International Journal of Computational Linguistics and Applications
vol. 7, no. 2, 2016, pp. 45–63
Received 08/02/2016, accepted 09/03/2016, final 19/09/2016
ISSN 0976-0962, http://ijcla.bahripublications.com

Pronominal Reference Type Identification and Event Anaphora Resolution for Hindi

Vandan Mujadia, Palash Gupta,
and Dipti Misra Sharma

IIIT-Hyderabad, India

Abstract

In this paper, we present hybrid approaches for pronominal reference type (abstract or concrete) identification and event anaphora resolution for Hindi. Pronominal reference type identification is one of the important parts for any anaphora resolution system as it helps anaphora resolver in optimal feature selection based on pronominal reference types. We use language specific rules and features in set of classifiers (ensemble learning) for pronominal type identification. We discuss event referring anaphors (pronouns) and their resolution using Paninian dependency grammar, language syntax, proximity of events, etc. We achieved around 90% accuracy in the pronominal reference type identification and around 71% F-score in the event anaphora resolution on Hindi dependency tree-bank corpus.

1 Introduction

Anaphora resolution is a complex and most ambiguous task which involves resolution of variety of phenomena like references to noun sequence, verb sequence, ellipsis and bridging,

etc. Any anaphora resolution system can be benefited if somehow it can gather information like whether given anaphor refers to an event or concrete entity. One can divide anaphors based on their reference types, either a concrete reference or an event / abstract reference. For a concrete reference, an anaphor refers to a concrete entity, it can be a noun phrase (which includes named entity, noun, etc.). While for an event / abstract reference, an anaphor refers to an event (verb phrase/clause) or it can have null reference. These reference types can help resolution system to choose different resolution strategies (in features and rules) based on different anaphoric expressions (in the same language). Therefore it is efficient to consider separate resolution modules for different types of pronominal references. In our approach, we separate out pronouns based on these reference types through pronominal reference type identifier, and then proceed for event anaphora resolution.

2 Related work

The majority of anaphora resolution studies have been carried out for English. These studies can be categorized into syntax-based methods, knowledge poor methods, machine learning methods or hybrid approaches [1]. Hobb‘s naive approach [2], Lappin-Leass [3] and Brennan’s [4] Centering Theory are well-known syntax and rule based approaches for anaphora resolution. SHRDLU [5], CogNIAC [6] and Kennedy [7] devise syntactic and relevance based heuristics for personal pronoun resolution. Mitkov ([8] and [9]) demonstrated a robust anaphora resolution method based on a set of boosting and impeding indicators on each antecedent of a pronoun. Connolly [10] presented supervised machine learning approach while [11] and [12] presented genetic algorithms in unsupervised methods for anaphora resolution. For English event anaphora resolution, [3] used slot grammar and [13] used hierarchical distance-dependent Bayesian model. One does not find much work on

Table 1: Pronominal type distribution

Anaphor (English tran)	#Concrete reference	#Event reference	#Null reference
अपना$_{mine}$	1551	0	17
अभी$_{now}$	0	0	267
आप$_{you}$	238	0	11
इसलिए$_{because-of}$	0	192	0
ऐसा$_{alike}$	36	277	107
खुद$_{own}$	124	0	0
जब$_{when}$	0	298	38
जहाँ$_{where}$	165	0	18
जो$_{what/that}$	1006	0	163
तब$_{then}$	0	196	0
तुम$_{you}$	263	0	8
मै$_{i}$	349	0	0
यह$_{this}$	3202	4159	514
यहाँ$_{here}$	507	0	47
यही$_{here-only}$	35	0	2
वह$_{he/she/it}$	5440	641	44
वहाँ$_{there}$	236	0	0
वहीं$_{there-only}$	8	0	0
हम$_{we/honorofic-i}$	632	0	0
इतना$_{thismuch}$	26	43	6
Total	13792	6336	1226

the pronominal reference type identification in literature. For Hindi, [14] described a hybrid approach for concrete anaphora resolution while (Mehla et al.)[1] and [15] listed various resolution approaches for event anaphora resolution. The corpora details are discussed in section 3. Section 4 describes approaches for pronominal reference type identification and event anaphora resolution with examples. We discuss the results and conclude the paper in sections 5 and 6 respectively.

