Chapter 7

Some Hindi Verbs

Verbs are the most important grammatical category in a language. Actions and states are denoted with the help of the verbs. The arguments of the verb specify various participants required by the verb. Verbs are classified both on semantic and syntactic basis. Semantically, verbs are classified into action verbs, state verbs, and process verbs. Syntactically they are classified into intransitives, transitives, and ditransitives. The morphological, semantic, and syntactic properties of verbs play an important role in deeper level analysis such as parsing. Verbs play a major role in interpreting the sentence meaning therefore, the study of verb argument structure and their syntactic behavior provides the necessary knowledge base for intelligent NLP applications.

As mentioned above, there are three types of verbs based on syntax, i.e., intransitives, transitives, and ditransitives. Most of the intransitive or transitive verbs slightly modify their forms to indicate causation of the actions denoted by them. These verbs are called causative verbs derived from intransitive or transitive verbs. The intransitive or transitive verbs and causative verbs derived from it, have similarity in the form of their stems and also have semantic relationship. Other than simple one-word verbs, there are complex verbs which are made up of more than one word. These complex verbs are divided into two categories: compound and conjunct verbs. Compound verbs consist of two verbs and conjunct verbs are made up of a nominal followed by a verb (Khachru, 2006). In this chapter, my focus is on causative verbs and conjunct verbs because causatives are morphologically realized in Hindi which pose interesting theoretical issues for their argument structure representation and conjunct verbs are highly productive in Hindi. Section 7.1 describes causative verbs in detail and how they are represented in dependency structure for Treebank. Section 7.2 describes conjunct verbs in detail where I have discussed the diagnostics to identify conjunct verbs and which of these diagnostics can be used to automate the identification process.

7.1 Causative Verbs

Causative verbs are differently realized in different languages. These verbs have been an interesting area of study in linguistics because of their morphological, syntactic and semantic behaviour. This section introduces a brief study of Hindi causative verbs: their classification, a linguistic model for their classification and their verb frames. The study presented here is already published as “A Preliminary Work on Causative Verbs in Hindi” [20] in Eighth Workshop on Asian Language Resources (ALR8) of COLING (International Conference on Computational Linguistics) 2010. While working on the annotation scheme for Hindi dependency Treebank, we were faced with the issue of annotation of sentences with causative verbs. As mentioned in Chapter 5, we had two options before us and we chose Possibility-II which says that there is a morphological relation between the base verb and the causative
verb. However, before arriving at our decision a study of the Hindi *causative* verbs and their classification was essential.

In the classification we show how different types of Hindi verbs have different types of *causative* forms. It will be a linguistic resource for Hindi *causative* verbs which can be used in various NLP applications. This resource enriches the already available linguistic resource on Hindi verb frames [22]. The classification is based on the morphology, semantics and syntax of the *causative* verbs. The morphology is captured by the word generation process; semantics is captured by the linguistic model followed for classifying the verbs and the syntax has been captured by the verb frames using relations given by Panini.

*Causative* verbs mean that *some actor makes somebody else do something or causes him to be in a certain state* [9]. The causal verb indicates the *causing of another to do something, instead of doing it oneself* [73]. Semantically *causative* verbs refer to a *causative situation* which has two components:

A. the causing situation or the antecedent;
B. the caused situation or the consequent.

These two combine to make a *causative situation* [126]. There are different ways in which causation is indicated in different languages. There are three types of causatives:

A. Morphological causatives,
B. Periphrastic causatives and
C. Lexical causatives [64].

**Morphological Causatives** indicate causation with the help of verbal affixes. Sanskrit, Hindi/Urdu, Persian, Arabic, Hebrew, Japanese, Khmer, and Finnish languages have morphological causatives (example is given in section 7.1.1). **Periphrastic causatives** indicate causation with the help of a verb which occurs along with the main verb. For example, in English in a sentence such as:

Ex- (7.1) **John made the child drink milk.**

In the above example 7.1, the verb *make* is expressing causation which is occurring along with the verb *drink* which in turn is expressing the main action. English, German, French, Korean, Tamil, Indonesian, and Thai are some of the languages which have periphrastic *causatives*. **Lexical causatives** are those in which there is no morphological similarity between the base verb root and the *causative* verb form. A different lexical item is used to indicate causation. For example, the *causative* of English *eat* is *feed.*
English and Japanese have lexical causatives. English has both periphrastic and lexical causatives. There are rare instances of lexical causatives in Hindi. An example of this is jaanaa ‘to go’ which is a base verb root and its lexical causative form is bhejnaa ‘to send’. However, the primary process of deriving causative verbs in Hindi is morphological and it is also highly productive.

7.1.1 Causative Verbs in Hindi

In Hindi, a base verb root changes to a causative verb when affixed by either an ‘-aa’ or a ‘-vaa’ suffix.

<table>
<thead>
<tr>
<th>Base verb</th>
<th>First causal</th>
<th>Second causal</th>
</tr>
</thead>
<tbody>
<tr>
<td>so ‘sleep’</td>
<td>sul-aa ‘put to sleep’</td>
<td>sul-vaa ‘cause to put to sleep’</td>
</tr>
</tbody>
</table>

In every step of causative derivation there is an increase in the valency of the verb (Kachru, 2006; 64)

Ex- (7.2) bachchaa (agent) soyaa
child slept
‘The child slept.’

Ex- (7.3) aayaa ne (causer) bachche ko (causee) sulaayaa
maid Erg. child Acc. put-to-sleep
‘The maid put the child to sleep.

Ex- (7.4) maa.N ne (causer) aayaa se (causee-causer [mediator]) bachche ko (causee)
mother Erg. maid by child Acc.
sulvaayaa
causd-to-put-to-sleep
‘Mother caused the maid to put the child to sleep.’

Hindi verbs are divided into two groups based on their behaviour in causative sentences: affective verbs and effective verbs [91]. The action of affective verbs benefits or affects the agent. Affective verbs will have both first and second causal forms. Verbs such as ronaay ‘to cry’ and dau.Dnaay ‘to run’ are affective intransitive verbs. Only verbs belonging to khaanaay ‘to eat’ class come under affective transitive verbs. The agent of the affective intransitive verb becomes the patient in the first causal and they both will take a ko postposition (Hindi case marker). Effective verbs and ditransitive verbs have only one causal form. The agent of the effective verb and ditransitive verb becomes the causative agent in the first causal. So this causative agent in the first causal
takes a se ‘with’ postposition (Hindi case marker). Verbs belonging to karna ‘to do’ class come under the effective verbs.

The major studies in Hindi causatives are Kachru [93], Kachru [92], Kachru [91], McGregor [114], Greaves [73], Kellogg [96], Agnihotri [9], Sahaay [154], Sahaay [153], Singh and Agnihotri [163] and Tripathi [173]. Kachru [93] has given the classification of Hindi verbs based on their causativization behavior. The others have mostly talked about the derivation process of the causative verbs. However the classification of causative verbs in Hindi remains an issue of discussion. Since they are morphologically related, the decision of what is the base verb form of these verbs remains a point of discussion. There are two approaches which are followed in deciding the base verb:

1. causative formation based only on morphology,
2. causative formation based on morphology and semantics.

### I. Based on Morphology

<table>
<thead>
<tr>
<th>Base verb (Intransitive)</th>
<th>First causal (Transitive)</th>
<th>Second causal (Causative)</th>
</tr>
</thead>
<tbody>
<tr>
<td>khul ‘open’</td>
<td>khol ‘open’</td>
<td>khulvaa ‘cause to open’</td>
</tr>
</tbody>
</table>

For verbs such as khol ‘open’, one needs to decide what is the root, whether it is khul ‘open’ or khol ‘open’. As mentioned above, at each step of morphological derivation, the number of arguments of these verbs increases by one. In approach I, khul ‘open’ is an intransitive verb which is taken as the base verb. khol ‘open’ and khulvaa ‘cause to open’ are derived from it by adding suffix ‘-aa’ and ‘-vaa’ respectively to the base verb [93, 92]. The arrow denotes the direction of derivation from base verb. The forward arrow denotes the increment of the argument from base to the causal forms.

### II. Based on Morphology and Semantics

<table>
<thead>
<tr>
<th>Intransitive (Intransitive)</th>
<th>Base verb (Transitive)</th>
<th>First causal (Causative)</th>
</tr>
</thead>
<tbody>
<tr>
<td>khul ‘open’</td>
<td>khol ‘open’</td>
<td>khulvaa ‘cause to open’</td>
</tr>
</tbody>
</table>

In approach II, khol ‘open’ is a transitive verb which is taken as the base verb. Here, other than morphology, the semantics of the verbs is also taken into consideration. Here khul ‘open’ and khulvaa ‘cause to open’ are derived from the base verb khol ‘open’. khulvaa ‘cause to open’ is derived from the
base verb by adding suffix ‘-vaa’ to it. *khul* ‘open’ is a derived intransitive form of *khol*. The agent of the base verb *khol* ‘open’ is not realized on the surface level of the derived intransitive verb *khul* ‘open’ though it is implied semantically. *khul*’ is an unaccusative verb. Thus, both unaccusative and *causative* forms of the verb are derived forms and the derivation in such cases is both forward and backward. From base verb to the derived intransitive it is a backward derivation which means there is a reduction of one argument from base verb to the derived intransitive verb [173, 149].

