undertakings. The primary test is the inborn multifaceted nature of phonetic wonders that makes it hard to speak to the successful highlights for the objective learning models [3].

India is a nation having assortment of dialects real one are Indo-Aryan Languages and Dravidian Languages. Some Corpus based NLP assignments for well-known dialects like English, Greek and so on has been worked with progress. Despite what might be expected, next to no has been done on

Indian dialects. One of the fundamental reasons is that no commented on prepared to utilize corpus sources accessible for such dialects. Also, Indian dialects are morphologically rich and agglutinative in nature that makes undertaking of making productive dialect particular apparatus troublesome.

### II. BACKGROUND THEORY

There exist numerous characteristic dialect Parsing strategies. These strategies are for the most part classified into three classifications: (I) administer based (ii) measurable based and (iii) summed up parsers. All the created parsers have a place with any of these classes and take after either 'top-down' or 'base up' approach. Measurable parsing systems are called "information driven" and administer based parsing procedures are called "sentence structure driven" methodologies [3].

#### A. Rule Based Parser

In lead-based approach the dialect particular guidelines are planned to recognize the best parse tree for a given syntax. Be that as it may, in this approach, as the creation rules are connected recursively which brings about covering. The issue can be unravelling effectively by utilizing Dynamic programming (DP) strategy. The reserve for sub parse trees in the DP-based parsers is known as the 'outline' and therefore the DP-based parsers are called 'graph parsers'. The CYK calculation and Early calculation have a place with control based parsers.

### **B. Statistical Based Parser**

Measurable parsing calculations gather factual information from effectively parsed sentences and resolves vagueness by encounter. The upside of measurable approach is that it covers the entire syntax use of the dialect. The execution of the factual parsers relies upon preparing corpus used to assemble measurable data about the sentence structure of the dialect. Rather than utilizing rules, measurable parsers pick the most ideal parse tree from hopefuls in view of the factual data. The disservice of this approach is that occasionally thinks of invalid succession of parse. CFG and Probabilistic Context Free Grammar (PCFG) based parsers are the cases for measurable parsers.

### **C. Generalized approach**

The structure behind both administer based and measurable parsing are comparable. Utilizing this preferred standpoint, Melamed recommended another summed up parsing calculation which depended on semi ring parsing thought. This summed up calculation comprises of five segments, for example, language, rationale, semi ring, and inquiry methodology and end condition. In which, punctuation characterizes terminal and non-terminal images, and additionally an arrangement of generation rules. Rationale characterizes the component of how the parser keeps running by creating new incomplete parse trees. The semi ring characterizes how fractional parse trees are scored. The hunt methodology characterizes the request in which incomplete parse trees are handled and the end condition characterizes when to stop the rationale fundamentally.

## **III. LIT TREATURE SURVEY FOR INDIAN LANGUAGES**

As contrast with remote dialects, a next to no work has been done in the common dialect preparing for Indian dialects. Different Parsers for Indian Languages like Hindi, Marathi, Bengali, Kannada, Telugu and Assamese are accessible however it is as yet a continuous procedure for Indian dialects. One of the essential measures of any parser is precision with the goal that exactness is additionally examined.

Joakim Nirve in 2009 [24] exhibited work to improve Malt Parser for three Indian dialects Hindi, Bangla and Telgu in NLP Tools Contest at ICON 2009. They accomplished second rank among took an interest framework. It is watched that enhanced marked connection scores are by 7-13 percent focuses and unlabelled connection scores by 2-5 percent focuses. A little test set of 150 sentences was utilized to examine the execution of the framework. The execution of the framework was somewhat better for Bangla and Hindi dialects however for Telugu it was lower than the pattern comes about It is watched that supportable change in

exactness can be accomplished by expanding the measure of preparing informational collection.

Prashanth Mannem in 2009[1] proposed a bidirectional reliance parser for Hindi, Telugu and Bangla dialects The created parser utilizes a bidirectional parsing calculation with two tasks projection and non-projection to manufacture the reliance tree. The execution of the parser was assessed in view of the test information sentences. He revealed that the framework accomplishes a marked connection score of 71.63%, 59.86% and 67.74% for Hindi, Telugu and Bangla separately on the Treebank with fine-grained reliance names. In light of the coarse-grained marks the reliance parser accomplished 76.90%, 70.34% and 65.01% correctness's individually.

