Developing Verb Frames for Hindi

Rafiya Begum, Samar Husain, Lakshmi Bai and Dipti Misra Sharma
Language Technologies Research Centre,
IIIT, Hyderabad, India.
{rafiya, samar}@research.iiit.ac.in, {lakshmi, dipti}@iiit.ac.in

Abstract

This paper introduces an ongoing work on developing verb frames for Hindi. Verb frames capture syntactic commonalities of semantically related verbs. The main objective of this work is to create a linguistic resource which will prove to be indispensable for various NLP applications. We also hope this resource to help us better understand Hindi verbs. We motivate the basic verb argument structure using relations as introduced by Panini. We show the methodology used in preparing these frames and the criteria followed for classifying Hindi verbs.

1. Introduction

Verbs are the most important grammatical category in a language. Actions, activities and states are denoted with the help of the verbs. The arguments of the verb specify various participants required by the verb. Verbs play a major role in interpreting the sentence meaning therefore, the study of verb argument structure and their syntactic behavior will provide the necessary knowledge base for intelligent NLP applications.

The relation of the verb with the other components of a sentence in a language can be encoded in different ways. Among them, the word order and the presence of case markers on the arguments are very frequently used by various languages. There are, however, languages in which the marking can be present on the verb itself rather than its arguments (Butt, 2006). Such relations frequently reflect the semantics of the verb, i.e. the syntactic behavior of the verb provides a good handle to understand its semantics. Languages generally also encode other information such as tense, aspect, modality, gender, number, person etc., generally with the verb, allowing for language specific variations.

This paper presents an ongoing effort of developing verb frames for Hindi and classifying them based on their semantic similarity and syntactic behavior. The paper is arranged as follows; In Section 2 we provide the motivation of our work. Section 3 gives a brief overview of the related work. We introduce our approach to Hindi verb classification in Section 4, previous approaches are also discussed in the same section. Section 5 talks about the Paninian grammatical framework. In Section 6 we discuss about the verb frames. Some verb classes are shown in Section 7. Finally, Section 8 concludes the paper.

2. Motivation

The primary motivation for developing frames for Hindi verbs and coming up with their classification is:

- To develop a knowledge base for various NLP applications, e.g. parsers, MT, language generation, etc.
- To create a linguistic resource to help us understand Hindi verbs better.

3. Related Work

Levin's verb classes (Levin, 1993) is an elaborate attempt to investigate English verbs. Drawing from earlier works dedicated to such an investigation, Levin has shown the correlations between the semantic and syntactic behavior of English verbs. **VerbNet (VN)** is a hierarchical, domain-independent, broad-coverage verb lexicon which extends Levin's verb classes (Levin, 1993) and provides the syntactic and semantic information for English verbs. It is an on-line lexicon which has been mapped to other major language resources. VN has more than 5,200 verbs and 237 verb classes (Kipper et al., 2000; Kipper, 2005). **PropBank (PB)** is a corpus, annotated with verbal propositions and their arguments. It has recently been extensively used for the semantic role labeling task (CoNLL shared task 2004-05 [1]). PB adds a layer of semantic annotation atop the syntactic structures. PB represents the verb argument relations by *Arg0, Arg1, Arg2* etc. depending on the verb (Kingsbury et al., 2002). **FrameNet (FN)** is an on-line lexical resource for English, based on frame semantics and supported by corpus evidence. FrameNet groups words according to the conceptual structures i.e. frames that underlie them (Baker et al., 1998).

All these resources have been extensively used for various NLP applications in English and have proved to be very useful in improving the state of the art for many of these applications. However, there have been hardly any attempts for most of the other languages. In this paper we introduce an attempt for the classification of Hindi verbs and developing their verb frames.

4. Hindi Verb Classification

4.1 Earlier Attempts

Earlier attempts on Hindi verb classification have mainly been of the three types. There have been efforts to classify the verbs according to their form. Suraj Bhan Singh (2003) has made a formal classification of Hindi main verbs based on their form and also compared them with English verbs.