7.1.2 Our Approach

7.1.2.1 Linguistic Model for Classifying Causative verbs

As discussed above, we have followed Paninian Grammatical framework in this model as the theoretical basis for our approach. To recap, the meaning of every *verbal root* (*dhaatu*) consists of: (a) *activity* (*vyapaara*) and; (b) *result* (*phala*). *Activity* denotes the actions carried out by the various participants (*karakas*) involved in the action. *Result* denotes the state that is reached, when the action is completed [36]. The participants of the action expressed by the verb root are called *karakas*. The *karta karaka* is the locus of the *activity*. Similarly *karma karaka* is the locus of the *result*. The locus of the *activity* implied by the verbal root can be *animate* or *inanimate*. Example 7.2 given above, is the example where the locus of the *activity* is *animate*. Example 7.5 given below, is the example where the locus of the activity is *inanimate*.

Ex- (7.5) *darvaazaa (patient) khulaa*

\[
\text{door} \quad \text{opened}
\]

‘Door opened.’

Ex- (7.6) *raam ne (agent) darvaazaa (patient) kholaa*

\[
\text{ram} \quad \text{Erg.} \quad \text{door} \quad \text{opened}
\]

‘Ram opened the door.’

Ex- (7.7) *maiM ne(agent) raam se(causee-causer [mediator]) darvaazaa(patient) khulваayaa*

\[
\text{I} \quad \text{Erg.} \quad \text{ram} \quad \text{by} \quad \text{door} \quad \text{caused-to-open}
\]

‘I made Ram open the door.’

When we come to the *causatives*, the notion of *prayojaka karta* ‘causer’, *prayojya karta* ‘causee’ and *madhyastha karta* ‘mediator causer’ are introduced (discussed above in chapter-5). *prayojaka karta* ‘causer’ is the initiator of the action. *prayojya karta* ‘causee’ is the one who is made to do the action by the *prayojaka karta* ‘causer’. *madhyastha karta* ‘mediator causer’ is the causative agent of the action. The *karta* of the base verb becomes the *prayojya karta* of the first *causative* verb and the *prayojaka karta* of the first *causative* becomes the *madhyastha karta* of the second *causative*. Notice that all the three,
prayojaka karta, prayojya karta and madhyastha karta are animate. This model takes both semantics and morphology into consideration.

7.1.2.1.1 Semantics

Ex- (7.8) chaabii ne (agent) taalaa(patient) kholaa
    key    Erg.  lock  opened
‘The key opened the lock.’

Ex- (7.9)* raam ne (agent) chaabii se (causee-causer [mediator]) taalaa(patient) khulvaayaa
    ram    Erg.  key by  lock  caused-to-open
‘Ram caused the key to open the lock.’

Ex- (7.10) raam ne (agent) mohan dvaaraa (causee-causer [mediator]) chaabii se (instrument)
            taalaa (patient) khulvaayaa
    ram    Erg.  mohan by  key  with
    lock  caused-to-open
‘Ram made Mohan open the lock with the key.’

In example 7.8, chaabii ‘key’ is the karta of the transitive verb khol ‘open’.
chaabii ‘key’ is an inanimate karta so this sentence can’t be causativized. The attempt to cast the key as a mediator in 7.9 is unacceptable. Example 7.9, is actually interpreted as example 7.10 where an inanimate noun with a se ‘with’ postposition acts as an instrument and not as a prayojya karta ‘causee’. So in example 7.10, chaabii ‘key’ is an inanimate noun and takes se ‘with’ postposition so chaabii se ‘with the key’ acts as an instrument and mohan ‘Mohan’ acts as the prayojya karta ‘causee’ [93]. It seems that only those verbs can be causativized which take an animate karta. Out of the above two given approaches, we are following approach II where both morphology and semantics are considered. In our approach we are saying that only those base verbs can be causativized which take an animate karta and it should also have volitionality [173, 149]. Those base verbs which take an inanimate karta can’t be causativized. So in our approach the prayojya karta ‘causee’ in the causative sentence is always animate as the karta of the base verb becomes the prayojya karta ‘causee’ in the causative sentences. In our approach, we have the notion of karmakartri which says an intransitive can be derived out of a basic transitive verb and the karma of the basic transitive verb becomes the karta of the derived intransitive verb. So the karta of the derived intransitive verb is called karmakartri. The derived intransitive verbs are like unaccusative verbs of English. Whereas in approach I, the intransitive base verbs that take both animate and inanimate karta can be causativized. But in the case of transitives only, base verbs which take animate karta can be causativized. Similarly, in ditransitives also, base verbs which take only animate karta can be causativized [93, 92]. We follow the dependency tagging scheme proposed by Begum et al., [21] for the development of a dependency annotation for Indian Languages. In this scheme prayojaka karta ‘causer’, prayojya karta ‘causee’ and madhyastha karta ‘mediator causer’ are represented as pk1, jk1 and mk1 respectively.
Some of the base verb forms and their causative sentences are given below with dependency relations marked in the brackets for the appropriate arguments:

Ex- (7.11) raam ne(k1) seb(k2) khaayaa
    ram  Erg.  apple   eat.Pst
    ‘Ram ate an apple.’

Ex- (7.12) siitaa ne(pk1) raam ko(jk1) seb(k2) khilaayaa
    sita   Erg.   ram Acc.  apple eat.Caus.Pst
    ‘Sita fed Ram an apple.’

Ex- (7.13) maa.N ne(pk1) siitaa se(mk1) raam ko(jk1) seb(k2) khilvaayaa
    ‘Mother caused Sita to feed Ram an apple.’

Ex- (7.14) naukar ne(k1) kaam(k2) kiyaa
    servant Erg. work do.Pst
    ‘The servant did the work.’

Ex- (7.15) maalik ne(pk1) naukar se(mk1) kaam(k2) karvaayaa
    master Erg. servant by work do.Caus.Pst
    ‘The master caused servant to do the job.’

Ex- (7.16) raam ne(k1) siitaa ko(k4) kitaab (k2) dii
    ram  Erg.  sita Dat.  book give.Pst
    ‘Ram gave a book to Sita.’

Ex- (7.17) mohan ne(pk1) raam se(jk1) siitaa ko(k4) kitaab(k2) dilaaaii
    mohan Erg.  ram by sita Dat. book give.Caus.Pst
    ‘Mohan made Ram give a book to Sita.’

Ex- (7.18) mujhko(k4a) chaan.Nd(k1) dikhaa
    I.Dat. moon appear.Pst
    ‘The moon became visible to me.’

Ex- (7.19) maiM ne(k1) chaan.Nd(k2) dekhaa
    I Erg.  moon see.Pst
    ‘I saw the moon.’
Ex- (7.20) maiM ne(pk1) raam ko(jk1) chaa.Nd(k2) dikhaayaa
mother Erg. ram Dat. moon see.Caus.Pst
‘Mother showed moon to Ram.’

Ex- (7.21) maiM ne(pk1) mohan se(mk1) raam ko(jk1) chaa.Nd(k2) dikhlaayaa
I Erg. mohan by ram Dat. moon see.Caus.Pst
‘Mother made Mohan show moon to Ram.’

7.1.2.1.2 Morphology

In this section we have given the derivation process of the Hindi causative verbs. We have studied 160 Hindi verbs and have come up with certain number of rules for the derivation process of causative verbs.

When causative affixes are added to the base verb roots then some of the base verb roots change in form and some don’t. Various causal affixes are added to each verb type to form causatives. An example of affix addition for each verb type is discussed below. The affixes that are added are given in bold. The changes in the base verb root are underlined and made bold in both root form and the causal form.

I. Type-1 and its causative forms:

Suffix ‘-aa’ is added to the verb root to form the first causal and ‘-vaa’ to form the second causal.

No Change in the Root:

<table>
<thead>
<tr>
<th>Chip</th>
<th>Chip-aa</th>
<th>Chip-vaa</th>
</tr>
</thead>
<tbody>
<tr>
<td>‘hide’</td>
<td>‘hide’</td>
<td>‘cause to hide’</td>
</tr>
</tbody>
</table>

Change in the Root:

<table>
<thead>
<tr>
<th>aa</th>
<th>a</th>
</tr>
</thead>
<tbody>
<tr>
<td>naach</td>
<td>nach-aa</td>
</tr>
<tr>
<td>‘dance’</td>
<td>‘make someone dance’</td>
</tr>
</tbody>
</table>
II. **Type-2 and its causative forms:**

Suffix ‘-aa’ is added to the verb root to form the first causal and ‘-vaa’ to form the second causal.

**No Change in the Root:**

- **likh** → **likh-aa** → **likh-vaa**
  - ‘write’ → ‘dictate’ → ‘cause to dictate’

**Change in the Root:**

- **khaa** → **khī-l-aa** → **khī-l-vaa**
  - ‘eat’ → ‘feed’ → ‘cause to feed’

- **piī** → **pi-l-aa** → **pi-l-vaa**
  - ‘drink’ → ‘make someone drink’ → ‘cause to make someone drink’

III. **Type-3 and its causative forms:**

Suffix ‘-vaa’ is added to the verb root to form the first causal.

**No Change in the Root:**

- **kharīid** → **kharīid-vaa**
  - ‘buy’ → ‘cause to buy’

**Change in the Root:**

- **gaa** → **ga-vaa**
  - ‘sing’ → ‘cause to sing’
IV. Type-4 and its causative forms:

Suffix ‘-aa/-vaa’ is added to the verb root to form the first causal.