Bharat Ram Ambati et al. in 2009 [24] investigated two information driven parsers called Malt and MST on three Indian dialects in particular Hindi, Bangla and Telugu. They combined both the preparation and advancement information and completed 5-overlap cross-approval best settings from the cross approval tests and these settings are connected on the test information of the challenge. At long last they assessed the individual and normal outcomes on both coarse-grained and fine-grained label set for these three dialects. They found that for every one of the dialects Malt performed better finished MST+ maxent. They additionally altered the usage of MST to deal with vibhakti and TAM markers for naming. They revealed that, the normal of best unlabelled connection, marked connection and named correctness's are 88.43%, 71.71% and 73.81% separately.

Akshar Bharati et al. in 2009 [5] proposed a straightforward parser for Indian dialects in a reliance structure. They depict syntactic parser which takes after a language driven approach. They depicted a sentence structure arranged model that makes utilization of semantic highlights to recognize relations. The proposed parser was displayed in light of Paninian linguistic approach which gives a reliance sentence structure system. They likewise looked at the genius postured parser execution against the past comparative endeavours and detailed its proficiency. They had looked at its execution against past comparative endeavours and announced its productivity. They have demonstrated how by utilizing basic yet strong tenets one can accomplish elite in the recognizable proof of different levels of reliance relations.

Meher Vijay Yeleti and Kalyan Deepak in 2009 [24] built up an imperative based Hindi reliance parser. In the proposed framework a sentence structure driven approach was supplemented by a controlled measurable technique to accomplish superior and strength. The created framework utilizes two phase limitation based half and half way to deal with reliance parsing. They have characterized two phases and this division prompts particular distinguishing proof and

determination of particular reliance relations at the two phases. They additionally utilized hard imperatives and delicate limitations to fabricate a proficient and powerful half and half parser. From the test they discovered that the best named and unlabelled connection correctness's for Hindi are 62.20% and 85.55% individually.

Phani Gadde et al. in 2010[10] portrays an information driven reliance parsing approach which utilizes data about the statements in a sentence to enhance the execution of a parser. The clausal data is included consequently utilizing a fractional parser. They showed the examinations on Hindi, a morphologically rich, free-word-arrange dialect, utilizing a changed rendition of MST Parser. They did every one of the examinations on the ICON 2009 parsing challenge information. They accomplished a change of 0.87% and 0.77% in unlabelled connection and named connection exactness's individually finished the pattern parsing correctness's.

Bharat Ram Ambati et al. in 2010 [2] breaks down the relative significance of various etymological highlights for information driven reliance parsing of Hindi. The investigation demonstrates that the best pick up in exactness originates from the expansion of transform syntactic highlights identified with case, tense, viewpoint and methodology. They had joined the highlights from the two parsers and accomplished a named connection score of 76.5%, which is 2 rate focuses superior to the past best in class.

Swati Ramteke et al in 2014[18] created vocabulary Parser for Devanagari content (Hindi), it indicates how a Hindi sentence is parsed into tokens and afterward discover the connection between tokens utilizing language structure and by utilizing semantic portrayal produce a parse tree. They utilized Rule based way to deal with settle the disambiguity of words. Labelling and tokenization calculations were created and executed for Devanagari content. The exactness of 89.33% was accomplished from Lexicon parser. From the examinations, it has been watched that the precision was low when they tried greater uncertainty sentences and sentences of future tense. Correspondingly, when they tried sentences of straightforward present and past tenses then the precision was high.

Sankar De et al. in 2009 [9] proposed a limitation based Dependency parsing framework and connected to a free-word arrange dialect Bangla . They have utilized a structure improvement and request fulfilment way to deal with reliance parsing in Bangla dialect. An outstanding and extremely compelling punctuation formalism with the expectation of complimentary word arrange dialect called Paninian Grammatical model was utilized for this reason. The essential thought behind this approach is to rearrange the

unpredictable and compound sentential structures in the first place, at that point to parse the straightforward structures so acquired by fulfilling the 'Karaka' requests of the Demand Groups (Verb Groups) and to re-join such parsed structures with proper connections and Karaka names. A Treebank of 1000 commented on sentences was utilized for preparing the framework. The execution of the framework was assessed with 150 sentences and correctness's accomplished are of 79.81%,90.32%, 81.27% and for marked connections, unlabelled connections and name scores individually.