[1] *ijiset.com/vol2/v2s1/IJISET_V2_I1_86.pdf*

3 Data

We evaluated our methods on the Hindi Dependency Treebank [16]. It is a linguistically rich news and story corpora having various level of annotations. It contains annotations from part-of-speech, morph, chunk/phrase to dependency based on the Computational Paninian Grammar (CPG) framework, as explained in [17] and [18]. We mapped [19]'s semantic features to the latest version of the HDTB. The pronominal reference type and event anaphora annotation were carried out on these corpora based on [20] annotation scheme. HDTB has around 1100 documents containing around 15478 sentences and around 21354 anaphors. Table 1 describes the pronoun distribution according to their reference types. Table 1 has four columns, column 1 describes the pronominal lemma (with respective English translation), and second, third and fourth columns respectively describe the distribution of pronouns according to reference types (concrete, event or null). We use 'null' as a reference type when there is no textual reference present in the given discourse.

4 Approach

In this section we describe our approaches for pronominal reference type identification and event anaphora resolution specific to Hindi. We use language specific rules and pronominal placement in sentential discourse as main features for pronominal type identification. For event anaphora resolution, we use Paninian dependency grammar, proximity of events as main features. For both of these subtasks, we do not intend to find anaphors from the text but systems trigger when POS (part-of-speech) label of given lexical is 'PRP' (pronoun tag).

4.1 *Pronominal reference type identification*

In Hindi, some pronominal forms give obvious clues to whether they refer to an abstract entity or a concrete entity, while some pronominal forms are more ambiguous. This statement can be verified from the table-1. Our algorithm identifies the pronominal reference type by combining the results from rule based and classifier based (ensemble) identifiers. Later, the identification function (Figure-1) is used to weight the outputs of rule and ensemble identifiers. The identification function, then, assigns the resultant pronominal type category for a pronoun. The rules of **Rule based pronominal type identifier** were arrived at by checking the frequencies of the anaphors reference categories (Hindi dependency treebank: table-1) and assigns the suitable pronominal type. These rules provide probabilities for pronouns in classes (concrete vs abstract). In **ensemble pronominal type identification**, we formed a cluster of classifiers. This cluster includes classifiers like decision-tree, linear model and SVM with various kernel functions. We use these classifiers from scikit-learn [21] python library. We use following best performing features in the above mentioned classifiers (60%-10%-30% data for training, tuning, testing respectively).

– **1.** Pronoun, **2.** Pronoun‘s lexical distance from last verb, **3.** pronoun‘s lexical distance from upcoming verb, **4.** Pronoun‘s lexical position in sentence, **5.** Pronoun root, **6.** Pronoun‘s sentence position in discourse, **7.** Pronoun‘s chunk position in sentence, **8.** Pronoun‘s gender-number-person, **9.** Pronoun‘s next word and its POS category **10.** Voice and sentence type of next and previous sentence, **11.** Pronoun‘s lexical distance from last occurred named entity, **12.** Lexical item count of pronouns chunk.

$$T = W_0 * T_0 + \sum W_i * T_i$$

Figure 1: Identification function

We use Identification function (T) [figure-1] to get the final reference type category, where W_0 and W_i represent the weights for respective classifiers (T_0, T_i). Here, W_0 is the weight for rule based identifier (T_0) and due to obvious evidence (table-1), we give it a higher weight than all other classifiers (ensemble based). Weights of other classifiers were derived from the development data. Each classifier (T_0 and T_i) gives +ve 1 if the predicted pronominal type is concrete, else -ve 1. Identification function (T) assigns weight $W_{0,i}$ to each classifier $T_{0,i}$ and gives numeric value, where +ve value resides for concrete pronominal type and -ve value for abstract pronominal type. The accuracies of each classifiers and pronoun wise accuracies are discussed in the result section.

4.2 *Event Anaphora Resolution*

For Hindi, event anaphors can be categorized into several classes based on their lexical forms. They are यह (this) with its forms (ये (this/these), यही (this only), इसे (it), इस (this), इसका (its), इसी (this), इसकी (its), इसके (its), इसमें (in this), इससे (from this)), वह (that), इसलिए(therefore, because-of), इतना (this much), ऐसा (like-this) and जब – तब (when-then). We defined event anaphora resolution as an identification problem where resolver tries to find the finite or non-finite verb as a referent for an event anaphor. This resolver works on SSF documents and triggers when it encounters 'PRP' as part-of-speech tag with pronominal reference type category as Event ('E').