[1] *http://www.lsi.upc.edu/~srlconll/*

They are classified into four types:

(a) Simple root (*saral dhaatu)*: These verbs are formed from single words. In Hindi *ubalanaa* 'boil' is an intransitive verb and *ubaalanaa* 'boil' is a transitive verb. English also has these verbs but the form remains same in both the transitive and the intransitive usage.
(b) Composite root (*saamaasik dhaatu*) is formed from two words which are related to each other in meaning and separated by an hyphen, e.g. *padha-likha* 'to become literate'.
(c) Complex verb (*mishra kriyaa*) is formed by combining a noun or an adjective with a verbalizer *kar* or *ho*. For instance, in *taariif karanaa* 'to praise', *taariif* 'praise' is a noun and *karanaa* 'to do' is a verb.
(d) Compound verb (*saMyukta kriyaa*) is formed with two verbs. The first forms the root and the second takes the tense and aspect information. The verb *ro padanaa* 'to start crying' is a compound verb.

This internal form or structure of the verb doesn't show any syntactic and semantic consequences.

The other two approaches deal with the syntactic structures. According to Kachru (1980), in Hindi there are three sets of inherent properties of verbs which have important syntactic consequences. These are:

(a) Stative vs. Inchoative vs. Active
(b) Volitional vs. Non-Volitional
(c) Factive vs. Non-Factive

Stative verbs indicate state of the subject. They are composed of an adjective or past participle and the verb 'be'. *khulaa honaa* 'to be open' is an example of stative verb. **Inchoative** verbs indicate change of state. They are either a simple verb or a complex verb. The complex verbs are composed of a nominal and a verb having the meaning of 'become' or 'come'. *khulanaa* 'to become open' and *yaad aanaa* 'to remember' are examples of inchoative verbs. **Active** verbs indicate actions. They are either causal verbs which are morphologically derived from the intransitive verbs or conjunct verbs composed of a nominal and the verb 'do'. *kholanaa* 'to open' and *yaad karanaa* 'to recall' are examples of active verbs. Accordingly, most intransitive and all dative-subject verbs are either stative or inchoative, and most transitive verbs are active.

Volitional verbs denote deliberate actions. **Non-Volitional** verbs denote states or accidental events. Most active verbs are volitional, whereas most inchoative and stative verbs are non-volitional. Verbs such as *jaananaa* 'to know', *pataa honaa* 'be aware' are **factive.** Verbs like *laganaa* 'feel', *samajhanaa* 'consider' are **non-factive**. The compliments of factive verbs are understood as facts, this is generally not true for non-factives.

Another approach related to syntactic structures is found in **Sahay (2004)** who classifies the Hindi verbs on their *karaka* [2] requirements. He enumerates different constructions that can be formed using karaka relations and classifies the verbs that participate in such constructions. Some of these constructions are:

(a) karta (agent/theme/force) + kriya (verb)
(b) karta + karma (theme) + kriya
(c) karta + adhikarana (location) + kriya
(d) karta + apaadaan (source) + kriya

All the above classification approaches focus on different aspects of the language. Singh focuses on word formation, Kachru on inherent properties of verbs having syntactic consequences, and Sahay, on sentence constructions. While classifying verbs each of these criterions are important. In this paper we present a more holistic approach to classifying Hindi verbs.

4.2 Our Approach

This section talks about our approach to classifying verbs in Hindi.

4.2.1. Initial Approach

We started the classification of Hindi verbs based on extracting the synonyms for a verb from a thesaurus, Brihad Hindi Kosh (Prasad et. al, 1952), and Hindi WordNet (Jha et al., 2001). Using them 100 verb classes were formed. The task of sub-classification was based on the following criteria:

- Frame differs in post-positions only.
- Frame differs in karaka relations.
- Member verbs participate in some other farmes than the class frame.

This initial attempt gave us important insights into the varied properties of Hindi verbs and their correlation to other verbs in the language. However, initial evaluation showed this methodology was very narrow in scope. More specifically, the methodology led to very few verbs in a class. The verbs in a class had very less variations. Analyzing and making generalizations within such a setup was extremely difficult. Nevertheless, such a classification helped us in generating verb frames which have eventually been used in the approach described in Section 4.2.2. The revised approach is much more holistic.