Aniruddha Ghosh et al. in 2009 [11] proposed a reliance parser framework for Bengali dialect. They have performed two separate examinations for Bengali. Measurable CRF based model took after by a control based post-handling method has been utilized. They have utilized ICON 2009 datasets for preparing the framework. The probabilistic grouping model prepared with the morphological highlights like root word, piece tag, vibhakti, POS-tag and reliance connection from the preparation set information. The aftereffect of the standard CRF based framework is sifted by a lead based post-handling module by utilizing the yield acquired through the control based reliance parser. The framework showed an unlabelled connection score (UAS) of 74.09%, named connection score (LAS) of 53.90% and marked precision score (LS) of 61.71% separately.

Sanjay Chatterji and et al. in 2009 [8] proposed a mixture approach for parsing Bengali sentences. The proposed framework depended on information driven reliance parser. So as to enhance the execution of the framework, some hand-created rules are recognized in light of the blunder designs on the yield of the pattern framework.

Akshar Bharati and Rajeev Sangal portrayed a sentence structure formalism called the 'Paninian Grammar Framework' that has been effectively connected to all free word Indian dialects [6]. They have portrayed a limitation based parser. Paninian system utilizes the documentation of karaka relations amongst verbs and things in a sentence. It is discovered that the Paninian structure connected to present day Indian dialects will give a rich record of the connection amongst vibhakti and karaka parts and that the mapping is exquisite and minimized.

In the year 2009, B.M. Sagar et al proposed a method for delivering setting free language structure for the Noun Phrase and Verb Phrase assertion in Kannada Sentences [20]. In this approach, a recursive drop parser is utilized to parse the setting free sentence structure. The framework works in two levels: First of all, it produces the CFG of the sentence. In the second level, a recursive plummet parser called Recursive Descent Parser of Natural Language Tool Kit (NLTK) was utilized to test the punctuation. To put it plainly, it is a language structure checking framework in

which for a given sentence parser decides if the sentence is linguistically right or wrong contingent on the Noun and Verb assertion. They have tried the framework with 200 example sentences and got empowering comes about.

Antony P J et al. in 2010 [12] have built up a Penn Treebank based measurable syntactic parsers for Kannada dialect. The notable sentence structure formalism called Penn Treebank structure was utilized to make the corpus for proposed factual syntactic parser. The parsing framework was prepared with 1,000 Kannada sentences. The created corpus has been now clarified with redress division and Part-Of-Speech data. The engineers utilized their own SVM based POS tagger generator for relegating appropriate labels to each and each word in the preparation and testing sentences.

The proposed syntactic analyser was executed utilizing directed machine learning and probabilistic setting free language structures approaches. Preparing, testing also, assessment was finished by help vector technique (SVM) calculations. Test perceptions appear that the execution of the proposed framework is altogether great and has exceptionally aggressive exactness.

B.M. Sagar et al. in 2010[19] proposed a Context Free Grammar (CFG) examination for basic Kannada sentences. They have clarified the written work of Setting Free Grammar (CFG) for a straightforward Kannada sentence with two sorts of illustrations. In the created framework, a dialect sentence structure is parsed with Top-Down and Bottom-Up parsers and they discovered that a Top-Down parser is more appropriate to parse the given linguistic creation

Rahman, Mirzanur and et al. in 2009 [17] have built up a setting free language for straightforward Assamese sentences. In these works they had thought about just set number of sentences for creating rules and just seven principle labels are utilized. They have broke down the issues that emerge in parsing Assamese sentences and create a calculation to explain those issues. The calculation is a change of Earley's Algorithm and they found the calculation straightforward and productive.

Navanath Saharia et al. in 2011 [21] depicted a parsing basis for Assamese content. They have depicted the down to earth investigation of Assamese sentences from a computational viewpoint. This approach can be utilized to parse the basic sentences with numerous thing, descriptive word, qualifier provisos.

Dhanashree Kulkarni et al. in 2014[14] has made an endeavour to compose setting free punctuation for basic Marathi sentences. Two arrangements of cases are taken to clarify the written work of CFG. Language structure is

parsed with Top Down and Bottom-Up Parser. Top Down parser is said to be more reasonable to parse syntactic creations This paper sets a phase to create mechanized sentence structure checking strategies for a given Marathi sentence and stresses fundamentally on portrayal of CFG considered.

