

Developing Links for Marathi Noun Phrase Morphology for Proposed Link Grammar Parser

Vaishali B. Patil¹, Bhausaheb V. Pawar²

ABSTRACT

This paper presents work done on the study of noun phrase morphology in order to develop various links for variations in the noun phrase structure of Marathi language. The paper describes development of links for articles, pre nominal modifiers, post nominal modifiers of noun phrase by studying its relative order in perspective of link grammar formalism based parser.

Keywords: Link Grammar Formalism, Natural Language Processing Applications, Marathi Noun Links

I. INTRODUCTION

Link Grammar is a theory of syntax which builds simple relations between pairs of words, rather than constructing constituents in tree like hierarchy. Link grammar provides a systematic way of parsing sentences by establishing links between the constituent words of a sentence. A link grammar consists of set of words (the terminal symbols of the grammar), each of which has a linking requirement. Linking requirements of each word are stored in a link dictionary. A sentence of the language defined by the grammar is acceptable if there exists a way to draw links among the words so as to satisfy three conditions Planarity, Connectivity and Satisfaction.

A well simplified program for parsing natural language, English with link grammar is exists. The program described in [8] reads in a dictionary and parse sentence

according to the given grammar. The program does an exhaustive search, it finds every way of parsing the given sentence with the given link grammar. The program makes use of effective data structure and heuristics to speed up parsing process.

English language is defined well in link grammar formalism. Around 60000 words which uses almost 107 link types are defined in dictionary that captures many phenomenon of English grammar like noun-verb agreement, imperatives, complex and irregular verbs and many more other things. Link grammar is a context free formalism for natural language English [11]. Absence of explicit constituents as well as high degree of lexicalization makes link grammar formalism more attractive

For Indian languages morphological analyzers and POS taggers are under development. For many Indian languages developing full fledged parser is still in its infancy.

Parsing is very fundamental requirement for many natural language processing tasks such as Information extraction [6, 12, 22], machine translation [1-3], speech recognition [25], question answering [4], automatic summarization etc. The parsed output of English language by using Link Grammar formalism is used in many applications. Link grammar formalism was used as foundation of the parser for many languages like German [10], Turkish [9], Persian [5], Arabic [26], Russian [23] etc, for few languages online parser is also available like English. In Indian Languages Link

¹ RCPET'S Institute of Management Research and Development, Shirpur, vaishali.imrd@gmail.com,
² Department of Computer Science, North Maharashtra University, Jalgaon, MS, India. bvpawar@hotmail.com

Grammar is used as an example based framework for Hindi language [7]. For Marathi language complete coverage parser is not available however based on context free grammar and left associative grammar formalism, many phenomenon have been captured for parsing Marathi [14-18, 19-21]. Developing complete parser for any natural language is difficult but certainly not impossible.

This work focuses on studying noun phrase morphology of Marathi language and developing the link types for proposed Marathi link grammar parser. The idea of link grammar based parser for the language Marathi is described in section 1. Section 2 and 3 gives a brief introduction of parsing English using Link Grammar and link types which are used for setting linking requirements of the language English. The representations of link types of English and Hindi languages are adapted for the language Marathi to some extent and links were modified or proposed new links where ever necessary with appropriate examples in section 4.

II. LG Parser for Marathi

We are developing a robust rule based parser depending on link grammar for the language Marathi. As link grammar is similar to dependency grammar which is more appropriate for free word order languages, we decided to use every advantage of this grammar formalism. A link grammar is developed by creating a dictionary of all the words in a language. By judging the roles of a particular word in different contexts, a list of possible linkages that can be associated with that word is ascertained. A sentence is said to be valid if all its words have their link bindings satisfied. Figure gives the brief idea of proposed Link Grammar Parser for Marathi.