4.2.2. Current Approach

We are currently classifying Hindi verbs and are also providing verb frames using karaka relations. We are referring to Levin's classes as a starting point for our classification. Since verb classes can be identified throughout language and are asserted to exist across languages since their basic meaning components can be applied cross-linguistically (Jackendoff, 1990). Note that we only take the broad semantic property of Levin's classes and not the verbs themselves. We then lookup the Hindi WordNet (Jha et al., 2001) and classification given by Sahay (2004) for identifying various class members. We also refer to the Hindi corpus to get the different syntactic variations of the class members. We are using the

2 *'karaka'* are relations defined by Panini for his grammar of Sanskrit. For a more detailed discussion see Bharati et al. (1995) and Begum et al. (2008).

following four criterions for classifying the Hindi verbs:

(a) Basic Semantics
(b) Semantic Sub-classification (if any)
(c) Morphological Relatedness
(d) Syntactic Behaviour and Verb Frames

(a) *Basic Semantics*: Verbs are initially grouped together according to some basic semantic similarity. For instance verbs such as *mil* 'to meet', and *laDa* 'to fight' have similar basic semantics, in that they signify group activities i.e. they require more than one participant. All such verbs are grouped together in a single class. (b) *Semantic Sub-classification*: These verbs may again be sub-classified within a class based on finer semantics, if there exists any such distinction. For instance, verbs relating to *eating* can be further sub-classified into simple eating verbs, verbs showing manner of eating and verbs relating to speediness while eating. (c) *Morphological Relatedness*: The morphological criterion looks for the possibility of deriving possible verb forms from the base verb of the class. For instance, intransitive verbs can have causative forms derived from them and transitive verbs can have intransitive and causative forms derived from them. Hindi verbs show the following morphological relatedness:

- Basic transitives which can have causative forms.

Transitive	**Causative-1**	**Causative-2**
khaa	*khilaa*	*khilavaa*
'to eat'	'to make to eat'	'to make to eat'

- Basic intransitives which can have transitive or causative forms.

Intransitive	**Causative-1**	**Causative-2**
daud	*daudaa*	*daudavaa*
'to earun'	'to make to run'	'to make to run'

- Basic transitives which can have intransitive forms. They are of two types:

  (i) intransitive form is derived from a transitive verb. This intransitive form takes a dative subject.

(1)*raam ko caand dikhaa*
'Ram' 'dat.' 'moon' 'to be seen'
'The moon was seen to Ram.'

Transitive	**Intransitive**	**Causative-1**	**Causative-2**
dekh	*dikh*	*dikhaa*	*dikhavaa*
'to see'	'to be seen'	'to show'	'to cause to show'

(ii)The intransitive form derived from a transitive verb implies the existence of an agent though there is no agent expressed in the sentence.

(2)*kapade dhul gaye*
'clothes' 'wash' 'have been'
'The clothes have been washed'

Transitive	**Intransitive**	**Causative-1**	**Causative-2**
dho	*dhul*	*dhulaa*	*dhulavaa*
'to wash'	'to be washed'	'to make to wash'	'to make to wash'

In (i) the subject of transitive and intransitive verb (dative subject) is the same whereas in (ii) the object of transitive is the subject of the intransitive verb.
Morphology of the verbs have significant syntactic consequences. The syntactic behaviour and a verb frame of an intransitive verb will vary from the transitive verb derived from it. In our approach morphology of a verb plays a major role in capturing the syntactic consequences. (d) *Syntactic Behavior*: Finally, the verbs are grouped based on their syntactic behavior. The syntactic behavior is decided based on the syntactic alternations for each verb. For each syntactic alternation the verb frame is formed. Thus, the class of verbs in this classification would share all the four criterion mentioned above.

5. Paninian Grammatical Framework

As mentioned earlier, we capture verb argument relations using the Paninian approach. The Paninian approach treats a sentence as a series of modifier-modified relations. A sentence is supposed to have a primary modified which is generally the main verb of the sentence. The elements modifying the verb participate in the action specified by the verb. The participant relations with the verb are called *karaka*, (Begum et al., 2008).