We built three different systems for event anaphora resolution. First system works on the predefined heuristics where the system demonstrates the working of anaphora resolution

(with and without dependency based rules, rl-Wdep and rl-WOdep). In second system, we use set of classifiers to learn various features for event anaphora resolution. Third one is a hybrid system (final system) where, the resolver finds the reference from both rule based and learning based systems and then an ensemble voter function gives the final reference for the given anaphor. Each system performance is showed in the result section.

Rule based event anaphora resolution (with dependency based rules)

Resolving* यह*(yaha[this]): It is obvious from Table 1 that the frequency of pronominal type यह ('this') is higher than other anaphors. They are most ambiguous pronouns for pronominal reference type identification as well as for resolution. Mostly, यह ('this') refers to a proposition of the same sentence which is expressed or narrated. We identified two types of resolution patterns for event referring यह ('this') based on dependency structure and relations. **In the first type**, a complementary clause acts as a child of an event anaphor यह ('this') with the dependency relation 'samanaadhikaran' (rs-label - same as), all labels which have 's' as postfix)[17]. The relation 'samanaadhikaran' shows the reason for occurrence of an event, thus, the sub-tree representing this complementary clause acts as a direct referent for the event anaphor.

(1) विदेश मंत्रालय दूसरे देशों में स्थित भारतीय दूतावासों को [यह]$_{\text{i}}$ निर्देश देने जा रहा है कि$_{rs}$ [वे हादसे के बारे में वहां के लोगों को सही जानकारी दें]$_{\text{i}}$ ।

The Foreign Ministry is going to$_{rs}$ direct the Indian embassies in other countries that$_i$ (they give the correct information to the people about the tragedy)$_i$.

In example - (1), pronoun यह(yaha) has a child कि(ki) with 'rs' relation and कि(ki) has a complementary clause वे हादसे के बारे में वहां के लोगों को सही जानकारी दें (they give the correct information to the people about the tragedy) as a child. According to our rule, we mark this complementary clause as a referent of event referring anaphor यह(yaha) [represented by index - i].

In the second type, यह ('this') may have reference of/in intra-sentential clause which is either expressed or narrated. In the intra-sentential clause, if parental subtree/node of यह ('this') has 'rh' (label describes the 'reason') dependency relation then this 'rh' relation containing subtree (finite verb) acts as a direct referent for the given event anaphor यह ('this').

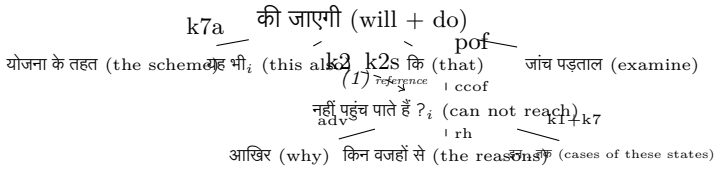

Figure 2

(2) इस योजना के तहत [यह]$_{\text{i}}$ भी जांच पड़ताल की जाएगी कि आखिर [किन वजहों से]$_{rh}$ [इन राज्यों के मामले आयोग तक नहीं पहुंच पाते हैं ?]$_{\text{i}}$

The scheme will also examine the reasons (this)$_i$ (why cases of these states can not reach to the Commission ?)$_i$

In (example 2 - Figure 2), as pointed earlier, the parent of pronoun यह(this) has a child कि (ki) with 'k2s' relation. Also कि (ki) has a clause आखिर किन वजहों से ... हैं ? (they give the right information to the people about the tragedy) as a

child with 'rh' (label, describes 'reason') dependency relation. Therefore, we mark this clause as a referent of event anaphor यह(this) [represented by index - i].

Resolving इसलिए ***(isaliye/because-of, therefore]):*** In Hindi, इसलिए (isaliye) is frequently used as a connective or an event referring anaphor. Though, it represents causal or sentential relationship, there are multiple possible structures for इसलिए (isaliye) depending on the other connectives with which it appears in sentential discourse. For event reference, इसलिए (isaliye) can have reference in either direction. The direction usually depends on position of reporting verbs like (कहा(tell/say), बोला(spoke), सुनाया(pronounced), बताया(told), etc.) and/or various demonstrative conjuncts like (i.e. कि (because/that), चूकि (since), क्यूंकि (because) etc). From the examples of Hindi dependency treebank, we can argue that this anaphor has occurrences for representing a 'reason' between clauses or for showing a 'goal' for occurred event.