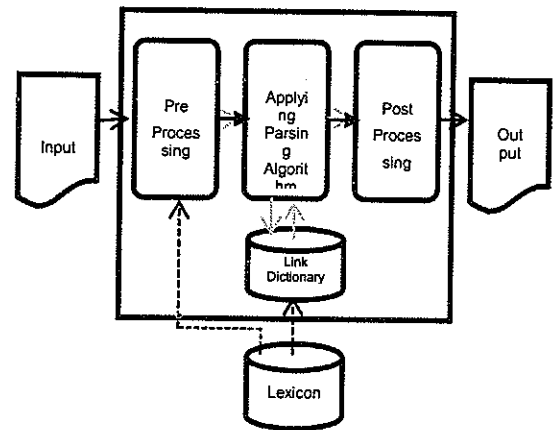


Fig 1 Block diagram of link grammar parser

The first step towards developing such parser is neat hand crafting of link dictionary. Link dictionary contains all the linking requirement of words in the form of disjuncts. Linking requirements can be fixed by studying the role of word in a sentence. The Parsing process may go through preprocessing process like separating case endings from the words for further processing and on, then parsing algorithm will be applied to find out the valid linkages for the sentences and post processing processes will help parsing in overall. Marathi shares features of Indo-Aryan languages such as SOV word order; however the word order is free in nature. There are approximately 62 million speakers of Marathi and it's the official language of Maharashtra state of the India. Marathi uses devnagari script for writing.

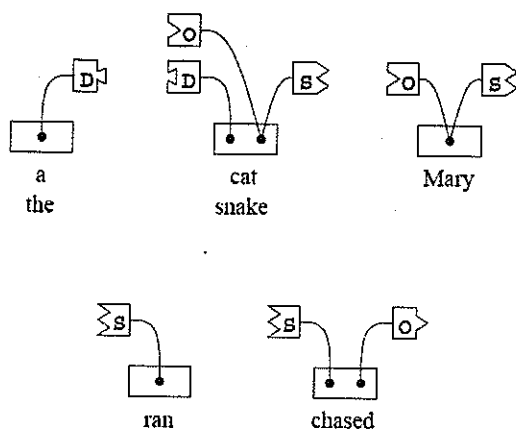
III. English Link Grammar

Link Grammar is a formal system showing how English grammar can be incorporated into a computer program and parsed by it competently. It identifies a sentence by diagnosing links among symbols. It is necessary to satisfy following three conditions Planarity – links (when drawn above the words) do not cross each other.

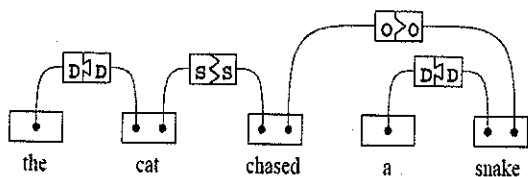
Connectivity – The links suffice to connect all the words of the sequence together.

Satisfaction – The links satisfy the linking requirements of each word in the sequence.

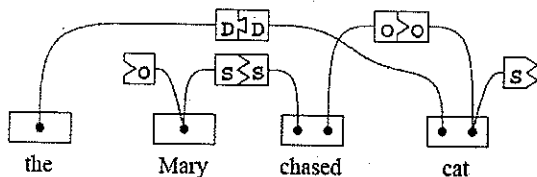
The requirements of each word are gathered in a dictionary. Figure gives the linking requirement of a simple dictionary for the words a, the, cat, snake, Mary, ran and chased [24]. Diagram above the word represents the linking requirement of that word,



The following figure shows how the linking requirements are satisfied in the sentence The cat chased a snake.



It is easy to diagnose that sequence of the words The Mary Chased cat is not acceptable for this grammar as it violates one of the three rule,



IV. The English Noun Phrase Links

English noun phrases typically consists of a noun/pronoun called the head of the noun phrase. It may further have the following optional constituents,

1. Determiners – a, an, the, this, all , one and also genitive cases
2. Pre nominal modifiers – such as adjectives, participles
3. Post-nominal modifiers – such as preposition phrase, to –infinitive

Table 1 gives the majorly used link types in English with their brief description.

Table 1 Link types in English

Link	Role of Link	Example
Determiners		
D	Connects determiner with noun	A BOY
DD	Connects the definite determiners to the number expression & adjectives acting as nouns	THE POOR
DG	Connects 'the' with proper nouns	THE GORGE BUSH
Pre-nominal Modifiers		
A	Connects pre noun adjectives to following nouns	The BIG FAT HEN seating on roof
AN	Connects noun modifiers to following noun	The TAX PROPSAL was rejected
AL	Connects determiners like 'all' or 'both' to following determiners	BOTH THE girls were pretty
G	Connects proper words together in series	BARACK HUSSAIN OBAMA
MG	Connects certain prepositions to modify proper nouns	The EMIR OF Kuwait
Post-nominal Modifiers		
M	Connects nouns to various kind of post noun modifiers	
	i) Prepositional Phrases	The MAN WITH the hat
	ii) Participle modifier	The WOMAN CARRYING
	iii) Prepositional relative	The MAN TO whom

V. THE MARATHI NOUN PHRASE

A. Order of Marathi NP

In this section we discuss the Marathi nominal modifiers and link types proposed for it by comparative study of English Link types. In Marathi generally modifiers precede the head noun which they modify except for some quantifiers. The relative order of the modifiers in a noun phrase [13] is as follows.

