Identification of Conjunct verbs in Hindi and its effect on Parsing Accuracy

Rafiya Begum, Karan Jindal, Ashish Jain,
Samar Husain, Dipti Misra Sharma

Language Technologies Research Centre, IIIT-Hyderabad, India
rafiyabegum@gmail.com
{karan_jindal, ashishjain}@students.iiit.ac.in
{samar,dipti}@mail.iiit.ac.in

Abstract. This paper introduces a work on identification of conjunct verbs in Hindi. The paper will first focus on investigating which noun-verb combination makes a conjunct verb in Hindi using a set of linguistic diagnostics. We will then see which of these diagnostics can be used as features in a MaxEnt based automatic identification tool. Finally we will use this tool to incorporate certain features in a graph based dependency parser and show an improvement over previous best Hindi parsing accuracy.

Keywords: Conjunct verbs, Diagnostics, Automatic Identification, Parsing, Light verb.

1 Introduction

There are certain verbs that need other words in the sentence to represent an activity or a state of being. Such verbs along with the other words, required for completion of meaning, are together called *Complex Predicates* (CP). CP exist in great numbers in South Asian languages [1], [2], [3]. A CP is generally made via the combination of nouns, adjectives and verbs with other verbs. The verb in the CP is referred as light verb and the element that the light verb combines to form a CP is referred as host [4]. [5] says that in Hindi/Urdu, the light verb is taken as a contributing 'semantic structure' which determines syntactic information such as case marking whereas *host* contributes the 'semantic substance', i.e. most of the meaning the complex predicate has. [6] has talked about four types of complex predicates: (a) In Syntactic Complex Predicates the formation takes place in the syntax. (b) In Morphological Complex Predicates, a piece of morphology is used to modify the primary event predication. Well known example is morphological causatives. (c) Light Verbs cross linguistically do not always form a uniform syntactic category. They are not always associated with a uniform semantics, but they always muck around with the primary event predication. (d) In Semantics, complex predicates represent the decomposition of event structure.

In CPs, 'Noun/Adjective+Verb' combinations are called conjunct verbs and 'Verb+Verb' combinations are called compound verbs. In this paper, we are focusing

on conjunct verbs in Hindi and their identification using set of diagnostics and then we will see which of these diagnostics can be used to automate the identification process using statistical techniques and showed their usefulness in data driven dependency parsing [41]. This work can also greatly help in automatically augmenting a lexical network such as the Hindi WordNet[1]. Previous automatic identification approaches made use of parallel English corpora [19], [20] which makes use of the property that single verb in English will break into two components i.e. noun/adjective and verb in Hindi. [21] also makes use of English corpus for extracting collocation based features. To the best of our knowledge ours is the first work towards automatic identification of conjunct verbs in Hindi using only Hindi corpus. We have achieved a maximum accuracy of 85.28%. Incorporating this as a feature in graph based dependency parsing shows an improvement of 0.39% in label and 0.28% in label attachment accuracy.

The paper is arranged as follows: Section 2 gives overview of conjunct verbs in Hindi. In Section 3, we describe behavioral Diagnostics to Identify Complex Predicates. In Section 4, we discuss the subjective evaluation of diagnostics. In Section 5 and 6, we define the system for automatic identification of conjunct verb and discuss experimental results respectively. We evaluate the effect of conjunct verb on parsing accuracy and compare it with the current state-of-the-art parser in Section 7. We conclude the paper is Section 8.

2 Conjunct Verbs in Hindi

Conjunct verb in Hindi is formed by combining a noun or an adjective with a verb. These verbs have the following structure [7]:

Noun/Adjective + Verb (Verbalizer)

The most frequent *verbalizers* in Hindi are *karnaa* 'to do'*, honaa* 'to be'*, denaa* 'to give'*, lenaa* 'to take'*, aanaa* 'to come'. Take (1) as a case in point.

(1) *raam ne siitaa ko* ***kshmaa kiyaa***
ram Erg. sita Acc. forgiveness do-Past
'Ram forgave Sita.'

(2) *raam ne shyaam kii* ***madad kii***
ram Erg. shyam Gen. help do-Past
'Ram helped Shyam.'