The notion of *karaka* relations is central to the Paninian framework. The *karaka* relations are syntactico-semantic relations between the verb and the other constituents of the sentence. They capture a certain level of semantics. The approach uses case markers (vibhakti information) for mapping the relation between the verb and its arguments. The six basic *karakas* are: (note that the English translations are only approximations and don't fully capture the concepts below)

(1) *karta*	(k1)	'agent/theme/force'
(2) *karma*	(k2)	'theme'
(3) *karana*	(k3)	'instrument'
(4) *sampradaan*	(k4)	'recipient'
(5) *apaadaan*	(k5)	'source'
(6) *adhikarana*	(k7p)	'location'

We must note here that although one can roughly map the last four *karakas* to their thematic role counterpart, *karma* and *karta* are different from 'theme' and 'agent' (although they might map with them sometimes). The reason for this divergence in the two notions (*karaka* and thematic role) is due to the difference in what they convey. Thematic role is purely semantic in nature whereas the *karaka* is syntactico-semantic, see Bharati et al. (1995), for a more detailed discussion).
Another important aspect of this approach is, that it considers the semantics of the verb for assigning *karta* and *karma* karakas. The semantic model of the Paninian

framework has a verbal root which denotes an action. Verbal root consists of two elements, *activity* and *result*. An activity denotes the actions of the various participants or *karakas* involved in the action and the result is the state which when reached, the action is complete. In this framework an action is usually complex as it is broken into sub-actions, (Bharati et al., 1995).

6. Verb Frames

The verb frames developed following this framework show the mandatory *karaka* relations for a verb. Each verb can have multiple senses and for each sense of a verb there can be a number of possible frames.

The following three resources have been primarily used for developing verb frames:

- Levin's verb classes
- A Hindi corpus[3]
- HWN (Jha et al., 2001)
- Sahay's verb classes

Verb::*aa*
SID::aa%VI%1
Verb_Sense::HWN(1,10,12)
Eng_Gloss ::to come
Example::*raam hyderabad aataa hai*
Ram Hyderabad come be-Pres
'Ram comes to Hyderabad'
Theta_Roles::AGENT DESTINATION V

Frame1::

arc-label	necessity	vibhakti	lexical-type
k1	mandatory	0	noun
k2	desirable	0|*para*	noun

The corpus is consulted to get the syntactic distribution in which the verb occurs and the HWN is referred to get the required sense information.

Given below is an example of a verb entry along with the verb frame:

Figure 5: Verb Frame for verb *aa* 'to come'

The following information is given for each verb entry:

(a) *Description of the verb*
(b) *Verb Frame*

(a) *Description of the verb:* In the description, we give the following information; *name of the verb, its sense id* (SID, an id is given according to the number of senses a verb has), *HWN sense id, English gloss, example sentence of the verb, theta roles and the* ***verb frame*** *(given in a tabular form).* In the figure 5 given above the verb is *aa* 'to come'. *SID* stands for sense id and it is represented as *aa%VI%1*. In *SID* we are capturing the name of the verb, the type of the verb and the sense number, all three separated by a percentage symbol. *aa* 'to come' is the verb, the type of the verb is *VI* which means verb intransitive and *1* is the sense number. *Eng_Gloss* stands for English gloss. Here 'to come' is the gloss of the verb *aa. Example* contains the Hindi example sentence containing the verb.

(b) *Verb Frame:* Verb frame is represented in a tabular form. A **verb frame** shows**:**

- *karaka* relations
- necessity of the argument i.e whether it is mandatory (m) or desirable (d).
- *vibhakti* (postpositions taken by the arguments)
- lexical category of the arguments.

In the figure we see that karaka relations for verb aa 'to come' is given. The arguments of the verb raam 'Ram' and hyderabad 'Hyderabad' are karta (k1) and karma (k2) respectively. The necessity of k1 (raam) and k2 (hyderabad) is mandatory and desirable respectively. k1 takes 0 vibhakti and k2 can take either 0 or para depending upon its selectional restrictions. The vibhakti of the arguments depends upon the TAM (tense, aspect amd modality). The lexical category of both the arguments is noun.

The frames are developed based on simple present tense and indicate habitual acts taking it as default. In fact, *karaka* relations and the postpositions in the frame reflect the behavior of the verb when it occurs in simple present ('*taa hai*' in hindi, eg. *khataa hai* 'eats'). This is done to bring in consistency while forming the various frames, in Hindi the postposition of an argument might change with the change in the TAM (tense, aspect and modality) information of the verb. These changes in the vibhaktis are not syntactic alternations but are transformations due to the change in the default TAM.