Demonstrative adjective; Possessive adjective; Quantifiers; Intensifiers; Adjective; relative clause; Noun; Head noun; Emphatic particle

The example of this order is,

Hi mazya bahini chi ek atishay sundar govyat rahanari maitrin. (this) (my) (sister)('s) (one) (very) (beautiful) (Goa) (staying) (Friend female)

There is a certain degree of optionality in the placement of these constituents. There has not been any systematic study of the preferred order of the modifiers relative to one another. However some generalization has been suggested which is useful in perspective of proposing link types to the words in constituents.

(i) Demonstrative adjective cannot be moved over numeral adjectives. For eg

He don mitra (These two friends)

(ii) Emphatic particles and limiters always follow the head noun. For eg.

Anu minu peksha lahan aahe (Anu Minu Compared to small is), Uday madhu ch yeil (tomorrow Madhu emph come)

(iii) The classifiers like color, measure of weight, measure of length always follow the noun of color and measure. For eg nilya ranga chi motar (blue colour. poss car), teen kilo bhazi (three kilogram vegetable)

(iv) Intensifier immediately precedes the noun. A sample list of intensifier and its use is given below,

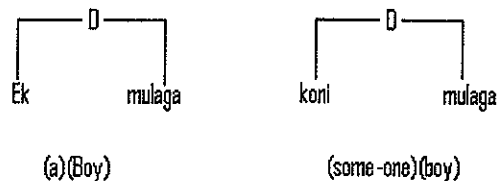
Khup – very, Far – very, Jara – a little, Ati – very, Atyant – extremely, Pushkal – quite / considerably

te ghar khup/atyant/jara/pushkal/ motha aahe (that house very/extremely/little/considerably big is)

B. Link for Marathi Articles

Marathi does not have a separate category of articles. The function of articles is carried out by numerals 'ek' (one) and the indefinite pronoun 'koni' (Some).

D – This Link connect 'ek' or 'koni' which behaves as article with the noun. For Eg,

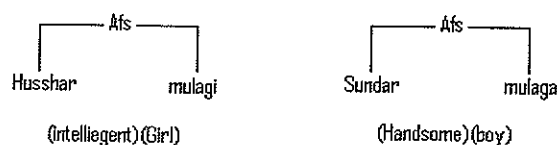


C. Link for pre nominal modifiers

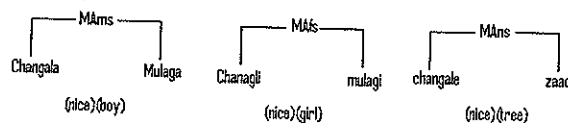
Relative order for Marathi noun phrase has been discussed in above section. Considering the facts following link types are proposed

A – This link connects the attributive adjective with the noun. Adjectives precede the noun they modify. English adjectives are uninflected they do not change morphology with the variations in the noun they qualify. But in Marathi adjectives could be inflected or uninflected. The link A is proposed for the uninflected adjectives of Marathi for eg. 'hushaar' (intelligent) is an uninflected adjective (eg hushaar mulga (intelligent boy), hushaar mulgi (intelligent girl)). A link is followed by three suffixes, the first one can either be 'm' male or 'f' female

or 'n' neuter denoting the gender of the head noun, the second suffix gives the number information i.e. 's' for singular and 'p' for plural and the third suffix 'j' can be used as a case ending marker of the head noun. Case ending marker are associated with head noun like 'mula ne' here 'ne' is case ending marker.

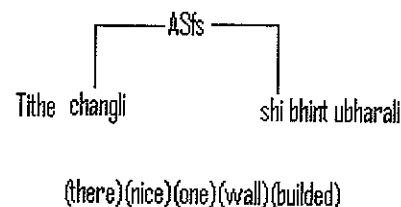


MA – This link connects the pre modifying adjective with the noun it modifies. MA link is proposed for the inflected adjectives and they too depend on gender, number and case ending of the head noun. Following are some examples of inflected adjectives 'changla mulga' (nice boy), 'changli mulgi' (nice girl), 'changle zad' (nice tree). The notion of suffix is some described as in above link type of gender 'neuter' is additional in Marathi for eg 'zaad' is neuter and so adjective used for 'zaad' too gets inflection 'changale'.