In example (1), *kshmaa* 'forgiveness' is a noun which is combined with the verb *karnaa* 'to do' to express the sense of the verb 'to forgive'. In example (2), conjunct verb is *madad karnaa* 'to help' and the noun *madad* 'help' is linked with the object *shyaam* 'Shyam' by the postposition (Hindi case marker) *kii* 'of'.

[1] Developed by the wordnet team at IIT Bombay, http://www.cfilt.iitb.ac.in/webhwn

There are two approaches [8] which define conjunct verbs: 'Lexical approach' and 'Semanticist approach'. The aim of Lexical approach is to offer either a formal or structural justification for the recognition of the category of conjunct verbs and to specify and delimit the noun or adjective plus verb sequence as conjunct verbs. [12], [13], [14], [15] have followed Lexical approach. In 'Semanticist Approach', they tried to explore the semantic structure of the language completely abandoning the lexicalist interpretationists's goal of specifying and delimiting the noun or adjective plus verb sequence as conjunct verbs. [16], [17] have followed semanticist approach.

In this paper we also discuss the syntactic analysis of the conjunct verbs i.e., how they are treated at the syntactic level annotation of the data.

3 Diagnostics to Identify Complex Predicates

The following are some of the diagnostics mentioned in the literature [4], [18] for deciding which Noun+Verb (N+V) combinations are conjunct verbs:

I. Coordination Test (D1): This test shows that nouns of conjunct verb don't allow coordination. However it is possible to conjoin the entire N+V combination.

(3)*l*og pratiyogita meN **rucii** aur **bhaag** **le** rahe the*
People competition in interest and participation take Prog be-Past
'People were taking interest and participation in the competition.'
(4) *log pratiyogita meN **rucii le** rahe the aur **bhaag** **le** rahe the*
People competition in interest take Prog was and participation take Prog was
'People were taking interest and participation in the competition.'

Example (3) is ungrammatical because *rucii* 'interest' and *bhaag* 'participation' are conjoined by *aur* 'and', whereas these nouns are part of CP. Sentence (4) is grammatical because here the N+V combination i.e., *rucii le* 'take interest' and *bhaag le* 'participate' has been conjoined with *aur* 'and'.

II. Constituent Response Test (Wh-Questions) (D2): CP internal nouns can't be questioned. Only N+V combination can be questioned.

(5) *raam ne **jamhaaii lii***
ram Erg yawn take-Past
'Ram yawned.'
(6)**raam ne kya lii?*
ram Erg what take-Past
'What did Ram take?'
(7) *raam ne kya kiya?*
raam Erg what do-Past
'What did Ram do?'

Example (6) is ungrammatical because only noun of CP i.e., *jamhaaii* 'yawn' given in example (5) has been questioned. Whereas in (7), the N+V combination, *jamhaaii le* 'take yawn' has been questioned.

III. Relativization (D3): CP internal nominals cannot be relativized.

(8)*_vah_ ***snaan [jo bahut pavitra hai]*** _raam ne gangaa taT par_ ***kiyaa***
that baath which lot pure is ram Erg ganga bank on do-Past
'The bath which Ram did on the bank of river Ganga is very pure.'

Sentence (8) is ungrammatical because *snaan* 'bath' which is noun internal to CP has been relativized by the relative clause.

IV. Adding the accusative case marker (D4): CP internal nominal will not allow the accusative marking.

(9)*r_aam ne us_ ***jamhaaii ko liyaa*** ...
ram Erg that yawn Acc take-Past ...
'Ram took that yawn…..'

Sentence (9) is ungrammatical because *jamhaaii* 'yawn' which is noun internal to CP has taken an accusative case marker.

V. Adding the Demonstrative Pronoun (D5): CP internal nominal will not take Demonstrative Pronoun.

(10) *raam ne yah nirdesh diyaa*
ram Erg. this order give-Past
'Ram gave this order.'

In sentence (10), the demonstrative pronoun *yah* 'this' is modifying the N+V combination i.e., *nirdesh diyaa* 'gave order' and not just the Noun, *nirdesh* 'order'. To justify the above diagnostics we did a survey of these tests among native speakers of Hindi Language.