It is clear that the entire structure just discussed is very rich. As of now we plan to exploit the frames and the verb classes (section 7) in parsing. They can also be used for various other applications which require a knowledge base, e.g. word sense disambiguation, Machine translation, etc.

7. Verb Classes

A few verb classes are discussed below to illustrate the entire classification approach and resultant verb frames for each class.

(1)Verbs of Social Interaction

Semantics:
These verbs signify group activities. This class includes a significant number of verbs relating to 'fighting' and 'verbal interactions'. If the subject of these verbs is a collective noun then it doesn't take a second participant. On the other hand, when the subject is a singular noun then the verb takes a second participant with a *se* vibhakti

3 We use the CIIL (Central Institute for Indian languages) corpus.

(postposition). All the participants of these verbs should have the same status. Both should have the capability to initiate and carry out the action.

Class Members:
(1) *yuddha kar* 'to fight a battle', (2) *lad* 'to fight', (3) *jhagad* 'to quarrel', (4) *ladaai kar* 'to have a fight', (5) *jhagadaa kar* 'to have a quarrel', (6) *baaten kar* 'to talk', (7) *bahas kar* 'to argue', (8) *vaad vivaad kar* 'to debate', (9) *caracaa kar* 'to discuss', (10) *mil* 'to meet', (12) *mulaakaat kar* 'to meet', (13) *khel* 'to play', (14) *prem kar* 'to love', (15) *shaadi kar* 'to marry', (16) *gale lag* 'to hug', (17) *talaakha le* 'to divorce', (18) *sahamat ho* 'to agree', (19) *raazii ho* 'to agree', (20) *asahamat ho* 'to disagree', (21) *samajhautaa kar* 'to negotiate', (22) *mazaak kar* 'to joke', (23) *dillagii kar* 'to banter'

Morphology:
The morphological criterion works by looking at various word forms that can be derived from the base verb:

Intransitive	**Causative**
lad	*ladaa/ladavaa*
'to fight'	'to cause to fight'
mila	*milaa/milavaa*
'to meet'	'to cause to meet'
khel	*khilaa/khilavaa*
'to play'	'to cause to play'

Notice that the verb frames not only capture the arguments of the base verb but also that of the causatives which, in Hindi, are derived through a morphological process.

Syntactic Behaviour and Verb Frames:

(3) *raam sitaa se milataa hai*
'Ram' 'Sita' 'with' 'meet' 'is'
'Ram meets with Sita.'

(4) *raam aur siitaa milate hai*
'Ram' 'and' 'Sita' ' meet' 'is'
'Ram and Sita meet.

(5) *ve milate hai*
'they' 'meet' 'is'
'They meet'

In (3) *raam* 'Ram' is the active participant whereas *sitaa* 'Sita' is passive. However, there is no such difference in (4). In (5) the subject is a collective noun, if we replace this by a singular noun, the sentence becomes ungrammatical. We can add *ek duusare se* 'with each other' and *aapasa me* 'among them', in (4) and (5). Example (3) is captured in Verb Frame-1 below. In (3) *raam* 'Ram' is *karta* (k1) whereas *siita* 'Sita' is an associative participant (ras, 'r[elation] as[sociative]') in the action. Both are mandatory arguments. *raam* takes *0* (null) vibhakti and *siita* takes *se* vibhakti. Both the participants are nouns. Examples (4) and (5), are captured by Verb Frame-2. *raam aur siita* 'Ram and Sita' in (4) and *ve* 'they' in (5) are collective nouns. Both are *karta (k1)* and take *0* (null) vibhakti. Note that Frame-2 shows that the argument of the verb is plural (n{pl}).

Verb Frame-1:

arc-label	necessity	vibhakti	lexical-type
k1	m	0	n
ras	m	se	n

Verb Frame-2:

arc-label	necessity	vibhakti	lextype
k1	m	0	n{pl}

The above frames are for the base verb *mila* 'meet'. The same verb frames applies for the other class members also. We also have the causative form of these verbs. These verbs are captured in a separate verb frame. There are two verb frames for causatives too.