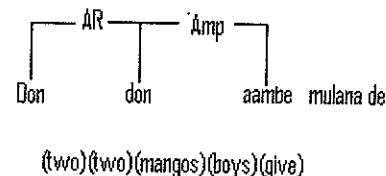


In Marathi certain form of adjectives do exists like reduplicated adjectives, some substrings like 'saa'/'she'/'shya' is attached with adjectives which do not have an equivalent structure in English so for such kind of adjectives we propose various links which are given below,

AS – This link connect the adjective with 'saa'/'she'/'shya' this link follows by three suffixes as discussed in link A.

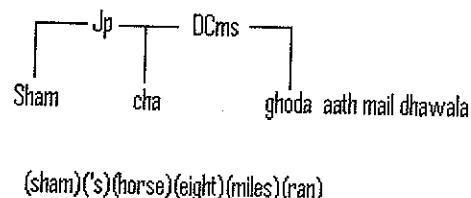


AR – In Marathi reduplicated adjectives are used like 'chan chan goshti' (nice nice stories). AR link is used to connect the first repetition of adjective and then second adjective links with noun followed by it.



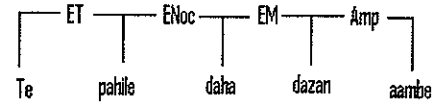
J – This link connects the nouns with case ending of possessive adjectives. Possessive adjectives precede the head noun as modifier case ending. Case ending 'cha'/'chi'/'che'/'chya' can be used as possession of noun for eg. 'sham cha ghoda' (sham's horse), 'sita chi pati' (sita's slate). Suffix p is proposed to show the possession of noun with case ending 'cha'/'chi'/'che'/'chya'.

DC – This link is proposed to connect the case ending 'cha'/'chi'/'che'/'chya' with the head noun. DC link uses suffix as explained for link A. Let us see the example of using J and DC.



Quantifiers are also the part of order in noun phrase morphology in Marathi quantifiers consists of the following co-occurring categories,

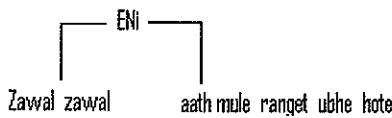
(Approximate) (Cardinal) (Collective)
 (Ordinal) (Multiplicative) (Measure)
 (Fractional)



(those)(first)(ten)(dozan)(mangos)

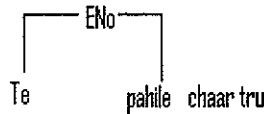
We propose following few link types for the categories of quantifiers,

ENi – This link connects the indefinite words like ‘sumare’ / ‘zawal zawal’ (approximately) to expression of quantity. It uses suffix I to show the indefiniteness of the word.



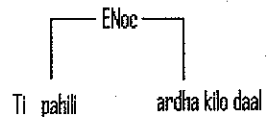
(approximately) (eight) (boys) (queue) (standing) (was)

ENo – This link connects the ordinal words like ‘pahile’ (first)/ ‘shevatache’ (last) / ‘madhale’ (in-between) to expression of quantity.



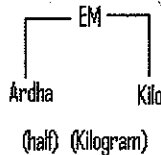
(those) (first) (four) (trucks)

ENoc – This link connects the ordinal words with the fractional measure like ‘ardha’ (half), ‘paav’ (quarter) etc. here c suffix used as count for measure.



(that) (first) (half) (kilogramme) (daal)

EM- this link connects the fractional with measure of noun.

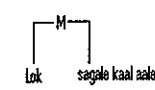


(half) (Kilogram)

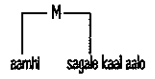
ET – This link connects the definite words ‘te’/’ti’/’tya’ with the ordinal words.

D. Links for Post nominal modifiers

Emphatic particles such as ‘ch’(only) / ‘pan’(also) / ‘suddha’ (as well) follow the noun phrase as discussed in relative order. We propose following links, M – this link connects nouns/pronouns to quantifier ‘sagale’(all) or ‘thoda’ (little)

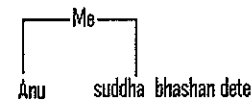


(people)(all)(yesterday)(came)



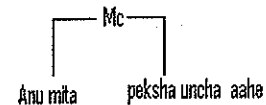
(we)(all)(yesterday)(came)

Me – This link connects emphatic particles ‘ch’ (only) / ‘hi’ (also) / ‘suddha’ (also)/’tar’(indeed)



(anu)(also)(speech)(gives)

Mc – The object of comparison may precede or follow the head noun. The suffix ‘-peksha’ is used as object of comparison. This link connects the suffix ‘-peksha’ (compared to) with noun.