No Change in the Root:

paros \rightarrow paros-vaa

‘serve’ \rightarrow ‘cause to serve’

Change in the root:

➢ e \rightarrow i; In addition, 'l' is inserted here between the root and the causative suffix.

de \rightarrow di-l-aa /di-l-vaa

‘give’ \rightarrow ‘cause to give’

In case type-5 and type-6 verbs, we can derive intransitive verbs out of transitive verbs. Here we have two types of word formations:

(i) causative formation,
(ii) Derived intransitive verb formation

Causative derivation is the forward derivation and intransitive derivation is backward derivation.

V. Type-5 and its causative forms:

Suffix ‘-aa’ is added to the verb root to form the first causal and ‘vaa’ to form the second causal. In this verb type there is no example where the verb root form doesn’t change.

Causative Formation: Change in the root

➢ e \rightarrow i
Derived intransitive formation: *Change in the above root:*

- \( i \leftarrow e \)

\[ \text{dekh} \quad \text{dekh-aa} \quad \text{dekh-vaa} \]

‘see’ ‘show’ ‘cause to show’

\[ \text{dik}\quad \text{dik-aa} \quad \text{dik-vaa} \]

‘appear’ ‘cause to appear’

VI.  *Type-6 and its causative forms:*

Suffix ‘-aa/-vaa’ is added to the transitive verb root to form the first causal.

**Causative formation: *No change in the root***

\[ \text{bhar} \quad \text{bhar-vaa / bhar-aa} \]

‘fill’ ‘cause someone to fill’

**Derived intransitive formation: *No change in the root***

\[ \text{bhar} \quad \text{bhar} \]

‘to fill’ ‘to fill’

**Causative formation: *Change in the root***

- \( o \rightarrow u; \text{In addition, 'l' is inserted here between the root and the causative suffix} \)

\[ \text{dhu} \quad \text{dhu-l-aa / dhu-l-vaa} \]

‘wash’ ‘cause to wash’
Derived intransitive formation: Change in the above root:

> u ← o; In addition, 'l' is inserted at the end of the root

dhu-I ← dh

‘be washed’ ‘cause to wash’

In the following section, we have given the classification of Hindi causative verbs.

7.1.3 Classification of Hindi Causative Verbs

Here Hindi verbs have been classified into 6 types based on their causativization behavior:

7.1.3.1 Type-1: Basic Intransitive verb

Basic intransitive verb has two causal forms, i.e., first causal and second causal form. First causal of the basic intransitive verb functions as a transitive verb. The subject of the basic intransitive verb becomes the object of the transitive verb or the first causal form. The subject of the first causal form becomes the causative agent of the second causal form. Sentence (7.2) is the example of basic intransitive and sentences (7.3) and (7.4) are its causative forms.

7.1.3.2 Type-2: Basic Transitive verb type-I (which is similar to khaanaa ‘to eat’ verb type given by Kachru [93])

7.1.3.3 Type-3: Basic Transitive verb type-II (which is similar to karnaa ‘to do’ verb type given by Kachru [93])

Type-2 and type-3 are transitive verbs which are divided into two types based on their causativization behavior. Basic transitive verbs of type-I, like khaanaa ‘to eat’ have two causal forms. First causal of khaanaa ‘to eat’ type verb also functions as ditransitive. Whereas transitive verbs of type-II, like karnaa ‘to do’ have one causal form. First causal of karnaa ‘to eat’ type verb functions as causative. Sentences (7.11-7.13) are examples for type-2 verb. Sentences (7.14-7.15) are examples for type-3 verb.
### 7.1.3.4 Type-4: Basic Ditransitive verb

Ditransitive verbs also have one causal form. Sentences (7.16-7.17) are examples for type-4 verb.

### 7.1.3.5 Type-5: Basic Transitive verb type-I, out of which intransitive verbs can be derived which takes a dative subject,

### 7.1.3.6 Type-6: Basic Transitive verb type-II, out of which intransitive verbs can be derived.

Type-5 and type-6 are transitive verbs which have causal forms depending on whether it is type-I (khaanaa ‘to eat’) transitive or type-II (karna ‘to do’) transitive and in addition both have a derived intransitive form. Type-5 takes a dative subject in the derived intransitive form. Sentences (7.18-7.21) are examples for type-5 verb. Sentences (7.5-7.7) are examples for type-6 verb. Other than the 4 classes classified by Kachru [93], we have come up with two more extra classes, i.e., type-5 and type-6. An example for each verb type that goes into the classification is given below:

1) **Type-1**

<table>
<thead>
<tr>
<th>Base verb</th>
<th>First causal</th>
<th>Second Causal</th>
</tr>
</thead>
<tbody>
<tr>
<td>so (‘sleep’)</td>
<td>sulaa (‘put to sleep’)</td>
<td>sulvaa (‘cause to put to sleep’)</td>
</tr>
</tbody>
</table>

2) **Type-2**

<table>
<thead>
<tr>
<th>Base verb</th>
<th>First causal</th>
<th>Second Causal</th>
</tr>
</thead>
<tbody>
<tr>
<td>khaa (‘eat’)</td>
<td>khilaa (‘feed’)</td>
<td>khilvaa (‘cause to feed’)</td>
</tr>
</tbody>
</table>

3) **Type-3**

<table>
<thead>
<tr>
<th>Base verb</th>
<th>First causal</th>
<th>Second Causal</th>
</tr>
</thead>
<tbody>
<tr>
<td>kar (‘do’)</td>
<td>karaa/karvaa</td>
<td>‘cause to do’</td>
</tr>
</tbody>
</table>
4) **Type-4**

**Base verb**

*de*

‘give’

**First causal**

*dilaa/dilvaa*

‘cause to give’

5) **Type-5**

**Intransitive**

*dikh*

‘appear’

**Base verb**

*dekh*

‘see’

**First causal**

*dikhaa*

‘show’

**Second Causal**

*dikhvaa*

‘cause to show’

6) **Type-6**

**Intransitive**

*khul*

‘open’

**Base verb**

*khol*

‘open’

**First causal**

*khulvaa*

‘cause to open’

We have taken 160 base verbs and given their causative forms and classified them (See Appendix IX). Out of 155 base verbs we have listed 27 base verbs and their causative forms under Type-1; 19 base verbs and their causative forms under Type-2; 25 base verbs and their causative forms under Type-3; 6 base verbs and their causative forms under Type-4; 1 base verb and their causative form under Type-5; 77 base verbs and their causative forms under Type-6.

### 7.1.4 Method

For this work, 160 base verbs were taken, their causative forms were given and were classified. These 160 verbs were selected based on their common usage in the language. Rules for deriving causative verb forms from their base forms were made. Verb frames for base verbs and their causative forms were developed. Based on the analysis of the base verbs, certain problem cases were reported and generalizations regarding causativization were made.

The implementation of the causative verbs is computationally possible. In the implementation of the causative verbs, the causative feature of a verb can be reflected in the morph analysis. Though we have not implemented this, these are two possible ways to implement causative information:

1. All the causative verb roots can be included in the root dictionary of the morph analyzer with an additional feature marking it a causative verb type.
(ii) For all causative verbs the following information can be marked; causative root, base root, verb type and causative suffix.

In (i), the information of base verb root from which the causative root is derived is missing which is captured in (ii). In the above mentioned two ways the latter gives more detailed information than the former.

### 7.1.5 Verb Frames

In this section we list out the syntactic frames for all the causative types discussed in the previous sections. Verb frame [22] is given for the base form and for its first and second causal form. For ease of exposition, below we show only the relevant information of a verb frame. Components not necessary for this present discussion have been left out. Here the structure of a verb frame is given in terms of dependency relation, postposition (Hindi case marker) and TAM. We have taken past tense (*yaa* is the past tense marker) in the TAM. Refer to examples given above for each type of causatives for a better understanding of the frames.