B. Venkata S. Kumari et al in 2012[15] presents this paper, they initially investigated Malt and MST parsers furthermore, grew best models, which they considered as the pattern models for their approach. Considering stars of both these parsers, they built up a half breed approach consolidating the yield of these two parsers in a natural way. They demonstrated that a basic framework like consolidating both MST and Malt in a natural way can perform better than both the parsers. They detailed their outcomes on both advancement and test information gave in the Hindi Shared Task on Parsing at workshop on MT also, parsing in Indian Languages, Colling 2012. Their framework's anchored named connection score of 90.66% and 80.77% for highest quality level and programmed tracks individually. The exactness's are third best and fifth best for highest quality level and programmed tracks individually.

Sambhav jain et al. in 2013 [13] presents an endeavours towards fusing outer information from Hindi Word Net to help reliance parsing. They lead parsing probes Hindi, using the data from idea ontologies accessible in Hindi Word Net to supplement the transform syntactic data officially accessible. The work is driven by the knowledge that idea ontologies catch a particular true part of lexical things, which is very particular and probably not going to be derived from transform syntactic data, for example, transform, POS-tag and piece. This supplementing data is encoded as an extra element for information driven parsing and analyses are led. They perform tests over datasets of various sizes. They accomplish a change of 1.1% (LAS) when preparing 1,000 sentences and 0.2% (LAS) on 13,371 sentences over the standard. The changes are measurably critical at  $p < 0.01$ . The higher enhancements on 1,000 sentences propose that the semantic data could address the information sparsity issue.

Pradipta Ranjan et al. in 2003[16] displayed a calculation for neighbourhood word gathering to remove settled word arrange conditions in Hindi sentences. Neighbourhood word gathering is accomplished by characterizing normal articulations for the word gatherings. Computational Paninian demonstrate. Additionally, nearby word gathering accomplished can be utilized to give contributions to pitch and ambiguities happening amid word gathering are moreover settled. Hindi being a free request dialect, settled arrange word amass extraction is fundamental for diminishing the heap on the free word arrange parser. The parser worldview being utilized is the prosody displaying units for content to discourse frameworks in Indian dialects.

Grammatical feature labelling is a fundamental prerequisite for nearby word gathering. They exhibit another calculation for grammatical feature labelling based on lexical arrangement requirements in Hindi. The calculation goes about as the principal level of grammatical form tagger, utilizing imperative spread, in view of ontological data and data from morphological investigation, and lexical standards.

Karan singla et al. [23] has tried different things with various parameters of data driven Malt Parser alongside the two-arrange pre-processing way to deal with manufacture a high calibre reliance parser for Hindi. The framework achieved best LAS of 90.99% for best quality level track and second-best LAS of 83.91% for computerized information.

Selvam M et al. in 2008 [22] proposed a factual parsing of Tamil sentences utilizing phrase structure half breed dialect display. They have fabricated a factual dialect display in view of Trigram for Tamil dialect with medium of 5000 words. In the try they demonstrated that measurable parsing gives better execution through trigram probabilities and huge vocabulary estimate. Keeping in mind the end goal to conquer a few impediments like spotlight on semantics as opposed to sentence structure, absence of help in free requesting of words and long-haul relationship of the framework, an auxiliary segment is to be consolidated. The created half breed dialect display depends on a grammatical form label set for Tamil dialect with in excess of 500 labels. The created state organized Treebank depended on 326 Tamil sentences which covers in excess of 5000 words. The state organized Treebank was prepared utilizing prompt head parsing strategy. Two experiments with 120 and 40 sentences have been chosen from prepared set and test set separately. They revealed that, the execution of the framework is superior to the syntax display.

Bharati, Akshar, et al. in 2008 [4] exhibited a paper is an endeavor at investigating and disengaging a few significant prompts introduce in the dialect which loan themselves to hearty reliance parsing. They report a progression of analyses. During the time spent these examinations they additionally thought about the execution of two uninhibitedly accessible reliance parsers and pointed their qualities and shortcomings. The outcomes acquired approve different semantic instincts which can be successfully utilized as a part of parsing. Specifically we take note of that conjoined vibhakti-name include and insignificant semantics can prompt intense change in the parser execution. Aside from this the outcomes likewise point towards some hard to learn phonetic developments.