4 Diagnostics Evaluation

We conducted a survey among 20 native language speakers of Hindi to ascertain the usefulness of the diagnostics described in the previous section in identification of conjunct verb (CV). We took conjunct verbs and applied the above diagnostics to see how they fare in a subjective evaluation. Table1 below shows the results of the test. '+ve'/'-ve' reflect the usefulness of diagnostics D1-D5 for each verb. A diagnostic is deemed '+ve' if it got the desired response from >50% of the subject. A noun/adjective-verb pair is accepted as a conjunct verb (indicated by 'yes') if >=3 diagnostics are '+ve', it is not accepted as a conjunct verb (indicated by 'no') if all the diagnostics are '-ve'. The decision is 'unsure' (indicated by 'maybe') if >=3 diagnostics are '-ve'. If a diagnostic is not applicable for a verb we use a hyphen ('-') to indicate this. The cells that show +ve/-ve indicate no majority in total number of

responses. For ease of exposition, Table1 shows the result only for only 7 verbs. The study considers a total of 20 verbs.

Table1. Results of the subjective evaluation

Noun+Verb	**D1**	**D2**	**D3**	**D4**	**D5**	**CV**
rucii le 'take interest'	+ve	+ve	+ve/-ve	+ve	+ve	Yes
maar khaa 'get beaten'	+ve	+ve	-ve	+ve	+ve	Yes
bhaag le 'participate'	+ve	+ve	+ve	+ve	-ve	Yes
snaan kar 'bathe'	-ve	+ve	+ve	+ve	+ve	Yes
chalaang maar 'jump'	-	+ve	+ve	+ve	+ve	Yes
bhojan kar 'eat'	-ve	-	-ve	+ve	-ve	may be
havaa khaa 'feel air'	+ve	+ve	+ve	+ve	+ve	Yes

After exploring the behavioral diagnostics to identify conjunct verb, we will now move on to automate this task of identification. The tool will try to use the diagnostics that can be incorporated.

5 Automatic Identification of Conjunct Verb

In the previous sections, various tests were explored for manual identification of conjunct verbs. Now, we will explain the methodology used for building a statistical tool for automatic identification of conjunct verb. We didn't focus on compound verbs (verb + verb) because already a high accuracy of 98% has been reported [22]. We have learned a binary classification using maximum entropy model, which will classify a noun/adjective-verb pair into either conjunct verb or literal class (non-conjunct verb).

5.1 Corpus

We have used two different dataset that are part of Hyderabad Dependency Treebank annotated according to CPG framework [9].

1. Dataset-1: Has 4500 manually annotated sentences (200k words approx.). It was released as part of the ICON'10 tools contest on Indian Language Parsing [39]. This dataset was used as a training data.

2. Dataset-2: Has 1800 sentences. It was released as part of the ICON'09 tools contest on Indian Language Parsing [40]. This dataset was used as a testing data.

Training data has around 3749 unique consecutive noun/adjective-verb pairs out of which 1987 are unique noun/adjective and 350 unique verbs. Semantic category of each object is mined from the Hindi WordNet. The language model consisting of trigrams of words is created for training data, which is later used for extraction of various features. Testing data has 3613 noun/adjective-verb pairs out which 998 are conjunct verbs and remaining are literal expressions.

5.2 Features

Each of noun/adjective-verb pair is represented as a vector of following feature set. The features are categorized into three categories (1) Lexical (word based features like f1, f2, f3), (2) Binary features (f4, f5), (3) Collocation based (f6, f7). These features will helps in classifying a noun/adjective-verb pair into literal or conjunct verb class.

a. Verb (f1): Some verbs govern whether an object-verb pair is conjunct verb or not as compared to other verbs. They are more likely to occur as light verbs. Example of such a verb is '*kar*' (to do) which accounts for large part of conjunct verb expressions. On the other hand verbs like '*chalnA*' (to walk) occur as literal expression in most cases. Hence, verb will be a good feature for classification task.

b. Object (Noun, Adjective) Lexical (f2): Some objects are more biased towards occurring with a light verb as compared to other objects. These objects have high chances of forming conjunct verb expression with a light verb as compared to other objects.

c. Semantic Category of Object (f3): In some of the theoretical work [5], [6] importance of semantic category of a noun/adjective in identifying conjunct verb has been shown. We incorporated this feature for nouns/adjectives by extracting it from the Hindi WordNet. We referred to the first sense of topmost ontological node of a noun/adjective. Some of the possible semantic categories are 'Artifact', 'Abstraction', 'State', 'Physical Object' etc. Total semantic categories are 83; noun/adjective will fall into any of these categories, so this will help in case of unknown nouns/adjectives.
For Example: in the expression '*viSvAsaGAwa-karana*' (meaning 'to betray'), the Semantic type of '*viSvAsaGAwa*' is "Anti Social".