(3) माकपा नेता ने कहा कि [इन पदार्थो पर कस्टम ड्यूटी घटाई गई है $_{rs}$]$_{\text{i}}$ [इसलिए$_{rh}$]$_{\text{i}}$ इनकी कीमतों पर कोई फ़र्क नहीं पड़ेगा ।

CPI-M leader said that (the customs duty on these items has been reduced)$_i$ therefore$_i$ their prices would not effected.

In Hindi dependency treebank 'rh' dependency relation represents the reason for an event occurrence, therefore if इसलिए (isaliye) has a 'rh' with it's' sibling subtree than this subtree can be a potential referent. In example-3, इन पदार्थो पर कस्टम ड्यूटी घटाई गई है (the customs duty on these items has been reduced) is a sibling subtree of an anaphor having 'rs' relation, therefore it can be the referent. While for inter sentential relations, event anaphor इसलिए (isaliye) mostly refers to last occurred event before the anaphor.

Resolving ऐसा ***(Esa-alike):*** Comparatively ऐसा(alike) has less occurrence than other pronouns in Hindi dependency treebank (Table-1). It is used for describing the reason for event, whether it has already happened or is currently happening. Hindi syntax helps a lot to find out correct reference for the event referring pronoun. Generally pronoun ऐसा(alike) uses conjunct like 'कि(ki)' to describe the reason for the event. The referent of ऐसा(alike) can be found in the subtree or the clause other than the pronominal clause which is located after conjunct (कि(ki)) in the same sentence.

(4) आडवाणी ने कहा कि अभी तक अपराध और अपराधियों को दंडित करने की परंपरा रही है लेकिन पहली बार [ऐसा]$_{\text{i}}$ हो रहा है कि$_{ki}$ [उन्हें पुरस्कृत किया जा रहा है]$_{\text{i}}$ ।

Advani said that so far, It has been a tradition that the crime and offenders has to be punished but this$_i$ is the first time that (they are being rewarded)$_i$.

In example-(4), event referring pronoun ऐसा(alike) is connected with a clause by conjunct कि(ki). Therefore, the child subtree उन्हें पुरस्कृत किया जा रहा है (they are being rewarded) of the conjunct is the referent of an anaphor.

Resolving जब – तब/तभी/तो ***(when - then/then/so):*** Hindi generally prefers the so-called relative correlative construction, wherein a relative pronoun is used along with another corresponding pronoun to correlate two clauses. जब – तब/तभी/तो (when - then/then/so) pronoun pairs are relative-correlative pronoun pairs. Even though they refer to time, we are treating this type of pronouns as an event referring pronoun because they refer to time that resides in event. Therefore we decided to mark this time indicating event as a referent for given event pronoun. Hindi syntax and constructions of these anaphor indicate the most prominent way to resolve these anaphors. Following three examples (5), (6), (7)

show the possible constructions of जब – तब/तभी/तो (when - then/then/so).

(5) [जब]$_i$ [मैने खाना खाया]$_i$ [तब]$_i$ उसने पानी पिया ।

When$_i$ (I had food)$_i$, then$_i$ he drank water.

(6) [मैने खाना खाया]$_i$ [तब]$_i$ उसने पानी पिया ।

(I had food)$_i$, then$_i$ he drank water.

(7) [जब]$_i$ [1911 में क्वीन मैरी यहां आई थीं]$_i$ [तब]$_i$ ब्रिटिश शासकों ने यहां के पानी के हौज के ऊपर एक छत बनवाई थी ।

When$_i$ (Queen Mary came here in 1911)$_i$ then$_i$ the British rulers had built a roof over the water of the pool.