(anu)(mita)(compared to)(tall)(is)

The summary for proposed link types of Marathi noun phrase is given in table 2 which could be extended or compressed depending upon its implementation, in future.

Table 2. Link Types of Marathi

Link Type for Marathi NP	Role of Link	Example
D	Connects ek/koni with noun	<i>Ek Mulga, Koni Mulga</i>
A with (G,N,C) suffix	Connects attributive Adj with noun (Uninflected Adj)	<i>Hushar Mulga, Hushar Mulgi</i>
MA with(G,N,C) suffix	Connects pre modifying adj with noun (inflected Adj)	<i>Changla mulga, changali mulgi, changale zaad</i>
AS with (G,N,C)	Connects adj with saa/she/shya case ending	<i>Pushkal saa, unch shya</i>
AR	Connects first repetition of adj with second repetition	<i>Chaan chaan, gar gar</i>
J with suffix p	Connects Possessive adj with case ending cha/chi/chya/che	<i>Sham cha ,Site chi</i>
DC with (G,N,C)	Connects possessive case ending cha/chi/chya/che with head noun	<i>Sham cha ghoda, Site chi paati</i>
M	Connects noun/pronouns to quantifiers sagale, thode	<i>Lok sagale aale hote Me kaam thode kele</i>
M with suffix e	Connects noun to emphatic particles ch/hi/sudhha/tar	<i>Anu sudhha bhashan dete</i>
M with suffix c	Connects noun with suffix -peksha	<i>Anu mita peksha uuncha aahe</i>
EN with suffix l	Connects infinite words like zawalzawal, sumare with expression of quantity	<i>Zawalzawal aath mule ranget ubhi hoti</i>
EN with suffix o	Connects ordinal words like pahile, mdhale, shevatache with expression of quantity	<i>Te pahile chaar truck ikade aale</i>
EN with suffix oc	Connects ordinal words with measure	<i>Pahili ardha kilo daal</i>
EM	Connects the fractional with measure	<i>Ardha kilo, aath fut</i>
ET	Connects definite words te/ti/tya with ordinal	<i>Te pahile daha dazan aambe</i>

VI. CONCLUSION

In this paper, we described noun phrase morphology of Marathi language and developed links that provide the syntactic relationship between the words in a noun phrase. We have considered links developed for the language English and Hindi with suitable modification to capture the characteristics of Marathi morphology. Many other variations could be discussed but due to lack of space we limited the entries of links here in this paper. Simultaneously developing links for other phenomenon like noun-verb agreement, relative clause, simple conjunctions is in progress.

VII. REFERENCES

- [1] ADJI B. T.; BAHARUDIN B.; NORSHUHANI Z. 2007a. Annotated disjunct in Link Grammar for Machine Translation, International Conference on Intelligent and Advanced Systems, 25-28 Nov. Page(s):205 - 208
- [2] ADJI B. T.; BAHARUDIN B., NORSHUHANI Z. 2007b. Building Transfer Rules using Annotated Disjunct: An Approach for Machine Translation.; 5th Student Conference on Research and Development, 12-11 Dec. Page(s):1 - 5
- [3] ADJI B. T.; BAHARUDIN B., NORSHUHANI Z. 2008. Applying Link Grammar Formalism in the Development of English-Indonesian Machine Translation System, AISC/ Calculemus/ MKM 2008, LNAI5144, pp17-23, Springer-verlag Berlin Heidelberg
- [4] BARAL C.; TARI L. 2006. Using AnsProlog with Link Grammar and WordNet for QA with deep reasoning, 9th International Conference on Information Technology, 18-21 Dec. Page(s):125 - 128
- [5] DEHDARI J.; LONSDALE D. 2005. A Link Grammar Parser for Persian, First International