(6)*mohan raam ko apne dost se*
'Mohan' 'Ram' '*ACC*' 'his' 'friend' 'with'
ladaataa hai
'cause to fight' '*is*'
'Mohan makes Ram to fight with his friend'

(7)*mohan raam aur shyaam ko ladaataa hai*
'Mohan' 'Ram' 'and' 'Shyam''*ACC*''cause to fight' 'is'
'Mohan causes Ram and Shyam to fight'.

(8) *mohan unko ladaataa hai*
'Mohan' 'them' 'cause to fight' 'is'
'Mohan causes them to fight'.

In (6), (7) and (8) the causative form *ladaa* 'to cause to fight' of the base verb *lad* 'to fight' is used. It is intransitive. Example (6) is captured in Verb Frame-3. In (6) *mohan* 'Mohan' is the *prayojak karta (pk1)* i.e. causer and *raam* 'Ram' is *prayojya karta (jk1)* i.e. causee. *mohan* 'Mohan' takes a null vibhakti, *ram* 'Ram' takes a *ko* vibhakti and *apne dost* 'his friends' takes *se* vibhakti. Examples (7) and (8) are captured in Verb Frame-4. *mohan* 'Mohan' in (7) and (8) is *pk1* and *mohan aur shyaam* 'Mohan and Shyam' in (7), and *unko* 'them' in (8) are *jk1*.

Verb Frame-3:

arc-label	necessity	vibhakti	lexical-type

pk1	m	0	n
jk1	m	ko	n
k2	m	se	n

Verb Frame-4:

arc-label	necessity	vibhakti	lexical-type
pk1	m	0	n
jk1	m	ko	n

A verb class can also have conjunct verbs as its member. A frame for a conjunct verb would differ from a normal frame. There are still some unresolved issues with the representation of their frame. We hope to solve the issues soon.

(2)Learn Verbs:

Semantics:
These verbs refer to the action of acquiring or learning information.

Class-Members:
(1) *siikha* 'learn', (2) *gyaan praapta kar* 'acquire', (3) *raTa* 'cram', (4) *yaad kar* 'memorize', (5) *padh* 'read, study'.

Morphology:

Transitive	**Causative-1**	**Causative-2**
siikha 'to learn'	*sikhaa* 'to cause to learn'	*sikhavaa* 'to cause to learn'
padh 'to read'	*padhaa* 'to cause to read'	*padhavaa* 'to cause to read'
raTa 'to cram'	**raTaa* 'to cause to cram'	**raTavaa* 'to cause to cram'

Syntactic Behaviour and Verb Frame:

(9) *siita ganita padhati hai.*
'Sita' 'maths' 'study' 'is'
'Sita studies Maths.'

(10) *vaha siita ko ganita padhaati hai.*
'she' 'Sita' '*ACC*' 'maths' 'teach' 'is'
'She teaches Maths to Sita.'

(11) *vaha raam se siita ko ganita padhavaati hai.*
'she' 'Ram' 'caus.''Sita' '*ACC*' 'maths' 'cause to teach' 'is'
'She makes Ram to teach Maths to Sita.'

Verb Frames for (9),(10) and (11) are denoted by Verb Frame 5, 6 and 7 respectively. In (9) the verb *padh* 'to learn' is a basic transitive verb. The arguments *sita* 'Sita' is *karta* (k1) and *ganita* 'maths' is *karma* (k2). Both have *null* vibhaktis. In (10), the same verb is in causative form. So *sita* 'she' which is the *karta* (k1) in (9), becomes *prayoja karta* in (10). *ganita* 'maths' has the same *karaka* role as (9). Finally, *vaha* 'she' is *prayojak karta* (pk1). In (11) the same verb has been used as double causative. All the arguments have the same *karaka* relations as (10) except an extra argument *raam* 'Ram' which is *madhyasta karta* (mk1) i.e. causer2. When there is a *madhyasta karta* in a sentence then the *prayojyak karta* (pk1) becomes causer1 and *madhyasta karta* (mk1) becomes causer2. *raam* 'Ram' which is a *madhyasta karta* (mk1) takes a *se* vibhakti, whereas all the other arguments take the same vibhakti as (10).