d. Post-Position Indicator (f4): is a Boolean feature which will indicate whether a noun/adjective is followed by a post position and then verb i.e. a post-position marker is present between noun/adjective and verb or not. Basic intuition behind this feature is that if a noun/adjective is followed by a post position than it's a possible candidate of being a part of verb argument structure. Hence, possibly the particular noun/adjective-verb pair doesn't belong to conjunct verb class, as mentioned in diagnostic number 4 (D4) in section3.

e. Demonstrative Indicator (f5): is a Boolean feature indicating presence of DEM (demonstrative tag) before noun/adjective-verb pair. This diagnostic is explained in section3 as D5.

f. Frequency of Verbs corresponding to particular Object (f6): If a noun/adjective is occurring with few verbs than it is highly probable that the given noun/adjective-verb pair is a multi-word expression.So the frequency of the number of different verbs occurring with a particular object will be a good indicator for conjunct verbs. For example: a noun '*svIkAra*' (to accept) occurs only with two different verbs –'*kar*' (to do) and '*hE*' and noun '*kAnUna*' (law) occurs with five different types of verbs – '*bawA*' (to tell), '*kar*', '*baxala*' (to change), '*lA*' (to bring) and '*paDa*' (to study). Therefore, '*svIkAra*' is more likely to form a conjunct verb expression.

g. Verb Argument Indicator (f7): This feature computes the average number of post-position occurring before a unique noun/adjective-verb pair. The reason for exploring this feature is that if an expression has large number of post position occurring before it then its verb's argument structure is likely to be satisfied because each post-position is preceded by a noun/adjective which may potentially be the argument of the verb. Hence this noun/adjective-verb pair is more probable to form a conjunct verb.

5.3 Maximum Entropy

The features extracted above are used for binary classification of a noun/adjective-verb expression into conjunct verb and non-conjunct verb using the maximum entropy model [23]. Maximum entropy has already been widely used for a variety of natural language tasks, including language modelling [24], [25], text segmentation [26], part-of-speech tagging [28], and prepositional phrase attachment [27].
The maximum entropy model estimates probabilities based on the principle that the model is consistent with the constraint imposed maintaining uniformity otherwise. The constraints are derived from training process which expresses a relationship between the binary features and the outcome [29] [30]. Some of the features on which training is performed are distinct valued features (f1, f2) while others are real valued feature (f6, f7). These features are mapped to binary features. We used maximum entropy toolkit[2] to conduct our experiments.

[2] http://homepages.inf.ed.ac.uk/s0450736/maxent toolkit.html

6 Experiments and Results

The trained system on the corpus of 4500 sentences is tested on 1800 sentences for measuring its accuracy. The binary classification of noun/adjective-verb test expressions into conjunct verbs and non-conjunct verbs are done. We took different set of features for our experiments by trial and error method to come up with the best model. The best model gives us the highest accuracy of around 85.28%. For the baseline for our task we included Verb (f1) and Object (f2) as feature. Table2 gives the overview of useful features which helped in improving the accuracy.

Table2 shows that when the semantic feature (f3) was introduced, it lead to an improvement of around '0.75%', which proves the relevance of this feature. Inclusion of both Boolean features f4 and f5 showed a large jump in accuracy of about '3.15%'. Recall that f3 and f4 corresponds to D4 and D5 in section3. Addition of feature f6 improved our system by '0.54%' showing dominance of particular objects (as discuss during f6 definition) in conjunct verbs. We have not considered features which will show the steep decrease in accuracy, e.g. feature f7 on addition shows a decrease of '7.78%' with respect to the best accuracy reached so far, and moreover it is even less than the baseline also. We define features (f1+f2+f3+f3+f5) and (f1+f2+f3+f3+f5+f6) as System-1 and System-2 respectively.