If a sentence contains relative-correlative pronouns pair then their temporal reference must be resided into the verb present between this pronoun pair. These verbs can be any, other then reporting/narrating verbs like (कहा(tell/say), बोला(spoke), सुनाया(pronounced), बताया(told), etc). Therefore, in example (5), (7), both relative correlative pronouns (जब (when) and तब (then)) are stated. The reference for both pronouns is मैने खाना खाया (I had food) in example (5). Same for example (7), where the reference for (जब (when) and तब (then)) is 1911 में क्वीन मैरी यहां आई थीं (Queen Mary came here in 1911). One can observe that both relative and correlative pronouns refer to the same event with only a directional difference. If only one of relative or correlative pronoun is present in the sentence, then the reference is the verb/event which precedes the anaphor other than reporting/narrating verbs. In example (6), given temporal pronoun तब (then) is referring to time, which resides in the event मैने खाना खाया (I had food).

```
Data: SSF text, Result: reference of given anaphor
while not at end of SSF document do
    read current pronoun node;
    if current pronoun has event reference then
        if Reporting verb [कहा, बोला, सुनाया, बताया, etc.] in current
        sentence then
            if Conjunct [कि',etc] in current sentence then
                if Position of reporting verb (dist) < position
                of conjunct(dist) > position of event
                pronoun(dist) then
                    select last event (after conjunct) as a
                    referent.;
                else
                    if Position of event pronoun is within first
                    few words then
                        select last event (after conjunct) as a
                        referent.;
                    else
                        select last verb/event before event
                        pronoun as a referent.;
                    end
                end
            else
                do nothing;
            end
        else
            if conjunct is in sentence and position of conjunct
            < position of event pronoun then
                select last event (after conjunct) as a referent.;
            else
                select last verb/event before event pronoun as a
                referent.;
            end
        end
    else
        Select the last clause of a previous sentence as referent;
    end
end
```

Algorithm 1: Event anaphora resolution

Rule based event anaphora resolution algorithm (without dependency based rules): The kind of pronouns like इतना (thismuch/not only this) can be resolved by just referring to the last occurred event or the nearest event based on the language syntax. Therefore, we came up with rule based event anaphora resolution algorithm (algorithm -1) which works for all categories of event anaphors without using dependency relations or structures. The referent of the event referring anaphor can be found in either of the directions based on the nature of anaphor.

(8) आधिकारिक सूत्रों ने बताया कि [जम्मू व कश्मीर में गुलमर्ग स्थित स्कींग रिजार्ट में बर्फ़ की पांच इंच मोटी परत बिछ गई ।]$_i$ [इतना]$_i$ ही नहीं पूरे राज्य , खासकर उत्तरी कश्मीर में हुई जबरदस्त बर्फ़बारी ।

(Official sources said the skiing resort of Gulmarg in Baramulla in Jammu and Kashmir, five inches of snow was covered with a thick layer)$_i$.Not only this$_i$ there were contumelious massive snowfall in whole Kashmir.

From the example (8), इतना (thismuch) is referring to quantitative event जम्मू व कश्मीर में गुलमर्ग स्थित स्कींग रिजार्ट में बर्फ़ की पांच इंच मोटी परत बिछ गई । (Official sources said the skiing resort of Gulmarg in Baramulla in Jammu and Kashmir, five inches of snow was covered with a thick layer), which is nearest event to the anaphor इतना (thismuch). The type of resolution also depends on the relative position of event anaphor, reporting/narrating verb and conjunct in the sentence. The algorithm is as described in (Algorithm-1) and which is self explanatory.

FEATURE LEARNING In discourse, event anaphors refer to abstract objects as well act as discourse connectives. From statistics, one can argue that event anaphors mostly refer to

last occurred event, they are either inter sentential or non-structural. Also humans have tendency that they hardly re-utter event through anaphor beyond 2 sentences due to their cognitive load carrying capability. This makes nearest verb as a potential candidate for event referring anaphors. We devised features (listed below) based on this assumption. As described in [22], pairwise features for decision tree learning help in resolving nominal references. So we also took pairwise features for event anaphora resolution i.e. a pair of event referring pronoun (its features) and verb with their features (listed below) and trained them with decision tree classifier [21] (60%–10%–30% data for training, tuning, testing respectively):

1. Pronoun root, **2.** Pronoun's lexical distance from last verb, **3.** pronoun's lexical distance from upcoming verb, **4.** Pronoun's GNP, **5.** Pronoun's dependency relation, **6.** Presence of sentential connectors, **7.** Distance and direction from pronoun to nearest verb , **8.** Nearest verb, **9.** Nearest verb GNP,voice and type, **10.** Dependency relations across sentence, **11.** Dependency tree paths, **12.** Presence of conjunct in same sentence, **13.** Presence of reporting verb in same sentence, **14.** Pronoun direction from reporting verb, **15.** Pronoun distance from conjunct, **16.** Pronoun distance from reporting verb, **17.** Pronoun direction from conjunct, **18.** conjunct distance from reporting verb of a same sentence.