Verb Frame-5:

arc-label	necessity	vibhakti	lexical-type
k1	m	0	n
k2	m	0	n

Verb Frame-6:

arc-label	necessity	vibhakti	lexical-type
pk1	m	0	n
jk1	m	ko	n
k2	m	0	n

Verb Frame-7:

arc-label	necessity	vibhakti	lexical-type
pk1	m	0	n
mk1	m	se	n
jk1	m	ko	n
k2	m	0	n

(3)Lodge Verbs:

Semantics:
These verbs refer to the types of living. They communicate various actions of living.

Class-Members:
(1) *Thahar* 'lodge, board', (2) *ruk* 'lodge', (3) *tik* 'lodge', (4) *rah* 'live, dwell, reside' (5) *nivaas kar* 'live', (6) *bas* 'settle', (7) *sharan le* 'shelter'

Morphology:

Intransitive	**Causative-1**	**Causative-2**

Thahar	*Taharaa*	*Taharavaa*
'to stay'	'to make to stay'	'to make to stay'
ruk	*rukaa*	*rukavaa*
'to stay'	'to make to stay'	'to make to stay'
tik	*tikaa*	*tikavaa*
'to stay'	'to make to stay'	'to make to stay'
rah	**rahaa*	**rahavaa*
'to live'	'to cause to live'	'to cause to live'

Syntactic Behaviour and Verb Frame :

(12) *vaha raam ke ghar par/me Thaharataa hai.*
'he' 'Ram' 'gen.' 'house' 'on/in' 'stay' 'is'
'He stays at Ram's house'.

(13) *vaha mohan ko raam ke ghar par/me*
'he' 'Mohan' '*ACC*' 'Ram' '*GEN*' 'house' 'on/in'
Thaharaataa hai.
'to cause to stay' 'is'
'He makes Mohan to stay at Ram's house'.

(14) *vaha mohan se raam ko hotel par/me*
'he' 'Mohan' 'caus.' 'Ram' '*ACC*' 'hotel' 'on/in'
Thaharavaataa hai.
'to cause to stay' 'is'
'He causes Mohan to make Ram to stay at a hotel'.

Verb Frame-8:

arc-label	necessity	vibhakti	lexical-type
k1	m	0	n
k7p	m	par|me	n

Verb Frame-9:

arc-label	necessity	vibhakti	lexical-type
pk1	m	0	n
jk1	m	ko	n
k7p	m	par|me	n

Verb Frame-10:

arc-label	necessity	vibhakti	lexical-type
pk1	m	0	n
mk1	m	se	n
jk1	m	ko	n
k7p	m	par|me	n

Verb Frames for (12), (13) and (14) are denoted by Verb Frames 8, 9 and 10 respectively. In (12) *vaha* 'he' and *ghar par/me* 'house at/in' are *karta* (k1) and *adhikaran* (k7p) respectively. Verb Frame 9 and 10 are for the causative form of the base verb *Thahara* 'stay' given in (12).

(4)Verbs of Ingesting

Semantics:
These verbs relate to the ingestion of food or drink.

Semantic Sub-Classes:

(i)Eat Verbs:

Semantics:
These verbs are the simple verbs of ingesting. *khaa* 'eat' refers to ingesting solids and *pI* 'drink' refers to ingesting liquids. They don't specify the manner of ingesting or the meal involved.

Class-Members:
(1) *pii* 'drink', (2) *khaa* 'eat'

Morphology:

Transitive	**Causative-1**	**Causative-2**
khaa	*khilaa*	*khilavaa*
'to eat'	'to make to eat'	'to make to eat'
pii	*pilaa*	*pilavaa*
'to drink'	'to make to drink'	'to make to drink'

Syntactic Behaviour and Verb Frame:

(15) *bacaa seb khaataa hai*
'child' 'apple' 'eat' 'is'
'The child eats apple'.

(16) *siita bacce ko seb khilaati hai*
'sita' 'child 'acc.' 'apple' 'to cause to eat' 'is'
'Sita fed an apple to the child'.

(17) *siita aayaa se bacce ko seb khilvaati*
'sita' 'maid' 'caus.' 'child''acc.''apple''to cause to eat'
hai
'is'
'Sita makes maid to feed an apple to the child'.