Table2. Showing system accuracy with different feature set

Feature set	Accuracy
f1 + f2	(81.59)
f1+f2+f3	(82.34)
f1+f2+f3+f4+f5	(84.74)
f1+f2+f3+f4+f5+f6	(85.28)
f1+f2+f3+f4+f5+f7	(77.44)

7 Effect of Conjunct Verb on Parsing Accuracy

It had been observed that Dependency framework is the better way to analyze morphological rich free word-order languages (MoRFWO) (such as Czech,Turkish, Hindi, etc). Various data driven [32], [33], [34] and hybrid approaches [31] has been tried but still the current state-of-the-art parsing accuracy hasn't reached to a level which is comparable to English. Complex linguistics phenomenon is considered as a most vital factor for low accuracy of Hindi parsing apart from long distance dependencies, non-projective sentences and less corpus size. In past various morphological [34], semantic [35] and clause boundary [32] features have been tried to give language specific features in data driven parsing. All these features help in increasing the overall Hindi dependency parsing accuracy, but the gap between labeled and unlabeled accuracy is still large. Previous works [34], [43] have pointed that error due to complex predicates are significant in Hindi dependency parsing. Recall that in a conjunct verb it is the noun/adjective-verb complex that forms the predicate thereby controlling the argument structure. This means that unlike a

sentence with a normal verb the predicate information in a sentence with conjunct verb is distributed.

In this section, we investigate the effect of using conjunct verb specific features on parser accuracy. MST [36], [37] Parser was used to parse sentence, the MaxEnt based tool described in section 5.3 provides the feature. An improvement of 0.39% in label and 0.28% in label attachment accuracy is achieved.

7.1 Experiments and Results

We considered the MST+MaxEnt setting mentioned in [38] as Baseline for our experiments. All the parsing related experiments are performed on Dataset-2 as described in section 5.1. Using the output of System-1 and System-2 as described in Section-6, we added conjunct verb feature in each consecutive noun/adjective-verb pair in the dataset. Feature is added in the feature column of CONLL [42] format by giving an extra indicator like 'pof' (for conjunct verb) and 'npof' (for non-conjunct verb), which led to an increase in parsing accuracy using MST. Total number of noun/adjective-verb pairs is 3613 out of which 962 and 942 are marked as 'pof' and remaining as 'npof' by System-1 and System-2 respectively. The parsing result is shown in Table3.

Table3. Average LA (Labeled Attachment), UA (Unlabeled Attachment) and L (Label) accuracies on 12-fold cross validation

	LA (%)	UA (%)	L (%)
Baseline	68.77	85.68	71.90
System 1	**69.05**	**85.68**	**72.29**
System 2	68.52	85.04	71.93

Table4. 2nd and 3rd column represents the number of correctly identified 'pof' and 'npof' labels. Baseline-1 and Baseline-2 gives the number of labels that are correctly identified by the Baseline System group into 'pof' and 'npof' labels in comparison to System-1 and System-2 respectively. These stats are the summation of 12 testing set which are tested during 12-fold cross validation.

	'pof' labels	'npof' labels
Baseline-1	715	1628
System-1	715+**36**	1628+**21**
Baseline-2	713	1630
System-2	713+**42**	1630+**15**

7.2 Observations:

System-1 shows an increase of 0.39% in label and 0.28% in label attachment accuracy, this increase accounts to the 0.3%, 1.87%, 2.94% and 0.43% increase in

labels accuracy of 'k1', 'k2', 'pof', 'k7p'[3] respectively. These labels occur in the same environment as 'pof', hence the confusion. Both the System-1 and System-2 helps in reducing the 'npof' label (like 'k1', 'k2', 'k7p' etc.) confusion for those chunks which are given conjunct verb feature, by correctly identifying 21 and 15 more labels compare to baseline respectively as shown in Table4. Similarly, number of correctly identified conjunct verb labels increase by 36 and 42 in System-1 and System-2 respectively. This increase shows the positive effect of giving label specific feature to noun/adjective-verb pairs. Even if there is an increase in both systems output, the overall accuracy of System-2 is less compare to both System-1 and Baseline results. This decrease is because of indirect wrong learning leading to ambiguity between different labels.

8 Conclusions and Future Work

We have analyzed some of the diagnostics for manual identification of conjunct verb and there relevance in automatic identification. We successfully showed the importance of these diagnostics in statistical techniques by observing the significant increase in overall accuracy of identifying conjunct verbs and there positive effect on parsing accuracy. In future we will try to automate behavioral diagnostics (like D1 and D3) on the availability of large corpus. Although some diagnostics like Constituent Response Test (Wh-Questions) cannot be automated, they can give some theoretical grounding to conjunct verb identification and can complement the statistical tool.