- Conference on Iranian Linguistics, Germany, June 2005
- [6] DING, J; BERLEANT D; XU J; FULMER A. W. 2003. Extracting Biochemical Interactions from MEDLINE using a Link Grammar Parser. 15th IEEE International Conference on Tools with Artificial Intelligence, 3-5 Nov (In Proceeding) Page(s):467 – 471.
- [7] GOYAL SHAILLY. 2007. Example Based Parsing for Resource Deficient Languages. PhD Thesis: IIT Delhi.
- [8] GRINBERG, D.; LAFFERTY J. D.; TEMPERLEY D. 1995. A Robust Parsing Algorithm for Link Grammars, CMU-CS-95-125
- [9] ISTEK, O. 2006. A Link Grammar for Turkish, Master's Thesis: Bilkent University.
- [10] KUBLER, S. 1998. Learning a Lexicalized Grammar for German. In Proceedings of NeMLaP 3/CoNLL98, Page(s) 11-18.
- [11] LAFFERTY, J.; SLEATOR D.; TEMPERLEY D. 1992. Grammatical Trigrams: A Probabilistic Model of Link Grammar. Proceeding Probabilistic Approaches to Natural Language.
- [12] MADHYASTHA, H. V.; BALKRISHNAN N; RAMKRISHNAN K. R. 2003. Event Information Extraction using Link Grammar, 13th International Workshop on Research Issues in Data Engineering: Multi-lingual Information Management, 10-11 March (In Proceeding) Page(s):16 – 22
- [13] PANDHARIPANDE, R. 1997. MARATHI. Rutledge publications, ISBN 0-415-00319-9, London.
- [14] PAWAR, B. V.; CHAUDHARI N.S. 1996. Comparison of Tomita's and CKY's Parsing Algorithms using CFG Formalism for simple Marathi Sentences", BRI'S JAST II (II) ,PP 20-35.
- [15] PAWAR, B. V.; CHAUDHARI N.S. 1998a. Context free grammar formalism for Marathi Language and its parsing using Earley's Approach. BRI'S JAST I (II) , PP 39-68.
- [16] PAWAR, B. V, 1998b. Context Free Grammar (CFG) Formalism for Simple Marathi Sentences and its Parsing Library progress (International) ISSN 0970-1052, Vol. -18, P.P. 1-11.
- [17] PAWAR, B. V., 1998c. Context Free Grammar (CFG) Formalism for Simple Marathi Sentences and its Parsing Using CYK's Parsing Algorithm, Library progress (International) ISSN 0970 –1050, Vol. – 18 No. 1-2 (Jan.-Dec.), P.P. 23-32.
- [18] PAWAR, B. V., 2000. Marathi Language Grammar Parsing using Tomita's Approach, Indian Linguistics, Journal of the Linguistic Society of India. Vol. 61, Numbers 1-4, PP 69-96.
- [19] PAWAR, B. V., 2001. LA Grammar Formalism and Parsing of Simple Marathi Sentences Using LA Algorithm, Indian Linguistics, Journal of the Linguistic Society of India. Vol. 62, Numbers 1-4, PP 141-154.
- [20] PAWAR, B. V., 2004. Comparison of Results of Tomita's Algorithm and LA Algorithm for parsing Marathi Sentences., Indian Linguistics, Journal of the Linguistic Society of India. Vol. 65, Numbers 1-4, PP 133 -140.
- [21] PAWAR, B. V., 2007. Comparison of Tomita's and Earley's Parsing Algorithms Using CFG Formalism for Simple Marathi Sentences, BRI's

Journal of Advances in Science & Technology (JAST), Vol. 10 No. I & II.

- [22] PHYU A.L.L.; THEIN N. 2007. Domain Adaptive Information Extraction Using Link Grammar and WordNet, The Fifth International Conference on Creating, Connecting and Collaborating through Computing, 24-26 Jan., Page(s): 47- 53
- [23] PROTASOV S. Russian Link Grammar, <http://s2.ru/parser/>
- [24] SLEATOR, D. D. K.; TEMPERLEY D. 1991. Parsing English with a Link Grammar, Technical Report CMU-CS-91-196.
- [25] ZHOU, L; SHI Y; FENG J; SEARS A. 2005. Data Mining for Detecting Errors in Dictation Speech Recognition. IEEE Transaction on Speech and Audio Processing, Volume 13, Issue 5, Page(s) 681-688.
- [26] Arabic Link Grammar,
<http://www.ling.ohiostate.edu/~jonsafari/arabiclg/bin/arabiclg.html>.

Author's Biography



Vaishali B. Patil has completed Master of Computer Application (MCA) from North Maharashtra University, India. She is currently pursuing her Ph.D in Computer Science at Department of Computer Science, North Maharashtra University, India under the supervision of Dr B. V. Pawar and has total 08 years of teaching experience in PG course. Her research interests include Natural Language Processing, Information Retrieval and Intelligent Tutoring Systems.



Dr. Bhausaheb Vyankatrao 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 23 years teaching experience. Presently he is working as Professor and Head of Computer Science Department 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 today 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.