**I. Frame of Type-1 and its Causative Forms:**

<table>
<thead>
<tr>
<th>Relation-Postposition</th>
<th>TAM</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Ex-7.2) k1-0</td>
<td>yaa</td>
</tr>
<tr>
<td>(Ex-7.3) pk1-ne</td>
<td>jk1-ko</td>
</tr>
<tr>
<td></td>
<td>yaa</td>
</tr>
<tr>
<td>(Ex-7.4) pk1-ne</td>
<td>mk1-se</td>
</tr>
<tr>
<td></td>
<td>jk1-ko</td>
</tr>
<tr>
<td></td>
<td>yaa</td>
</tr>
</tbody>
</table>

**II. Frame of Type-2 and its Causative Forms:**

<table>
<thead>
<tr>
<th>Relation-Postposition</th>
<th>TAM</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Ex-7.11) k1-ne</td>
<td>k2-0</td>
</tr>
<tr>
<td></td>
<td>yaa</td>
</tr>
<tr>
<td>(Ex-7.12) pk1-ne</td>
<td>jk1-ko</td>
</tr>
<tr>
<td></td>
<td>k2-0</td>
</tr>
<tr>
<td></td>
<td>yaa</td>
</tr>
<tr>
<td>(Ex-7.13) pk1-ne</td>
<td>mk1-se</td>
</tr>
<tr>
<td></td>
<td>jk1-ko</td>
</tr>
<tr>
<td></td>
<td>k2-0</td>
</tr>
<tr>
<td></td>
<td>yaa</td>
</tr>
</tbody>
</table>
III. Frame of Type-3 and its Causative Forms:

<table>
<thead>
<tr>
<th>Relation-Postposition</th>
<th>TAM</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Ex-7.14) k1-ne</td>
<td>k2-0 yaa</td>
</tr>
<tr>
<td>(Ex-7.15) pk1-ne jk1-se</td>
<td>k2-0 yaa</td>
</tr>
</tbody>
</table>

IV. Frame of Type-4 and its Causative Forms:

<table>
<thead>
<tr>
<th>Relation-Postposition</th>
<th>TAM</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Ex-7.16) k1-ne</td>
<td>k4-ko k2-0 yaa</td>
</tr>
<tr>
<td>(Ex-7.17) pk1-ne jk1-se</td>
<td>k4-ko k2-0 yaa</td>
</tr>
</tbody>
</table>

V. Frame of Type-5 and its Causative Forms:

<table>
<thead>
<tr>
<th>Relation-Postposition</th>
<th>TAM</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Ex-7.18) k4a-ko</td>
<td>k1  aa</td>
</tr>
<tr>
<td>(Ex-7.19) k1-ne</td>
<td>k2-0 aa</td>
</tr>
<tr>
<td>(Ex-7.20) pk1-ne jk1-ko</td>
<td>k2-0 yaa</td>
</tr>
<tr>
<td>(Ex-7.21) pk1-ne mk1-se</td>
<td>jk1-ko k2-0 yaa</td>
</tr>
</tbody>
</table>

VI. Frame of Type-6 and its Causative Forms:

<table>
<thead>
<tr>
<th>Relation-Postposition</th>
<th>TAM</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Ex-7.5) k1-0</td>
<td>aa</td>
</tr>
<tr>
<td>(Ex-7.6) k1-ne</td>
<td>k2-0 aa</td>
</tr>
<tr>
<td>(Ex-7.7) pk1-ne jk1-se</td>
<td>k2-0 yaa</td>
</tr>
</tbody>
</table>
7.1.6 Issues and Observations

There are some verbs which cannot be causativized. Motion verbs like aa ‘come’ and jaa ‘go’ cannot be causativized. After analysing a certain amount of corpora we have observed that not all motion verbs behave like the above verbs. aanaa ‘to come’ and jaanaa ‘to go’ verbs can’t be causativized because these verbs always occur as main verbs and take the following verbs as manner adverbs: chalnaa, bhaagnaa, daudnaa. For instance, chalkar aayaa ‘came running’ and dau.Dkar gayaa ‘went running’. Those motion verbs which occur as manner adverbs and modify another motion verb can be causativized and those verbs which occur as main verbs and never occur as manner adverbs of another motion verb can’t be causativized. Natural process verbs like khil ‘blossom’, garajnaa ‘thunder’ and uga ‘rise’ also can’t be causativized.

There are three types of the verb nikal ‘leave’. All the three are used as intransitives.

- **Derived intransitive**: sense → drain out

  Ex- (7.22) *paanii kamre se nikal gayaa*
  
  water room from leave go.Pst.
  
  ‘Water drained out of the room.’

- **Basic Intransitive**: sense → walked out

  Ex- (7.23) *raam kamre se baahar nikal gayaa*
  
  ram room from out leave go.Pst.
  
  ‘Ram walked out of the room.’

- **Basic Intransitive which involves natural process**

  Ex- (7.24) *gangaa gangotrii se nikaltii hai*
  
  ganga gangotri from emerge.Impf. be.Pres.
  
  ‘Ganga emerges from Gangotri.’
The first type is a derived intransitive which is derived from the base transitive verb *nikaal* ‘remove’. This base transitive verb root can be causativized. The second type is basic intransitive which can also be causativized. The third type which is natural process can’t be causativized. This shows how important the property of animacy for making causatives is.

7.1.7 Dependency Analysis of Causatives

As mentioned in chapter 5, we have two possibilities for the dependency analysis of causatives:

- **Possibility-I**: Annotate their arguments based on the syntactic properties of various participants
- **Possibility-II**: Annotate the arguments starting from the semantics of the verb.

The dependency tree for the example 7.25 according to Possibility-I is given below:

Ex- (7.25) *maa.N ne (k1) aayaa se (k3) bachche ko (k4) khaanaa (k2) khilvaayaa.*

mother Erg. maid by child Dat. food caused-to-feed

'Mother caused the maid to feed the child'.

Figure 7.1 Causatives Possibility-I

Possibility-I gives the syntactic analysis. In the above tree, *khilvaa* ‘casue to feed’ is the verb root and the arguments of the verb *khilvaa* ‘cause to feed’ are marked as follows: *maa.N* ‘mother’ is marked as *karta* because of *ne* vibhakti; *aayaa* ‘maid’ is marked as *karana* because of *se* vibhakti; *bachcha* ‘child’ is marked as *sampradana* because of *ko* vibhakti.
The dependency tree for the example 7.25 according to Possibility-II is:

```
khilvaayaa 'caused to feed' <root = khaa 'eat'>

pk1        mk1       jk1       k2
maa.N ne   aayaa se  bachche ko  khaanaa
'mother Erg.' 'by aayaa' 'to child' 'food'
```

Figure 7.2 Causatives Possibility-II

Possibility-II marks causativity. It is semantically richer and captures morphological relatedness. In the above tree, the verb root is khaa ‘eat’, i.e., the base verb root of the causative verb khilvaa ‘cause to feed’ is khaa ‘eat’. The causative verb khilvaa ‘cause to feed’ is derived from base verb khaa ‘eat’. As there is morphological relatedness between the base verb khaa ‘eat’, and causative verb khilvaa ‘cause to feed’, we mark the arguments of the verb khilvaa ‘cause to feed’ as follows: maa.N ‘mother’ is marked as prayojaka karta ‘causer’ (pk1); aayaa ‘maid’ is marked as madhyastha karta ‘mediator causer’ (mk1); bachchaa ‘child’ is marked as prayojya karta ‘causee’ (jk1).

Out of the above two, we have chosen Possibility-II as it gives an appropriate representation of the causatives. In Possibility-I, we are not able to capture the relation between causative verbs and their base verbs whereas Possibility-II allows us to do this. This also maps to PropBank analysis.

After discussing causatives, we proceed to conjunct verbs in section 7.2. It is one of the important topics and requires major discussion like causatives of Hindi.

### 7.2 Conjunct Verbs in Hindi

Verbs which are composed of more than one unit are referred to as Complex Predicates (CP). A CP is generally composed by combining nouns, adjectives and verbs with other verbs which are primarily support verbs referred to as vectors. There are two types of Complex Predicates (CP): (i) ‘Noun/Adjective+Verb’ combinations, called conjunct verbs, and (ii) ‘Verb+Verb’ combinations, called compound verbs. CPs are highly frequent in South Asian languages [82, 122, 10]. The verb in the CP is referred to as light verb and the element that the light verb combines to form a CP is referred to as the host [123].
According to Butt [53], the light verb, in Hindi/Urdu, 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. Butt [51] has talked about four types of complex predicates: (a) In Syntactic Complex Predicates, the formation happens in the syntax. (b) In Morphological Complex Predicates, a part of morphology is used to modify the primary event predication. Example for this is morphological causatives. (c) Light Verbs cross linguistically do not always form a consistent syntactic category. (d) In Semantics, complex predicates represent the decomposition of event structure.

‘Noun+Verb’ constructions in Hindi, and in fact in other Indian languages, are a highly productive process. However, a major issue that one encounters in literature on conjunct verbs is that which Noun+Verb sequence should be considered as a conjunct verb and which should not be. In other words, whether the given sequence of noun and verb is a conjunct verb or is a verb and its object. It, obviously, is an issue for us as well, both for Treebank annotation and parsing. There is a substantial amount of literature on which N+V makes a conjunct verb and why. The literature proposes a number of diagnostics for identifying conjunct verbs. We will first look at the diagnostics presented by various scholars for identifying a ‘noun+verb’ combination as a conjunct verb.

We will also show how many of these diagnostics may not always work for conjunct verb identification and therefore, using them for Treebank annotation may not be very effective. 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 [30].

This work can also greatly help in automatically augmenting a lexical network such as the Hindi Wordnet. Previous automatic identification approaches made use of parallel English corpora [125, 166] which makes use of the property that single verb in English will break into two components, i.e., noun/adjective and verb in Hindi. Venkatapathy and Joshi [183] also make 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 study presented here is already published as “Identification of Conjunct verbs in Hindi and its effect on Parsing Accuracy” [19] in 12th International Conference on Intelligent Text Processing and Computational Linguistics (CICLING) 2011.

### 7.2.1 Conjunct Verbs in Hindi

The most frequent verbs (verbalizers) which occur in a conjunct verb in Hindi are karnaa ‘to do’, honaa ‘to be’, denaa ‘to give’, lenaa ‘to take’, aanaa ‘to come’, etc. In our scheme, we mark the relation between a noun/adjective and its verbalizer as ‘pof’ (Part OF) (Refer chapter 3 and chapter 5).
In example 7.26, kShamaa ‘forgiveness’ is a noun which is combined with the verb karnaa ‘to do’ to express the sense of the verb ‘to forgive’.