Verb Frames for (15), (16) and (17) are denoted by Verb Frames 5, 6 and 7 respectively given above. In (15) *baccaa* 'child' and *seb* 'apple' is *k1* and *k2* respectively. Example (16) and (17) are the causative form of the verb *khaa* 'eat' given in (15).

(ii)Chew Verbs:

Semantics:
They refer to the manner of ingesting.

Class-Members:
(1) *cabaa* 'chew', (2) *kutar* 'gnaw, nibble', (3) *caat* 'lick', (4) *cuga* 'pick, peck', (5) *sip* 'ghuunt le', (6) *sudak* 'slurp', (7) *cuus* 'suck'

Syntactic Behaviour and Verb Frame:

(18) *billii doodh caatati hai*
'cat' 'milk' 'lick' 'is'
'Cat licks milk'.

(iii)Gulp Verbs:

Semantics:
These verbs refer to the complete, and usually speedy, consumption of something.

Class-Members:
(1) *bhakos* 'gobble', (2) *gatak* 'gulp', (3) *nigal* 'swallow'

Syntactic Behaviour and Verb Frame:

(l9) *vaha goliyaa nigalataa hai*
'he' 'tablet' 'swallow' 'is'
'He swallows tablets.'

8. Conclusion and Future Directions

This paper introduces a more holistic approach to Hindi verb classification. The approach considers the basic semantics, morphology and syntactic alternations of a verb to classify it into a class. We hope that such an initiative will prove to be beneficial for various NLP applications. Around 300 verb frames which show the basic argument structure of these verbs have been prepared. We are in the process of classifying them into different classes. We briefly discussed some of these classes in this paper.

9. References

Baker, C. F., Fillmore, C. J., and Lowe, J. B. (1998). The Berkeley FrameNet project. *In Proceedings of the COLING-ACL*, Montreal, Canada.

Begum, R., Husain, S., Dhwaj, A., Sharma, D.M., Bai, L., and Sangal, R. (2008). Dependency Annotation Scheme for Indian Languages. In *Proceedings of The Third International Joint Conference on Natural Language Processing (IJCNLP)*, Hyderabad, India.

Bharati, A., Chaitanya, V., and Sangal, R. (1995). *Natural Language Processing: A Paninian Perspective*, Prentice-Hall of India, New Delhi, pp. 65-106.

Butt., M. (2004(. The Light Verb Jungle. In G. Aygen, C. Bowern & C. Quinn eds. *Papers from the GSAS/Dudley House Workshop on Light Verbs.* Cambridge, Harvard Working Papers in Linguistics, p. 1-50.

Greaves, E. (1983). *Hindi Grammar*. Asian Educa-tional Services, New Delhi, pp. 335-340

Jackendoff, R. (1990). *Semantic Structures*. MIT Press, Cambridge, Massachusetts.

Jha, S., Narayan, D., Pande, P., Bhattacharyya, P., A WordNet for Hindi. (2001). *International Workshop on Lexical Resources in Natural Language Processing*, Hyderabad, India, January 2001.

Kachru, Y. (*ed.*) (1980). *Aspects of Hindi Grammar*. South Asia Books.

Kellogg, S. H. (1972). *A Grammar of the Hindi Language.* Munshiram Manoharlal, New Delhi, pp. 271-279.

Kingsbury, P. and Palmer, M. (2002(. From Treebank to PropBank. In *Proceedings of the 3rd International Conference on Language Resources and Evaluation*, Las Palmas, Canary Islands, Spain.

Kipper, K. (2005). *VerbNet: A broad-coverage, comprehensive verb lexicon.* Ph.D. thesis, Computer and Information Science Dept., University of Pennsylvania, Philadelphia, PA.

Kipper, K., Dang, H. T., and Palmer, M. (2000). Class-based construction of a verb lexicon. In *AAAI/IAAI*, pages 691-696.

Levin, B. (1993). English verb classes and alternations, A Preliminary Investigation. The University of Chicago Press.

Mohanan, T. (1994). *Arguments in Hindi*. CSLI Publications.

Prasad, K. et al. (1952). *Brihad Hindi Kosh*, Varanasi Jnanamandal

Singh, S. B. (2003). *English-Hindi Translation Grammar,* Prabhat Prakashan, Delhi, pp. 167-186.