Indian dialects are Resource poor and less favored where explained corpora isn't accessible, so to make labeled corpus is dreary and tedious assignment. Subsequently building up a well syntactic parser is a testing task. The challenges in

syntactic examination of a content are called auxiliary vagueness. Distinctive kinds of difficulties amid syntactic investigation are:

### 3.1 Scope Ambiguity: The first level of ambiguity is scope ambiguity

#### **Apurva is drinking tea and reading newspaper.**

The scope of the subject "apurva" is ambiguous. The question in the sentence is "to which activity apurva refers to i.e. whether apurva is only drinking tea or he is reading newspaper or doing both the activities. How much of the text subject is qualified? This is known as scope ambiguity means what is the region of influence i.e. the scope of the subject here. How much text does it qualifies?

### 3.2 Attachment ambiguity:

Attachment ambiguity arises from uncertainty of attaching a phrase or clause to a part of a sentence. Here are **She Never Saw a Dog and Didn't Smile.**

The sentence is a decent case of a connection uncertainty, which I've diagrammed beneath. The planned perusing is on the left, in which never joins "high", to the whole verb express observed a puppy and didn't grin. The amusing perusing is on the right, where never joins "low", to only the VP saw a puppy.

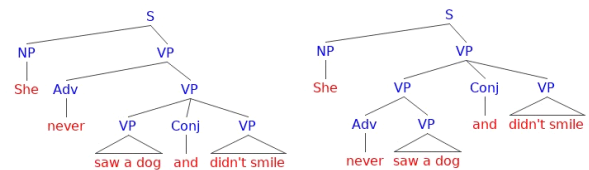


Fig.1

One exception to this is the genitive case of Hindi (का के की ka, ke, kii). But the genitive case marker always links two NPs

### आरती ने आरती के लिए थाली सजाई I

Aarti ne Aarti ke liye thali sajai.

Arti decorated plate for Arti.

Here the first Aarti is the name of a girl and the second Aarti means worship. Translating this as worship decorate plate for worship is not correct, though Arti decorated plate for Arti. Is passable. [7].

## IV. PERFORMANCE MEASURES

The Performance of any Syntactic Parser is assessed by utilizing the measures, for example, Precision, Recall, and F-measure and so on. In this respects, diverse number of sentences are utilized for testing and the preparation corpus are assemble from Word Net.

Amid preparing and testing information, the parser was prepared on the whole discharged information with the best performing highlight set and the un-explained test information was parsed with the model got.

## V. CONCLUSION

The Performance of any Syntactic Parser is assessed by utilizing the measures, for example, Precision, Recall, and F-measure and so on. In this respects, diverse number of

sentences are utilized for testing and the preparation corpus are assemble from Word Net.

Amid preparing and testing information, the parser was prepared on the whole discharged information with the best performing highlight set and the un-explained test information was parsed with the model got.

**Table: 1 survey of Indian languages**

No.	Parser name	year	Publication Detail	Language	Method/algorithm/tool	CORPUS	Accuracy
1	Lexicon Parser for syntactic and semantic analysis of Devanagari sentence using Hindi wordnet	2014	International Journal of Advanced Research in Computer and Communication Engineering (Swati Ramteke et al.)	Hindi	Rule based	1500 tokens	89.33%
2	Specifying context free grammar for Marathi sentences	2014	International Journal of Computer Applications (Dhanashree Kulkarni et al.)	Marathi	Both top-down and bottom-up parsers	Simple Marathi sentences	Top-down parser is more efficient
3	Exploring Semantic Information in Hindi WordNet for Hindi Dependency Parsing	2013	The sixth international joint conference on natural language processing (Sambhav jain et al.)	Hindi	HWN ontology	Hindi Dependency Treebank	Better accuracy is achieved
4	Hindi dependency parsing using a combined model of MALT and MST	2012	Proceedings of the Workshop on Machine Translation and Parsing in Indian Languages (B. Venkata S. Kumari et al.)	Hindi	Hybrid approach (MALT +MST)	gold standard automatic tracks	(for gold standard) LAS-90.66% (for automatic tracks) LAS-80.77%
5	Two-stage approach for Hindi dependency parsing using malt parser	2012	Proceedings of the Workshop on Machine Translation and Parsing in Indian Languages (Karan singla et al.)	Hindi	MALT parser	gold standard track Automated data	gold standard track LAS90.99% Automated data LAS-83.91%
6	Hindi Parser- Based on CKY algorithm	2012	Int. J. Computer Technology and Applications (Nitin Humber and Ambries Shrivastava)	Hindi	CKY algorithm	Hindi Sentences	Good