The ensemble voter function was:

$$T(EA) = W_0 * T_0 + W_1 * T_1 + \sum W_i * T_i.$$

Hybrid approach for event resolution In hybrid event anaphora resolution, we combined our all above described systems, and used the ensemble voter function for final resolution output. The ensemble voter function for Hybrid event

Table 2: Classifier-wise accuracy [pronominal type identification]

Classifier Name	#Accuracy
$linear\_model\_BayesianRidge$	0.94
$linear\_model.Ridge(alpha = .5)$	0.93
$svm\_SVC(kernel =' linear')$	0.92
$svm\_LinearSVC()$	0.91
$svm.LinearSVC()$	0.91
$tree.DecisionTreeClassifier$	0.91
$tree.DecisionTreeRegressor$	0.90
$svm.SVR$	0.90

anaphora resolution is stated above as T(EA) [figure-3]. The functions $T_{0,1,i}$ returns references (verb/event) for the given event anaphor. Function W stand for predefined weight assigned to respective type of resolution module. These weights are learnt from various experiments. T_0 stands for rule based resolution module with dependency features (Wdep) and W_0 for its corresponding weight. While T_1 stands for rule based resolution module (Algorithm 1 - WOdep) and its corresponding weight as W_1. T_i and W_i refer to learning based classifiers and their weights respectively which includes classifiers like decision-tree, linear model and svm with various kernel functions. The output references and weights of all the modules act as a factor for preferential voting. So, for given event anaphor, T(EA) function gives list of references with their corresponding weights. From this list **max function** outputs the reference by picking up the highest weighted reference from the list.

5 Results

Table (2) shows the classifier-wise average accuracies for pronominal reference type identification on Hindi dependency Treebank data. Pronoun wise average accuracies for pronominal type identification are listed in table (3). The accuracies in Table (2) are of learning based module of different classifiers.

Table 3: Accuracy pronominal type identification

Pronoun/ Eng-form	Accuracy	Pronoun/ Eng-form	Accuracy
अपना$_{mine}$	0.95	तुम$_{you}$	0.66
अभी$_{now}$	1.00	मै$_{i}$	1.00
आप$_{you}$	0.88	यह$_{this}$	0.74
इतना$_{this-much}$	0.94	यहाँ$_{here}$	0.86
इसलिए$_{because-of}$	0.94	यही$_{here-only}$	0.86
ऐसा$_{alike}$	0.94	वह$_{he/she/it}$	0.93
खुद$_{owns}$	1.00	वहाँ$_{there}$	0.86
जब$_{when}$	0.79	वहीं$_{there-only}$	0.88
जहाँ$_{where}$	0.86	हम$_{we/hor-i}$	1.00
जो$_{what/that}$	0.79		

Table 4: Event anaphora resolution F-Score

Pronouns	FS ([15])	FS (rl-Wdep)	FS (rl-WOdep)	FS (hybrid)
इतना$_{this-much}$	-	.50	.65	.65
इसलिए$_{because-of}$	.53	.71	.70	.82
जब$_{when}$	-	-	.70	.70
तब$_{then}$	-	-	.72	.72
यह$_{this}$	.63	.62	.56	.68
वह$_{that}$	-	.50	.48	.56
ऐसा$_{alike}$	-	.70	.78	.87
Total	.59	.61	.65	.71

One can observe that linear classifiers give best overall accuracy but SVM and decisionTree based classifiers outperform other classifiers for some pronominal types like यह (this) and वह (that). Table - 3 shows the final combined (rule based + learning based (figure -1)) accuracies for pronominal type identification.

Table (4) contains the pronoun typewise F-scores, in which first column shows the results presented by [15] for limited number of pronouns. Second, third and fourth columns show the F-scores for rule based with dependency features, rule

based without dependency features and combine hybrid system respectively. From results one can observe that our system outperforms the existing event anaphora resolution system[15]. As stated in [14], Paninian based dependency labels play major role for concrete anaphora resolution in Hindi, but for event anaphora resolution, process mainly depends on language syntax, proximity of verbs, conjuncts and anaphor (from table (5)).

6 Future work

In future, we would try to verify our claims by applying these systems upon other Indian Languages with minor modifications. Also we can try to use thes pronominal reference type identifier and anaphora resolution in applications like machine translation, especially in ILMT [23] systems for discourse centric translation, question answering, summarization, dialogue processing, etc.