Since identification of conjunct verbs is a difficult area, coming up with clear diagnostics for doing the same in our guidelines was one of the most challenging part of this study. There is plenty of discussion on this issue in linguistic literature and we surveyed it thoroughly to come up with reliable diagnostics to identify conjunct verbs both while annotating and also, if possible, automatically. Most of the diagnostics that exist in the literature work up to a point and do not yield very reliable results. In the following section, we present the diagnostics found in the literature and whether we could use them for our purpose. We also present our study on some reliable diagnostics which could be used to automatically identify a conjunct verb and how it helps in better parsing for Hindi.

### 7.2.2 Diagnostics to Identify Complex Predicates

Although, literature talks about various diagnostics, it is also not clear how far these diagnostics work for conjunct verb identification and whether there is a consensus on strict adherence to these diagnostics for identifying Noun+Verb (N+V) sequence as a conjunct verb. The following are some of the diagnostics mentioned in the literature ([123], [43]) 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.

**Ex- (7.27) **

\[\text{log} \quad \text{niind} \quad \text{aur} \quad \text{jamhaaii} \quad \text{le} \quad \text{rahe} \quad \text{the} \quad [43] \]

people sleep and yawn taking was
‘People were sleeping and yawning.’

**Ex- (7.28) **

\[\text{log} \quad \text{niind} \quad \text{le} \quad \text{rahe} \quad \text{the} \quad \text{aur} \quad \text{jamhaaii} \quad \text{le} \quad \text{rahe} \quad \text{the} \quad [43] \]

people sleep taking was and yawn taking was
‘People were sleeping and yawning.’

According to [43], the example 7.27 is ungrammatical because *niind* ‘sleep’ and *jamhaaii* ‘yawn’ are conjoined by *aur* ‘and’, whereas these nouns are part of Complex Predicates (CPs). Sentence 7.28 is grammatical because here the *N+V* combination, i.e., *niind* le ‘sleep’ and *jamhaaii* le ‘yawn’ have been conjoined with *aur* ‘and’. However, one also finds example, such as 7.29 given below:

**Ex- (7.29) **

\[\text{log} \quad \text{bai} \quad \text{-} \quad \text{bai} \quad \text{The} \quad \text{niind} \quad \text{aur} \quad \text{jamhaaii} \quad \text{le} \quad \text{rahe} \quad \text{the} \quad [43] \]

people sitting-sitting sleep and yawn taking was
‘People were sleeping and yawning.’

If we go by Bhattacharya’s [43] claim then the above example (7.29) becomes ungrammatical or *niind* le ‘sleep’ and *jamhaaii* le ‘yawn’ are not *conjunct* verbs. Moreover, many Hindi speakers find example 7.29 to be acceptable. Since *niind* le ‘sleep’ and *jamhaaii* le ‘yawn’ are non-compositional, they become a strong contender for *conjunct* verbs. Therefore, syntactically also *niind* ‘sleep’ and *jamhaaii* ‘yawn’ doesn’t prove to be a strong object of the verb *le* ‘take’. Syntactically, if we take passivization test to see if the sentences containing *niind* le ‘sleep’ and *jamhaaii* le ‘yawn’ are acceptable or not after passivizing.

**Ex- (7.30) **

\[\text{?logoM} \quad \text{dvaaraa} \quad \text{niind} \quad \text{lii} \quad \text{gayi} \quad \text{people-Obl} \quad \text{by} \quad \text{sleep} \quad \text{taken} \quad \text{Pasv.} \]

‘Sleep was taken by the people.’

**Ex- (7.31) **

\[\text{?logoM} \quad \text{dvaaraa} \quad \text{jamhaaii} \quad \text{lii} \quad \text{gayi} \quad \text{people-Obl} \quad \text{by} \quad \text{yawn} \quad \text{taken} \quad \text{Pasv.} \]

‘Yawn was taken by the people.’
The question arises whether the examples 7.30 and 7.31 are good or bad? They don’t look to be good. So, coordination as a test is a weak test. Passivization in this case looks awkward and doesn’t seem to be acceptable by Hindi speakers, so *niind le ‘sleep’ and jamhaaii le ‘yawn’ become a strong contender for conjunct verbs. Therefore coordination itself becomes a weak diagnostic since the following points also helped us in concluding that *niind le ‘sleep’ and jamhaaii le ‘yawn’ become a strong contender for conjunct verbs:

1) Semantically the above verbs are non-compositional, and
2) Syntactically passivization is not possible.

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

Ex- (7.32) *raam ne roTi lii
   ram Erg roti took
   ‘Ram took roti.’

Ex- (7.33) raam ne kyaa lii?
   ram Erg what took
   ‘What did Ram take?’

Ex- (7.34) raam ne jamhaaii lii
   ram Erg yawn took
   ‘Ram yawned.’

Ex- (7.35) *raam ne kyaa lii? [43]
   ram Erg what took
   ‘What did Ram take?’

Ex- (7.36) raam ne kyaa kiyaa? [43]
   ram Erg what did
   ‘What did Ram do?’

According to [43], example 7.35 is ungrammatical because only the noun *jamhaaii ‘yawn’ of the CP given in example 7.34 is the focus of the question in example 7.35, whereas example 7.36 is grammatical because the N+V combination, *jamhaaii lii ‘yawned’ given in example 7.34 is the focus of the question in example 7.36.

23
Example in 7.33, is grammatical as the questioned element here is \textit{roTii} ‘Roti’ which is an object noun given in example 7.32. So the question in example 7.33, used to question \textit{roTii} ‘roti’ given in example 7.32, is grammatical whereas the example given in 7.35 is ungrammatical when it is used to question \textit{jamhaaii} ‘yawn’ given in 7.34. Question given in example 7.36 is grammatical as it is used to question \textit{Noun + Verb} combination \textit{jamhaaii lii} ‘yawned’. This diagnostic was acceptable by some of the Hindi speakers and some objected it. As mentioned above, \textit{jamhaaii le} ‘yawn’ is non-compositional as well as passivization of the sentence (refer example 7.31) having this verb looks awkward. Hence, \textit{jamhaaii le} ‘yawn’ becomes a strong contender for \textit{conjunct} verb. Therefore Constituent Response Test diagnostic itself becomes a weak diagnostic since some of the Hindi speakers have accepted it. But, \textit{jamhaaii lii} ‘yawn’ becomes a strong contender for \textit{conjunct} verb because semantically, it is non-compositional and syntactically, passivization is not possible.

III. Relativization (D3): \textit{CP} internal nominals cannot be relativized.

\textbf{Ex- (7.37)} *\textit{raam ne vaha snaan [jo bahut pavitra hai] gangaa taT par kiyaa}
\textit{ram Erg. that bath which lot pious is ganga bank on did}
‘The bath which Ram did on the bank of river Ganga is very pious.’

\textbf{Ex- (7.38)} *\textit{vaha dhyaan [jo bahut mushkil hai] raam ne apne kaam par diyya}
\textit{that concentration which lot difficult is ram Erg self work on gave}
‘The concentration which Ram paid on his work is very difficult.’

According to Mohanan [123], this diagnostic is acceptable for deciding the N+V combination as \textit{conjunct} verb. According to this diagnostic, examples 7.37 and 7.38 are ungrammatical because \textit{snaan} ‘bath’ and \textit{dhyaan} ‘concentration’ which are nouns internal to \textit{CP} have been relativized by the relative clauses. This diagnostics was unacceptable by many Hindi speakers out of 20 speakers and for some speakers it was acceptable. Example 7.38 was relatively less acceptable by speakers when compared to example 7.37. In case of example 7.38, if we look at the argument structure of \textit{de} ‘give’, we see that there is an argument \textit{apne kaam par} ‘on one’s work’ which is not an argument taken by the verb \textit{de} ‘give’. The argument \textit{apne kaam par} takes a \textit{par} ‘on’ \textit{vibhakti} which is a locative case marker and the verb \textit{de} ‘give’ doesn’t take location as its argument. By this, we can say that \textit{apne kaam par} ‘on one’s work’ is an argument of \textit{dhyaan de} ‘concentrate’. Thus, \textit{dhyaan de} ‘concentrate’ is a strong contender for \textit{conjunct} verb. In case of passivization also, example 7.40 relatively doesn’t look good when compared to example 7.39. Here, passivization of \textit{snaan karna} ‘to bathe’ was acceptable by most Hindi speakers but not the passivization of \textit{dhyaan denaa} ‘to concentrate’. Hence, \textit{dhyaan de} ‘concentrate’ is the strong contender for \textit{conjunct} verb in comparison to \textit{snaan karna} ‘to bathe’. Here, I am also analysing the types of arguments that can’t be taken by the verb and can be taken by the N+V combination. Therefore Relativization diagnostic itself becomes a weak diagnostic.
Ex- (7.39) ?raam dvaaraa gaMgaa taT par snaan kiyaa gayaa
ram by ganga bank on bath done Pasv.
‘Bath was done by Ram on the banks of the river.’

Ex- (7.40) ?raam dvaaraa kaam par dhyaan diyaa gayaa
ram by work on concentration gave Pasv.
‘Concentration was done on work by Ram.’

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

Ex- (7.41) *raam ne us jamhaaii ko liyaa ... [43]
ram Erg that yawn Acc. took ...
‘Ram took that yawn…..’