We tried to include some context through feature like f7, but they didn't help. Since, additional context proves helpful in many tasks; we will have to explore this feature. The parsing accuracy showed improvement by incorporating the features given by our tool. Other NLP application tasks such as Machine Translation can also be tried.

9 References

1. Hook, P.E.: "The Hindi compound verb: What it is and what it does?" In: Readings in Hindi-Urdu linguistics, K.S. Singh (ed.), Delhi: National Publishing House (1974)
2. Mohanan, T.: "Wordhood and Lexicality", NLLT 13, 75-134 (1995)
3. Alsina, A.: Complex Predicates. CSLI Publications. Stanford (1995)
4. Mohanan, T.: Arguments in Hindi. Center for the Study of Language and Information, Leland Stanford Junior University, United States (1994)
5. Butt, M.: Conscious Choice And Some Light Verbs In Urdu. In: Verma, M.K.(eds.) Manohar (1993)
6. Butt, M.: Complex Predicates Compendium, Tromso, May (2005)
7. Agnihotri, R.K.: Hindi, An Essential Grammar. Routledge, London and New York, pp.121-126 (2007)
8. Bahl, K.C.: Studies in the Semantic Structure of Hindi. Motilal Banarasidass, Bihar (1974)

[3] k1, k2 can be roughly translated as agent and theme respectively. 'pof' is the relation between noun/adjective-verb in a conjunct verb, 'k7p' shows place relation. The dependency labels in the Treebank are syntactico-semantic in nature. For more details refer [10].

9. Bharati, A., Chaitanya, V., Sangal, R.: Natural Language Processing: A Paninian Perspective, Prentice Hall of India, New Delhi, pp. 65-106 (1995)
10. Begum, R., Husain, S., Dhwaj, A., Sharma, D.M., Bai, L., Sangal, R.: Dependency Annotation Scheme for Indian Languages. In: The Third International Joint Conference on Natural Language Processing (IJCNLP), Hyderabad, India (2008)
11. Bhatt, R., Narasimhan, B., Palmer, M., Rambow, O., Sharma, D.M., Xia, F.: A Multi Representational and MultiLayered Treebank for Hindi/Urdu. In: The Third Linguistic Annotation Workshop (The LAW III) in conjunction with ACL/IJCNLP. Singapore (2009)
12. Kellogg, Rev, S.H.: A Grammar of the Hindi Language, Routledge and Kegan Paul Ltd. (London 1875)
13. Bailey, T.G.: Teach Yourself Urdu, J.R. Firth and A.H. Harley (editors) (London, 1956)
14. Guru, K.P.: Hindi Vyakarana, Kasi Nagari Pracarini Sabha, (Kasi, 1922)
15. Sharma, A.: A Basic Grammar of Modern Hindi. Government of India, Ministry of Education and Scientific Research (Delhi, 1958)
16. Bahri, H.: Hindi Semantics, The Bharati Press (Allahabad, 1959)
17. Bahl, K.C.: A Reference Grammar of Hindi, (The University of Chicago, 1967)
18. Bhattacharyya, P., Chakrabarti, D., Sarma, V.M.: Complex predicates in Indian languages and wordnets. In: Language Resources and Evaluation 40(3-4): 331-355 (2006)
19. Mukerjee, A., Soni, A., Raina, A.M.: Detecting Complex Predicates in using POS Projection across parallel Corpora Aligned Hindi sentence. In: Workshop on Multiword Expressions: Sydney, pp 11–18 (2006)
20. Sinha, R.M.K.: Mining Complex Predicates In Hindi Using A Parallel Hindi-English Corpus.In: ACL International Joint Conference in Natural Language Processing, pp 40 (2009)
21. Venkatapathy, S., Joshi, A.: .Relative Compositionality of Noun+Verb Multi-word Expressions in Hindi. In: ICON (2005)
22. Chakrabarti, D., Mandalia, H., Priya, R., Sarma, V., Bhattacharyya, P.: Hindi Compound Verbs and their Automatic Extraction. In: International Conference on Computational Linguistics, pp. 27-30 (2008)
23. Ratnaparkhi, A.: Maximum Entropy Models for Natural Language Ambiguity Resolution. Ph.D. thesis (1998)
24. Rosenfeld, R.: Adaptive Statistical Language Modeling: A Maximum Entropy Approach. Ph.D. thesis, Carnegie Mellon University (1994)
25. Chen, S.F., Rosenfeld, R.: A Gaussian prior for smoothing maximum entropy models. Technical Report CMUCS-99-108, Carnegie Mellon University (1999)
26. Beeferman, D., Berger, A., Lafferty, J.: Statistical models for text segmentation. Machine Learning, 34(1-3):177-210 (1999)
27. Ratnaparkhi, A., Reynar, J., Roukos, S.: A maximum entropy model for prepositional phrase attachment. In: ARPA Human Language Technology Workshop, pages 250-255 (1994)
28. Ratnaparkhi, A.: A maximum entropy model for part-of-speech tagging. In: Empirical Methods in Natural Language Conference (1996)
29. Nigam, K., Lafferty, J., McCallum, A.: Using Maximum Entropy for Text Classification. In: IJCAI-99 Workshop on Machine Learning for Information Filtering, pp. 61-67. (1999)
30. Berger, A., Pietra, D.: A Maximum Entropy Approach to Natural Language Processing. Computational Linguistics, Vol., 22. p. 39-71 (1996)
31. Bharati, A., Husain, S., Vijay, M., Deepak, K., Sharma, D.M., Sangal, R.: Constraint Based Hybrid Approach to Parsing Indian Languages. In: The 23rd Pacific Asia Conference on Language, Information and Computation (PACLIC 23). Hong Kong (2009)
32. Gadde, P., Jindal, K., Husain, S., Sharma, D.M., Sangal, R.: Improving Data Driven Dependency Parsing using Clausal Information. In: NAACL-HLT 2010, Los Angeles, CA (2010)