7	A First Step Towards Parsing of Assamese Text	2011	Special Volume: Problems of Parsing in Indian Languages (Navanath Saharia et al.)	Assamese	Rule Based	ICON 2009 datasets	78.82%
8	On the role of morphosyntactic features in Hindi dependency parsing	2010	Proceedings of the First Workshop on Statistical Parsing of Morphologically-Rich Languages. Association for Computational Linguistics (Bharat Ram Ambati et al.)	Hindi	Data Driven	ICON 2009 datasets	LAS-76.5%
9	Improving data driven dependency parsing using clausal information	2010	11th Annual Conference of the North American Chapter of the Association for Computational Linguistics (Phani Gadde et al.)	Hindi	Data Driven dependency approach	ICON 2009 contest data	LAS-74.39% UAS-91.87% LS-76.21%
10	Penn Treebank-Based Syntactic Parsers for South Dravidian Languages using a Machine Learning Approach	2010	International journal on Computer Application (IJCA), (Antony P J)	Kannad	Stastical parser	Penn treebank 1000 kannad sentences	Good
11	Context Free Grammar (CFG) Analysis for simple Kannada sentences	2010	Special Issue of IJCTT Vol.1 Issue 2, 3, 4; 2010 for International Conference. (B.M. Sagar et al.)	Kannada	Both Top Down and Bottom-Up parsers	Kannad sentences	Top-Down parser is more efficient
12	Two methods to incorporate local morphosyntactic features in Hindi dependency parsing	2010	Proceedings of the First Workshop on Statistical Parsing of Morphologically-Rich Languages. Association for Computational Linguistics (Bharat Ram Ambati et al.)	Hindi	MALT and MST parse	Hindi Treebank	Good
13	Parsing Indian languages with malt parser	2009	Proceedings of the ICON09 NLP Tools Contest: Indian Language Dependency Parsing: (Joakim Nirve)	Hindi, Bagla, Telgu	Transition-based approach MALT parser	Training set of: Hindi-1651 Bangla-1130 Telgu-1615 Test sentences-150	UAS:H-90%, B-90% and T85% LAS:15-25% low
14	Simple parser for Indian languages in a dependency framework.	2009	Proceedings of the Third Linguistic Annotation Workshop. Association for	Hindi	Grammar driven approach	Hyderabad dependency treebank Total 2100	Precision96.2% Recall - 82.6%

			Computational Linguistics. (Akshar Bharati et al.)			words Training set 1300 Test set 800	
15	Parsing of part-of-speech tagged Assamese Texts	2009	IJCSI International Journal of Computer Science Issues (Rahman, Mirzanur et al.)	Assamese	Earley's Algorithm	Assamese sentences	Earley's algorithm is simple and effective
16	Dependency Parser for Bengali: the JU System at ICON 2009	2009	Proceedings of ICON09 NLP Tools Contest: Indian Language Dependency Parsing, (Aniruddha Ghosh et al.)	Bengali	Rule Based	ICON 2009 datasets	UAS-74.09% LAS-53.90% LS-61.71%
17	Grammar Driven Rules for Hybrid Bengali Dependency Parsing	2009	Proceedings of ICON-2009 7th International Conference on Natural Language Processing, Macmillan Publishers, India (Sanjay Chatterji et al.)	Bengali	Hybrid approach MALT parser	ICON 2009 datasets	Highly effective rules
18	Constraint based Hindi dependency parsing	2009	Proceedings of ICON09 NLP Tools Contest: Indian Language Dependency Parsing. (Meher Vijay Yeleti, Kalyan Deepak)	Hindi	Hybrid approach	ICON 2009 datasets Hindi data	LSA-62.20 UAS -85.55
19	Bidirectional Dependency Parser for Hindi, Telugu and Bangla	2009	Proceedings of ICON09 NLP Tools Contest: Indian Language Dependency Parsing, (Prashanth Mannem)	Hindi, Telgu Bagla,	Bidirectional Dependency parser algo	ICON 2009 datasets	LAS for: Hindi-71.63% Telgu-59.86% bangla-67.74%
20	Solving the Noun Phrase and Verb Phrase Agreement in Kannada Sentences	2009	International Journal of Computer Theory and Engineering, (B.M. Sagar et al.)	Kannad	Recursive Descent Parser	200 sentences	Good
21	Parsing Free Word Order Languages in the Paninian Framework	2009	(Akshar Bharati, Rajeev Sangal)	Hindi	Constraint based parser	ICON 2009 datasets	Efficient and effective parser
22	Structure Simplification and Demand Satisfaction Approach to Dependency Parsing in Bangla	2009	Proceedings of ICON09 NLP Tools Contest: Indian Language Dependency Parsing, Hyderabad, India (Sankar et al.)	Bangla	Constrained based dependency parser	Treebank dataset training sentences1000 test sentences-150	LAS-79.81% UAS-90.32% LS-81.27%