According to [43], example 7.41 is ungrammatical because jamhaaii ‘yawn’ which is noun internal to the CP jamhaaii lii ‘yawned’, has taken an accusative case marker ko. This diagnostic seems to work in most of the cases. Most of the native Hindi speakers don’t accept a sentence as grammatical, if there is an accusative case marker ‘ko’ with the noun of the N+V combination. By this, we can say that Accusative case marker test is a reliable diagnostic to decide whether a given N+V combination is a conjunct verb or not. Hence, jamhaaii lenaa ‘to yawn’ is a strong contender for conjunct verb.

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

Ex- (7.42) raam ne yaha nirdesh diyaa
ram Erg. this order gave
‘Ram gave this order.’

In the example 7.42, the question is whether yaha ‘this’ is modifying just the Noun or the N+V combination? But most of the native Hindi speakers consider that yaha ‘this’ is modifying just the noun, i.e., nirdesh ‘order’ and not the N+V combination, i.e., nirdesh denaa ‘to order’. Here, the passivization of the sentence having this N+V combination too looks good and was also acceptable by most of the Hindi speakers:
Ex- (7.43) raam dvaaraa yaha nirdesh diyaa gayaa
     ram by gave order gave Psv.
     ‘This order was given by Ram.’

Semantically nirdesh denaa ‘to order’ is non-compositional. At the same time, denaa ‘to give’ doesn’t take a clausal object but in the example below diyaa ‘give’ is taking a clausal object:

Ex- (7.44) raam ne yaha nirdesh diyaa (kii kal se sab kaam kareMge)
     ram Erg. this order gave that tomorrow from everyone work do.Fut.
     ‘Ram gave this order that everybody will work from tomorrow.’

By the example 7.44, we can say that the clausal object in the above sentence is the argument of nirdesh denaa ‘to order’. Hence, nirdesh denaa ‘to order’ is a strong contender for conjunct verb. Since the following points that nirdesh denaa ‘to order’ is non-compositional and takes a clausal object help in concluding that nirdesh denaa ‘to order’ is a strong contender for conjunct verb. Hence, Adding the Demonstrative Pronoun diagnostic itself becomes a weak diagnostic.

VI. Adjectival Modification (D6): CP-internal nominals allow no modification.

Ex- (7.45) * siita ko (bahut krodh) aayaa.
     sita Acc. lot anger came
     ‘Sita was very angry.’

Ex- (7.46) siita ko (bahut) (krodh aayaa).
     sita Acc. lot anger came
     ‘Sita was very angry.’

According to Bhattacharya et al. [43] and Mohanan [123], example 7.45 is ungrammatical because here only CP internal nominal krodh ‘anger’ is being modified by the adjective bahut ‘lot’, whereas example 7.46 is grammatical as the whole N+V sequence krodh aayaa ‘get angry’ is being modified by bahut ‘lot’. So according to this diagnostic, noun of N+V combination can’t take adjectival modifiers. According to the majority of the Hindi speakers, bahut ‘lot’ is just modifying krodh ‘anger’ and not the N+V sequence krodh aayaa ‘get angry’. This is similar to the above diagnostic (D5) above, where a demonstrative pronoun is modifying the noun. In the diagnostic-D3 also a relative clause is modifying the noun of the N+V combination. According to most of the native Hindi speakers, just the nouns of the N+V combinations in most of these three diagnostics (D3, D5 and D6), can be modified. Hence, all these
diagnostics are weak. Here, krodh aanaa ‘to get anger’ is non-compositional. krodh ‘anger’ and aanaa ‘to come’ are two words and have their own meanings but when combined together they give a different meaning. Therefore, krodh aanaa ‘to get angry’ is a strong contender for conjunct verb.

VII. apnaa-apnaa ‘self-self’ Test (D7): We can add apne-apne ‘self’ to the object noun and not to the noun of the N+V (CP) combination.

Ex- (7.47) *unhoM ne sab apnii apnii shaadii kii
they Erg. everyone self self marriage did
‘They all got themselves married.’

Ex- (7.48) apne apne savaal puuCho
self self questions ask
‘Ask your questions.’

Example 7.47 is ungrammatical because apnii apnii ‘own self’ is modifying the noun of the CP, whereas example 7.48 is grammatical as apnii apnii ‘own self’ is modifying the object noun. For most of the Hindi speakers, example 7.47 was acceptable with the apnii apnii ‘own self’ modifying the noun shaadii ‘marriage’ of the N+V combination shaadii karnaa ‘to marry’. The passivization of the sentence having the N+V combination shaadii karnaa ‘to marry’ (example 7.49) was also acceptable to some extent. But, the passivization of another sentence having shaadii karnaa ‘to marry’ (example 7.50) looks bad:

Ex- (7.49) ?unake dvaaraa shaadii kii gayii
they-Gen. by marriage did Pasv.
‘Marriage was done by them.’

Ex- (7.50) ?raam dvaaraa siitaa se shaadii kii gayii
ram by sita with marriage did Pasv.
‘Marriage was done by Ram with Sita.’

Hence, shaadii karnaa ‘to marry’ becomes a strong contender for the conjunct verb. Therefore, apnaa-apnaa ‘self-self’ diagnostic itself becomes a weak diagnostic. This is also an example of noun modification of N+V combination.
VIII. Ellipses (D8): In answer to yes-no questions in Hindi, the predicate of a clause can stand for the entire clause.

Ex- (7.51) *kyaa raam ne aaj bahut kaam kiyaa?  
what ram Erg. today lot work did  
‘Did Ram do lot of work today?’

Ex- (7.52) *haa.N, kiyaa  
yes did  
‘Yes, he did.’

Answer to the yes-no interrogative in 7.51 is given in 7.52. Everything in the answer is omitted except the predicate.

Ex- (7.53) *kyaa raam ne mohan par bharosaa kiyaa?  
What ram Erg. mohan on reliance did  
‘Did Ram trust Mohan?’

Ex- (7.54) *haa.N kiyaa.  
yes did  
‘Yes, he did.’

Ex- (7.55) haa.N bharosaa kiyaa.  
yes trust did  
‘Yes, he trusted.’

According to Mohanan [123], sentence 7.54 is ungrammatical because the noun internal to a CP cannot be omitted in this manner. Whereas sentence 7.55 is grammatical because the noun internal to a CP is present in the sentence. But for most of the Hindi speakers, omitting the noun internal to a CP while answering the yes-no interrogative is fine. The passivized sentence of the verb bharosaa karnaa seems well with most of the Hindi speakers:

Ex- (7.56) raam dvaaraa mohan par bharosaa kiyaa gayaa  
ram by mohan on reliance did Pasv.  
‘Mohan was trusted by Ram’

But, if we look at the example 7.56, there is an argument mohan par ‘on Mohan’ which is not the argument taken by the verb karnaa ‘to do’. The argument, mohan par ‘on Mohan’ takes a par ‘on’ vibhakti which is a locative case marker and the verb kar ‘do’ doesn’t take location as its argument. By
this, we can say that *mohan par* is an argument of *bharosaa kar* ‘trust’. Thus, *bharosaa kar* ‘trust’ is a strong contender for conjunct verb. So *Ellipses* diagnostic itself becomes a weak diagnostic.

IX. Movement (D9):

Ex- (7.57) *
*raam ne  *bharosaa  *mohan par  *kiyaa
  ram Erg. trust mohan on did
  ‘Ram trusted Mohan.’

*Mohanan, 1994*

In 7.57, all the nominals except the one internal to *CP* are free to scramble. According to [43] and [123], the reason for ungrammaticality in 7.57 is that the nominal ‘*bharosaa*’ *trust*, within the *CP*, has moved away from the verb *kiyaa* ‘did’.

Ex- (7.58)  *raam ne  *bharosaa  *kiyaa  *mohan par
  ram Erg. reliance did mohan on
  ‘Ram trusted Mohan.’

Example 7.58 is grammatical because the N+V sequence *bharosaa kiyaa* has been scrambled here. This again is a very weak diagnostic as according to many Hindi speakers, the noun of the N+V combination can be moved away from the verb. For example, it can even be said:

Ex- (7.59)  *bharosaa  raam ne  *mohan par  *kiyaa
  trust ram Erg. mohan on did
  ‘Ram trusted Mohan.’

In the above example 7.59, ‘*bharosaa*’ *trust*, from the *CP  bharosaa kiyaa* has been moved away from the verb *kiyaa*. For the sake of focus, the noun of the CP has been moved away from the the verb *kiyaa*. We have already proved in the above diagnostics – (D8) that *bharosaa karnaa* ‘to trust’ is a strong contender for the conjunct verb. So *Movement* diagnostic itself becomes a weak diagnostic.

Diagnostics, by and large don’t help us in identifying conjunct verbs. I had lots of discussions on these diagnostics with a group of native Hindi speakers who had given their opinions on these diagnostics which I have given above along with each diagnostics. Out of the 9 diagnostics mentioned above, I came up with two diagnostics, i.e., *Adding the Demonstrative Pronoun* (D5) and *apnaa-apnaa ‘self-self’ Test* (D7) diagnostics. The rest of the 7 diagnostics are taken from the literature ([123]; [43]). Out of the remaining 7 diagnostics taken from the literature, only *Adding Accusative case marker* (D4) is a reliable diagnostic to decide whether a given N+V combination is a *conjunct* verb or not. The two diagnostics
given by me i.e., *Adding the Demonstrative Pronoun* (D5) and *apnaa-apnaa ‘self-self’ Test* (D7) are also weak.