33. Nivre, J., Hall, J., Kubler, S., McDonald, R., Nilsson, J., Riedel, S., Yuret, D.: The CoNLL 2007 Shared Task on Dependency Parsing. In: CoNLL Shared Task Session of EMNLP-CoNLL (2007)
34. Ambati, B.R., Husain, S., Nivre, J., Sangal, R.: On the Role of Morphosyntactic Features in Hindi Dependency Parsing. In: NAACL-HLT 2010 workshop on Statistical Parsing of Morphologically Rich Languages (SPMRL 2010), Los Angeles, CA (2010)
35. Ambati, B.R., Gade, P., GSK, C.: Effect of Minimal Semantics on Dependency Parsing. In: RANLP 2009 Student Research Workshop (2009)
36. McDonald, R., Pereira, F., Ribarov, K., Hajic, J.: Non-projective dependency parsing using spanning tree algorithms. In: HLT/EMNLP, pp. 523–530 (2005)
37. McDonald, R., Crammer, K., Pereira, F.: On-line large-margin training of dependency parsers. In: ACL 2005. pp. 91–98 (2005)
38. Ambati, B.R., Gadde, P., Jindal, K.: Experiments in Indian Language Dependency Parsing. In: ICON09 NLP Tools Contest: Indian Language Dependency Parsing, pp 32-37 (2009)
39. Husain, S., Mannem, P., Ambati, B., Gadde, P.: The ICON-2010 tools contest on Indian language dependency parsing. In: ICON-2010 tools contest on Indian language dependency parsing. Kharagpur, India (2010)
40. Husain, S.: Dependency Parsers for Indian Languages. In: ICON09 NLP Tools Contest: Indian Language Dependency Parsing. Hyderabad, India (2009)
41. Bharati, A., Husain, S., Ambati, B., Jain, S., Sharma, D., Sangal, R.: Two Semantic features make all the difference in Parsing accuracy. In: International Conference on Natural Language Processing (2008)
42. Buchholz, S., Marsi, E.: CoNLL-X shared task on multilingual dependency parsing. In: Tenth Conference on Computational Natural Language Learning (CoNLL), pp.149–164. (2006)
43. Ambati, B.R., Husain, S., Jain, S., Sharma, D.M., Sangal, R.: Two methods to incorporate local morphosyntactic features in Hindi dependency parsing. In: NAACL-HLT 2010 workshop on Statistical Parsing of Morphologically Rich Languages (SPMRL 2010) Los Angeles, CA (2010)