23	Experiments in Indian Language Dependency Parsing	2009	Proceedings of ICON09 NLP Tools Contest: Indian Language Dependency Parsing, Hyderabad, India, ( Bharat Ram Ambati et al.)	Hindi, Bangla, Telugu	Data driven MST and MALT parsers	ICON 2009 datasets	UAS-88.43% LAS-71.71% LS-73.81%
24	Two semantic features make all the difference in parsing accuracy	2008	In Proceedings of the 6th International Conference on Natural Language Processing (ICON-08), CDAC Pune, India. 2008. (Bharati, Akshar, et al.)	Hindi	MALT and MST parser	Hindi Dependency treebank 1200 sentences	LAS-69.64% UAS-88.67%
25	Structural Parsing of Natural Language Text in Tamil Using Phrase Structure Hybrid Language Model	2008	International Journal of Computer, Information and Systems Science, and Engineering. (Selvam M et al.)	Tamil	phrase structure hybrid language mode	5000 words	Good
26	Part of speech tagging and local word grouping techniques for natural language parsing in Hindi	2003	Proceedings of the 1st International Conference on Natural Language Processing(Pradipta Ranjan et al.)	Hindi	Paninian model	Hindi sentences	Improved performance of a parser

## VI. REFERENCES

- [1] A. Abhilash and P. Mannem, "Bidirectional Dependency Parser for Indian Languages."
- [2] B. R. Ambati, S. Husain, J. Nivre, and R. Sangal, "On the Role of Morphosyntactic Features in Hindi Dependency Parsing," First Work. Stat. Parsing Morphol. Rich Lang. (SPMRL 2010), no. June, pp. 94–102, 2010.
- [3] P. J. Antony and K. P. Soman, "Computational Morphology and Natural Language Parsing for Indian Languages: A Literature Survey," *Ijcsset*, vol. 3, no. 4, pp. 136–146, 2012.
- [4] A. Bharati, S. Husain, B. Ambati, S. Jain, D. Sharma, and R. Sangal, "Two semantic features make all the difference in parsing accuracy," *Proc. 6th Int. Conf. Nat. Lang. Process.*, vol. 8, no. November, 2008.
- [5] A. Bharati, M. Gupta, V. Yadav, K. Gali, and D. M. Sharma, "Simple parser for Indian languages in a dependency framework," *Proc. Third Linguist. Annot. Work.*, no. August, pp. 162–165, 2009.
- [6] A. Bharati and R. Sangal, "Parsing free word order languages in the Paninian framework," *Proc. 31st Annu. Meet. Assoc. Comput. Linguist.*, pp. 105–111, 1993.
- [7] P. Bhattacharyya, "Natural Language Processing: A Perspective from Computation in Presence of Ambiguity, Resource Constraint and Multilinguality," *CSI J. Comput.*, vol. 1, no. 2, pp. 1–11, 2012.
- [8] S. Chatterji and D. Roy, "Grammar Driven Rules for Hybrid Bengali Dependency Parsing," 2009.
- [9] S. De, "Structure Simplification and Demand Satisfaction Approach to Dependency Parsing in Bangla." *Proceedings of ICON09 NLP Tools Contest: Indian Language Dependency Parsing, Hyderabad, India, 2009.*
- [10] P. Gadde, K. Jindal, S. Husain, D. M. Sharma, and R. Sangal, "Improving Data Driven Dependency Parsing using Clausal Information," *Hum. Lang. Technol. 2010 Annu. Conf. North Am. Chapter Assoc. Comput. Linguist.*, no. June, pp. 657–660, 2010.
- [11] A. Ghosh, "Dependency Parser for Bengali: the JU System at ICON 2009," 2009.
- [12] A. P. J. N. J. Warriar Assistant Professor, and S. K. P Professor, "Penn Treebank-Based Syntactic Parsers for South Dravidian Languages using a Machine Learning Approach," *Int. J. Comput. Appl.*, vol. 7, no. 8, pp. 975–8887, 2010.
- [13] S. Jain, N. Jain, A. Tammewar, R. Ahmad, B. Dipti, and M. Sharma, "Exploring Semantic Information in Hindi WordNet for Hindi Dependency Parsing," no. October, pp. 189–197, 2013.
- [14] D. Kulkarni, "Specifying Context free Grammar for Marathi Sentences," vol. 99, no. 14, pp. 38–41, 2014.
- [15] B. V. S. Kumari and R. R. Rao, "Hindi Dependency Parsing using a combined model of Malt and MST," no. December, pp. 171–178, 2012.
- [16] A. B. P. R. Ray V. Harish and S. Sarkar, "Part of Speech Tagging and Local Word Grouping Techniques for Natural Language Parsing in Hindi," *Proc. ICON 2003*, 2003.
- [17] M. Rahman, S. Das, and U. Sharma, "Parsing of part-of-speech tagged Assamese Texts," *J. Comput. Sci.*, vol. 6, no. 1, 2009.
- [18] S. Ramteke, K. Ramteke, and R. Dongare, "Lexicon Parser for