To justify the above diagnostics we did a survey of these tests among another group of native speakers of Hindi Language. We conducted a manual experiment where we selected 20 people who are working as Hindi language editors and have good knowledge of Hindi. We took the above listed diagnostic tests and manually added a set of sentences with conjunct verbs to each diagnostic test. Then we distributed these tests among the 20 people and asked them to tick mark the sentences which were acceptable according to them. Based on these results, we formulated a table showing the results. Below given is the report on diagnostics evaluation which we did taking the first 5 diagnostics. The diagnostics evaluation is shown in the table 7.1 given below.

We also conducted another experiment where we have tried to automatically identify the conjunct verbs with the help of certain features and have used Maximum entropy model for the identification.

### 7.2.3 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. We took conjunct verbs and applied the above diagnostics to see how they fare in a subjective evaluation. Table 7.1 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, Table 7.1 shows the result only for only 7 verbs. The study considers a total of 20 verbs.
Table 7.1. Results of the subjective evaluation

<table>
<thead>
<tr>
<th>Noun+Verb</th>
<th>D1</th>
<th>D2</th>
<th>D3</th>
<th>D4</th>
<th>D5</th>
<th>CV</th>
</tr>
</thead>
<tbody>
<tr>
<td>rucii le</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve/ve</td>
<td>+ve</td>
<td>+ve</td>
<td>Yes</td>
</tr>
<tr>
<td>‘take interest’</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>maar khaa</td>
<td>+ve</td>
<td>+ve</td>
<td>-ve</td>
<td>+ve</td>
<td>+ve</td>
<td>Yes</td>
</tr>
<tr>
<td>‘get beaten’</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bhaag le</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve</td>
<td>-ve</td>
<td>Yes</td>
</tr>
<tr>
<td>‘participate’</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>snaan kar</td>
<td>-ve</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve</td>
<td>Yes</td>
</tr>
<tr>
<td>‘bathe’</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>chalaang</td>
<td>-</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve</td>
<td>Yes</td>
</tr>
<tr>
<td>maar ‘jump’</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bhojan kar</td>
<td>-ve</td>
<td>-</td>
<td>-ve</td>
<td>+ve</td>
<td>-ve</td>
<td>may be</td>
</tr>
<tr>
<td>‘eat’</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>havaa khaa</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve</td>
<td>Yes</td>
</tr>
<tr>
<td>‘feel air’</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

After exploring the behavioral diagnostics to identify the conjunct verb, we now move on to automate this task of identification.

7.2.4 Automatic Identification of Conjunct Verb

In the previous sections, various tests were explored for manual identification of conjunct verbs. Now, we explain the methodology used for building a statistical tool for automatic identification of conjunct verbs. We didn’t focus on compound verbs (verb + verb) as we didn’t work on it. However, conjunct verbs are more problematic and [54] has already reported a high accuracy of 98% for compound verbs. We trained a binary classification model using maximum entropy. The model classifies a noun/adjective-verb pair into either conjunct verb or literal class (non-conjunct verb).

7.2.4.1 Corpus

We used two different datasets that are part of CPG framework [36] based annotated Dependency
Treebanks for Hindi. Dataset-1 is part of the multi-layered/multi-representational Treebank for Hindi [42, 188], whereas Dataset-2 is part of HyDT (Hyderabad Dependency Treebank for Hindi) [21].

1) Dataset-1: It has 4500 manually annotated sentences (200k words approx.). This dataset belongs to the multi-layered/multi-representational treebank for Hindi [42, 188] which was released for the tools’ contest on IL parsing in ICON 2010 [85]. This dataset was used as a training data.

2) Dataset-2: Has 1800 sentences. This dataset belongs to the HyDT (Hyderabad Dependency Treebank for Hindi) which was released as part of the ICON’09 tools contest on Indian Language Parsing [86]. 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 [89]. 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.

7.2.4.2 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 help 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’ (do) which accounts for large part of conjunct verb expressions. On the other hand verbs like ‘calnaa’ (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 [53, 51] 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 [89]. 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 ‘vishvaasghaat-karna’ (meaning ‘to betray’), the Semantic type of ‘vishvaasghaat’ is “Anti Social”.

**d. Post-Position Indicator (f4):** It 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 then 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 section 7.2.2.

**e. Demonstrative Indicator (f5):** It is a Boolean feature indicating presence of DEM (demonstrative tag) before noun/adjective-verb pair. This diagnostic is explained in section 7.2.2 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 ‘sviikaar’ (to accept) occurs only with two different verbs – ‘kar’ (do) and ‘hat’ (is), and noun ‘kaanuu’ (law) occurs with five different types of verbs – ‘bataa’ (tell), ‘kar’ (do), ‘badal’ (change), ‘lall’ (bring), and ‘padha’ (study). Therefore, ‘sviikaar’ is more likely to form a conjunct verb expression.

**g. Verb Argument Indicator (f7):** This feature computes the average number of post-positions 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.

### 7.2.4.3 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 [146]. Maximum entropy has already been widely used for a variety of natural language tasks, including language modelling [150, 58], text segmentation [18], part-of-speech tagging [147], and prepositional phrase attachment [148].

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 [127, 23]. Some of the features on which training is performed are distinct valued features (f1, f2) while others
are real valued features (f6, f7). These features are mapped to binary features. We have used maximum entropy toolkit\(^1\) to conduct our experiments.

### 7.2.5 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. Table 7.2 gives the overview of useful features which helped in improving the accuracy.

Table 7.2 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 f4 and f5 corresponds to D4 and D5 in section 7.2.2. Addition of feature f6 improved our system by ‘0.54%’ showing dominance of particular objects (as discussed 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 ‘7.84%’ 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+f4+f5) and (f1+f2+f3+f4+f5+f6) as System-1 and System-2 respectively.

<table>
<thead>
<tr>
<th>Feature set</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>f1 + f2</td>
<td>(81.59)</td>
</tr>
<tr>
<td>f1+f2+f3</td>
<td>(82.34)</td>
</tr>
<tr>
<td>f1+f2+f3+f4+f5</td>
<td>(84.74)</td>
</tr>
<tr>
<td>f1+f2+f3+f4+f5+f6</td>
<td>(85.28)</td>
</tr>
<tr>
<td>f1+f2+f3+f4+f5+f7</td>
<td>(77.44)</td>
</tr>
</tbody>
</table>

| **Table 7.2.** Showing System Accuracy with Different Feature Set |

\(^1\) http://homepages.inf.ed.ac.uk/s0450736/maxent\_toolkit.html
7.2.6 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 [71, 131, 12] and hybrid approaches [26] have been tried but still the current state-of-the-art parsing accuracy hasn’t reached to a level which is comparable to English. Complex linguistic phenomena are considered as a vital factor in the low accuracy of Hindi parsing apart from long distance dependencies, non-projective sentences and a smaller corpus size. In the past, various morphological [12], semantic [14] and clause boundary [71] features have been tried as potential language specific features for 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 [12, 13] have pointed out those errors due to complex predicates are significant in Hindi dependency parsing. Recall that in a *conjunct* verb it is the noun/adjective-verb combination 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 a *conjunct* verb is distributed.

In this section, we investigate the effect of using *conjunct* verb specific features on parser accuracy. MST [112, 113] Parser was used to parse sentence, the MaxEnt based tool described in section 7.2.4.3 provides the feature values. An improvement of 0.39% in label and 0.28% in label attachment accuracy is achieved.

7.2.6.1 Experiments and Results

We considered the MST+MaxEnt setting mentioned in [15] as a Baseline for our experiments. All the parsing related experiments are performed on Dataset-2 as described in section 7.2.4.1. Using the output of System-1 and System-2 as described in Section-7.2.5, we added a *conjunct* verb feature for each consecutive noun/adjective-verb pair in the dataset. The feature is added in the feature column of CONLL [49] 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. The 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 results are shown in Table 7.3.

<table>
<thead>
<tr>
<th></th>
<th>LA (%)</th>
<th>UA (%)</th>
<th>L (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>68.77</td>
<td>85.68</td>
<td>71.90</td>
</tr>
<tr>
<td>System 1</td>
<td>69.05</td>
<td>85.68</td>
<td>72.29</td>
</tr>
<tr>
<td>System 2</td>
<td>68.52</td>
<td>85.04</td>
<td>71.93</td>
</tr>
</tbody>
</table>

Table 7.3. Average LA (Labeled Attachment), UA (Unlabeled Attachment) and L (Label) accuracies on 12-fold cross validation
Table 7.4: The 2nd and 3rd columns represent the number of correctly identified ‘pof’ and ‘npof’ labels. Baseline-1 and Baseline-2 give 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 statistics are the summation of 12 testing sets which are tested during 12-fold cross validation.