References

1. Barbu, C.: Bilingual pronoun resolution: experiments in English and French. PhD thesis, University of Wolverhampton (2003)
2. Hobbs, J.: Resolving pronoun references. In: Readings in natural language processing, Morgan Kaufmann Publishers Inc. (1986)
3. Lappin, S., McCord, M.: Anaphora resolution in slot grammar. Computational Linguistics **16**(4) (1990) 197–212
4. Brennan, S.E., Friedman, M.W., Pollard, C.J.: A centering approach to pronouns. In: Proceedings of 25th ACL. (1987)
5. Winograd, T.: Understanding natural language. Cognitive psychology **3**(1) (1972) 1–191
6. Baldwin, B.: Cogniac. In: Proceedings of a Workshop on Operational Factors in Practical, ACL (1997) 38–45
7. Kennedy, C., Boguraev, B.: Anaphora for everyone: pronominal anaphora resoluation without a parser. In: Proceedings of the 16th CL-Volume 1, ACL (1996) 113–118
8. Mitkov, R.: Robust pronoun resolution with limited knowledge. In: Proceedings of the 17th international conference on Computational linguistics-Volume 2, Association for Computational Linguistics (1998) 869–875

9. Mitkov, R., Evans, R., Orasan, C.: A new, fully automatic version of mitkov's knowledge-poor pronoun resolution method. In: Computational Linguistics and Intelligent Text Processing. Springer (2002) 168–186
10. Connolly, D., Burger, J.D., Day, D.S.: A machine learning approach to anaphoric reference. In: New Methods in Language Processing. (1997) 133–144
11. Byron, D.K., Allen, J.F.: Applying genetic algorithms to pronoun resolution. In: AAAI/IAAI. (1999) 957
12. OrĂsan, C., Evans, R., Mitkov, R.: Enhancing preference-based anaphora resolution with genetic algorithms. In: Natural Language Processing—NLP 2000. Springer (2000) 185–195
13. Yang, B., Cardie, C., Frazier, P.: A hierarchical distance-dependent bayesian model for event coreference resolution. arXiv preprint arXiv:1504.05929 (2015)
14. Dakwale, P., Mujadia, V., Sharma, D.: A hybrid approach for anaphora resolution in hindi. In: International Joint Conference on Natural Language Processing, Nagoya, Japan. (2013) 977–981
15. Dakwale, P.: Anaphora Resolution in Hindi. PhD thesis, International Institute of Information Technology Hyderabad (2014)
16. Bhatt, R., Narasimhan, B., Palmer, M., Rambow, O., Sharma, D.M., Xia, F.: A multi-representational and multi-layered treebank for hindi/urdu. In: Proceedings of the Third Linguistic Annotation Workshop, Association for Computational Linguistics (2009) 186–189
17. Begum, R., Husain, S., Dhwaj, A., Sharma, D.M., Bai, L., Sangal, R.: Dependency annotation scheme for indian languages. In: Proceedings of IJCNLP. (2008)
18. Bharati, A., Chaitanya, V., Sangal, R., Ramakrishnamacharyulu, K.: Natural language processing: a Paninian perspective. Prentice-Hall of India (1995)
19. Jena, I., Bhat, R.A., Jain, S.: Animacy annotation in Hindi treebank. In: Proceedings of the LAWVII, ACL (2013)
20. Dakwale, P., Sharma, H., Sharma, D.M.: Anaphora annotation in hindi dependency treebank. In: Proceedings of the 26th Pacific Asia Conference on Language, Information, and Computation. (2012) 391–400
21. Pedregosa, F.e.a.: Scikit-learn: Machine learning in Python. Journal of Machine Learning Research **12** (2011) 2825–2830
22. Soon, W.M., Ng, H.T., Lim, D.C.Y.: A machine learning approach to coreference resolution of noun phrases. CL **27** (2001)
23. Consortium, I., et al.: Ilmt system. IIT Hyderabad (2007)

Vandan Mujadia
Language Technologies Research Centre,
IIIT-Hyderabad, India
E-mail: <vmujadia@gmail.com>

Palash Gupta
Language Technologies Research Centre,
IIIT-Hyderabad, India
E-mail: <palash.gupta03@gmail.com>

Dipti Misra Sharma
Language Technologies Research Centre,
IIIT-Hyderabad, India
E-mail: <diptims@gmail.com>