- syntactic and semantic analysis of Devanagari sentence using Hindi wordnet,” Int. J. Adv. Res., vol. 3, no. 4, 2014.
- [19] B. M. Sagar, G. Shobha, and R. Kumar, “Context Free Grammar (CFG) Analysis for simple Kannada sentences,” Proc. Int. Conf. [ACCTA-2010] Spec. Issue IJCCT, vol. 1, no. 2, p. 2, 2010.
- [20] B. M. Sagar, G. Shobha, and R. K. P., “Solving the Noun Phrase and Verb Phrase Agreement in Kannada Sentences,” Int. J. Comput. Theory Eng., vol. 1, no. 3, pp. 288–292, 2009.
- [21] N. Saharia, “A First Step Towards Parsing of Assamese Text.”
- [22] M. Selvam, A. M. Natarajan, R. Thangarajan, and A. S. Approach, “Structural Parsing of Natural Language Text in Tamil Using Phrase Structure Hybrid Language Model,” vol. 2, no. 3, pp. 463–469, 2008.
- [23] K. Singla, A. Tammewar, N. Jain, and S. Jain, “Two-stage Approach for Hindi Dependency Parsing Using MaltParser,” no. December, pp. 163–170, 2012.
- [24] “ICON09 NLP TOOLS CONTEST: INDIAN LANGUAGE DEPENDENCY PARSING Hyderabad , India,” no. December, 2009.
- [25] P. Ghosalkar, S. Malagi, V. Nagda, Y. Mehta, and P. Kulkarni, “English Grammar Checker,” IJCSE no. 3, pp. 98–100, 2016.
- [26] T. Suryakanthi, “Parsing Complementizer Phrases In Machine Translation System,” IJCSE, vol. 4, no. 10, pp. 1702–1709, 2012.

### Authors Profile

*Manisha Prajapati* pursued Master of Science from Sardar Patel University of Anand, Gujarat in 2003. She is currently pursuing Ph.D. and currently working as Head of Department in Gujarat Technological University, Ahmedabad, Gujarat.



*Mr Archit Yajnik* pursued Ph.D. degree from Maharaja Sayaji Rao University of Vadodara, India in year 2010. He is currently Working as Additional Professor in the Department of Mathematics at SMIT, Sikkim, India. He has published more than 10 research papers in reputed International conference proceedings. His main research work focuses on Natural Language Processing, Wavelet Analysis, Pattern Recognition and Artificial Neural Networks.