<table>
<thead>
<tr>
<th></th>
<th>‘pof’ labels</th>
<th>‘npof’ labels</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline-1</td>
<td>715</td>
<td>1628</td>
</tr>
<tr>
<td>System-1</td>
<td>715+36</td>
<td>1628+21</td>
</tr>
<tr>
<td>Baseline-2</td>
<td>713</td>
<td>1630</td>
</tr>
<tr>
<td>System-2</td>
<td>713+42</td>
<td>1630+15</td>
</tr>
</tbody>
</table>

Table 7.4 Summation of 12 testing set tested during 12-fold cross validation

7.2.6.2 Observations

System-1 shows an increase of 0.39% in label and 0.28% in label attachment accuracy, this increase accounts for the 0.3%, 1.87%, 2.94% and 0.43% increase in labels accuracy of ‘k1’, ‘k2’, ‘pof’, ‘k7p’ respectively. These labels occur in the same environment as ‘pof’, hence the confusion. Both System-1 and System-2 help 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 compared to baseline respectively as shown in Table 7.4. 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 compared to both System-1 and Baseline results. This decrease is because of indirect wrong learning leading to ambiguity between different labels.

7.2.7 Statistics of Conjunct Verbs in Hindi Treebank:

We have extracted conjunct verbs from multi-layered/multi-representational Treebank for Hindi [42, 188]. The statistics of the conjunct verbs is given below:

- Total No. of unique conjunct verbs in the Hindi Treebank : 1256 (types)
- Total conjunct verb instances in the Hindi Treebank : 3140 (tokens)

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].
The statistics for some conjunct verbs are given below in the table 7.5:

<table>
<thead>
<tr>
<th>No.</th>
<th>Conjunct Verbs</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>shuruu_kar ‘start’</td>
<td>36</td>
</tr>
<tr>
<td>2.</td>
<td>pesha_kar ‘present’</td>
<td>32</td>
</tr>
<tr>
<td>3.</td>
<td>shuru_ho ‘start’</td>
<td>31</td>
</tr>
<tr>
<td>4.</td>
<td>giraftaar_kar ‘arrest’</td>
<td>26</td>
</tr>
<tr>
<td>5.</td>
<td>vichaar_kar ‘think’</td>
<td>25</td>
</tr>
<tr>
<td>6.</td>
<td>aaropa_lagaa ‘accuse’</td>
<td>24</td>
</tr>
<tr>
<td>7.</td>
<td>faisalaa_kar ‘decide’</td>
<td>23</td>
</tr>
<tr>
<td>8.</td>
<td>koshish_kar ‘try’</td>
<td>21</td>
</tr>
<tr>
<td>9.</td>
<td>ghoshaNaa_kar ‘announce’</td>
<td>19</td>
</tr>
<tr>
<td>10.</td>
<td>istemaal_kar ‘use’</td>
<td>18</td>
</tr>
<tr>
<td>11.</td>
<td>saamanaa_kar ‘confront’</td>
<td>16</td>
</tr>
<tr>
<td>12.</td>
<td>baat_kar ‘talk’</td>
<td>15</td>
</tr>
<tr>
<td>13.</td>
<td>nazar_aa ‘visible’</td>
<td>14</td>
</tr>
<tr>
<td>14.</td>
<td>sambodhit_kar ‘address’</td>
<td>12</td>
</tr>
<tr>
<td>15.</td>
<td>mahasuus_kar ‘feel’</td>
<td>11</td>
</tr>
<tr>
<td>16.</td>
<td>khatma_ho ‘finish’</td>
<td>9</td>
</tr>
<tr>
<td>17.</td>
<td>hissaa_le ‘participate’</td>
<td>8</td>
</tr>
<tr>
<td>18.</td>
<td>ba.Dhaavaa_de ‘encourage’</td>
<td>7</td>
</tr>
<tr>
<td>19.</td>
<td>istifaa_de ‘resign’</td>
<td>6</td>
</tr>
<tr>
<td>20.</td>
<td>tayyaar_ho ‘get ready’</td>
<td>5</td>
</tr>
</tbody>
</table>
Table 7.5. Statistics of Conjunct Verbs in Hindi Treebank

<table>
<thead>
<tr>
<th></th>
<th>Verb</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>21</td>
<td>‘nishaanaa_banaa’</td>
<td>4</td>
</tr>
<tr>
<td>22</td>
<td>‘target’</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>chetaavani_de</td>
<td>3</td>
</tr>
<tr>
<td>24</td>
<td>‘warn’</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>vilaMba_ho</td>
<td>2</td>
</tr>
<tr>
<td>26</td>
<td>‘delay’</td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>tasviira_khiiMcha’</td>
<td>1</td>
</tr>
<tr>
<td>28</td>
<td>‘take picture’</td>
<td></td>
</tr>
</tbody>
</table>

7.2.8 Applying Diagnostics on Noun/Adj+V Combinations:

We have chosen 100 Noun/Adj+V combinations and applied the above diagnostics to them. We applied the diagnostics to the 100 Noun/Adj+V combinations and checked in the corpus if such constructions occurred, i.e., constructions which violate the rule of conjunct verb formation. If such constructions occurred then these Noun/Adj+V combinations were not considered as conjunct verbs and if such constructions didn’t occur then these Noun/Adj+V combinations were considered as conjunct verbs. For example, if we come across a construction like, ‘raam ne us jamhaaii ko liyaa’ ‘Ram took that yawn…..’ violates the rule of conjunct verb formation. If such a construction according to D4 occurs in the corpus then it is not considered as conjunct verb. For ease of exposition, we have taken 25 verbs in the table 5.6 given below and applied the diagnostics. If such constructions occurred then we have marked –ve under that diagnostic corresponding to that Noun/Adj+V combination and if such constructions didn’t occur then we have marked +ve under that diagnostic corresponding to that Noun/Adj+V combination. A Noun/Adj+V combination is accepted as a conjunct verb (indicated by ‘yes’) if >=5 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 >=5 diagnostics are ‘-ve’. If a diagnostic is not applicable for a verb we use a hyphen (‘-’) to indicate this.
<table>
<thead>
<tr>
<th>N/ADJ+V</th>
<th>D1</th>
<th>D2</th>
<th>D3</th>
<th>D4</th>
<th>D5</th>
<th>D6</th>
<th>D7</th>
<th>D8</th>
<th>D9</th>
<th>Conjunct Verb</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) charchaa_kar ‘discuss’</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve</td>
<td>-ve</td>
<td>-ve</td>
<td>Yes</td>
</tr>
<tr>
<td>(2) jaankaarii_de ‘give information’</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve</td>
<td>-ve</td>
<td>Yes</td>
</tr>
<tr>
<td>(3) pragati_ho ‘progress happen’</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve</td>
<td>-ve</td>
<td>-ve</td>
<td>Yes</td>
</tr>
<tr>
<td>(4) vaadaa_kar ‘promise’</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve</td>
<td>-ve</td>
<td>Yes</td>
</tr>
<tr>
<td>(5) haar_maan ‘accept defeat’</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve</td>
<td>-ve</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve</td>
<td>-ve</td>
<td>Yes</td>
</tr>
<tr>
<td>(6) aaropa_lagaa ‘blame’</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve</td>
<td>-ve</td>
<td>+ve</td>
<td>+ve</td>
<td>-ve</td>
<td>Yes</td>
</tr>
<tr>
<td>(7) vishvaas_kar ‘believe’</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve</td>
<td>-ve</td>
<td>Yes</td>
</tr>
<tr>
<td>(8) rakshaa_kar ‘protect’</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve</td>
<td>-ve</td>
<td>Yes</td>
</tr>
<tr>
<td>(9) dastakhat_kar ‘sign’</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve</td>
<td>-ve</td>
<td>Yes</td>
</tr>
<tr>
<td>(10) savaal_kar ‘question’</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve</td>
<td>-ve</td>
<td>Yes</td>
</tr>
<tr>
<td>(11) anurodh_kar ‘request’</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve</td>
<td>-ve</td>
<td>Yes</td>
</tr>
<tr>
<td>(12) hatyaa_kar ‘murder’</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve</td>
<td>-ve</td>
<td>Yes</td>
</tr>
<tr>
<td>(13) kaam_kar ‘work’</td>
<td>-ve</td>
<td>-ve</td>
<td>-ve</td>
<td>-ve</td>
<td>-ve</td>
<td>-ve</td>
<td>-ve</td>
<td>-ve</td>
<td>-ve</td>
<td>No</td>
</tr>
<tr>
<td>(14) chori_kar ‘steal’</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve</td>
<td>-ve</td>
<td>Yes</td>
</tr>
<tr>
<td>(15) upayog_kar ‘use’</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve</td>
<td>-ve</td>
<td>Yes</td>
</tr>
<tr>
<td>(16) taariif_kar ‘praise’</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve</td>
<td>-ve</td>
<td>Yes</td>
</tr>
<tr>
<td>(17) javaab_de ‘answer’</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve</td>
<td>-ve</td>
<td>Yes</td>
</tr>
<tr>
<td>(18) jimmedaarii_le ‘take responsibility’</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve</td>
<td>-ve</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve</td>
<td>-ve</td>
<td>Yes</td>
</tr>
<tr>
<td>(19) pataa_lagaa ‘find_out’</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve</td>
<td>-ve</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve</td>
<td>-ve</td>
<td>Yes</td>
</tr>
<tr>
<td>(20) nazar_aa ‘to be visible’</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve</td>
<td>-ve</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve</td>
<td>+ve</td>
<td>-ve</td>
<td>Yes</td>
</tr>
<tr>
<td>(21) taiyyaar_ho</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In the next chapter we have discussed about the experiments and results that have been done taking the data from Hindi Treebank (HyDT).