A COMPARATIVE STUDY OF ENGLISH AND HINDI VERB SEMANTICS

by

Kumar Sanjeev, Rafiya Begum, Radhika Mamidi

in

LSI Platinum Jubilee Conference

Report No: IIIT/TR/2005/29

Centre for Language Technologies Research Centre
International Institute of Information Technology
Hyderabad - 500 032, INDIA
December 2005
A COMPARATIVE STUDY OF ENGLISH AND HINDI VERB SEMANTICS

Kumar Sanjeev, Rafiya Begum and Radhika Mamidi
Language Technologies Research Centre
International Institute of Information Technology
Gachibowli, Hyderabad. India.
{sanjeev_kumar,rafiya}@students.iiit.net, radhika_m@iiit.net

Abstract

This paper follows the prevailing assumption in current approaches that the rules linking arguments to syntactic positions are universal. By this, we mean that verbs with similar lexical semantic features across languages map onto a similar set of basic syntactic structures. However, recent works on cross-linguistic study of verb frames have shown different syntactic structures for the verbs with same lexical semantics [Levin 1998]. Our study explores syntactic and semantic variation while mapping English verbs onto Hindi verbs. We will explore the cause for such variations by looking at the semantic complexities of the verbs. This preliminary study will help us in structuring lexicon. It will also help in text information processing tasks like machine translation, information extraction, etc.

1. Introduction

Verbs are important grammatical category as they form the pivot of the sentence. A sentence without a verb is unthinkable. The knowledge of a speaker about a lexical item suggests that he or she has more lexical knowledge than the knowledge of word-specific idiosyncratic properties. All the grammatical information with respect to tense, aspect, modality, number, gender, etc is carried by the verbs in a sentence. These are the reasons why the study of verbs has acquired immense importance in Linguistics and NLP.

The verb is the binding agent in a sentence. The noun phrases in a clause link to the main verb of the clause according to the verb’s selection preferences. Look at the following examples:
In the above examples\(^1\) we have taken English verb ‘give’ and Hindi verb ‘dena’ in 1(a) and 1(b) respectively. The verbs ‘give’ and ‘dena’ are equivalent and their frames are also same. Only the word order is different between Hindi (SOV) and English (SVO).

\(^1\) Abbreviations used:

Abl-Ablative; Acc-Accusative; Adj-Adjective; Adv-Adverb; Dat-Dative; Erg-Ergative; Fem-Feminine; Gen-Genitive; Inst-Instrument; Loc-Location; Msc-Masculine; Prep-Preposition; Perf-Perfective; PROG-Progressive; Sg-Singular; V-Verb; TAM-tense,aspect and modality.
have used two resources: Levin’s verb classes and VerbNet. These two resources deal with English verbs. But, we think that these two resources can be helpful in the study of Hindi verb syntax and semantics.

2. Resources

The main aim of our study of verb semantics and syntax is to contribute to structure the lexicon and to allow for a better organized, more homogeneous description of the semantics. From a more formal point of view, the main aims are:

• the identification of meaning components forming the semantics of verbs
• the specification of more subtle meaning elements that differentiate closely related verbs and
• the study of the cooperation between syntax and semantics.

2.1 Levin’s Verb Semantic Classes and Syntactic Alternations

Levin [1993] shows the correlations between the semantics of verbs and their syntactic behavior for a large set of English verbs (about 3200). More precisely, she shows that some facets of the semantics of verbs have strong correlations with the syntactic behavior of these verbs and with the interpretation of their arguments.

Levin\(^2\) first precisely delimits the different forms of verb syntactic behavior. Each of these forms is described by one or more alternation (e.g. alternations describe passive forms, there-insertions and reflexive forms). Then, she proposes an analysis of English verbs according to these alternations: each verb is associated with the set of alternations it undergoes. A preliminary investigation showed that there are sufficient correlations between some facets of the semantics of verbs and their syntactic behavior to allow for the formation of classes. From these observations, Levin has then defined about 200 verb

\(^2\) http://www.ilc.cnr.it/EAGLES96/rep2/node10.html
semantic classes, where, in each class, verbs share a certain number of alternations.

Examples of some of the alternations are given below:

a. **Middle alternation:**
   
   *The butcher cuts the meat. / The meat cuts easily.*
   
   Generic interpretation, understood but unexpressed agent.

b. **Causative/Inchoative alternations:**
   
   *Janet broke the cup. / The cup broke.*
   
   *Tim dropped the ball. / The ball dropped.*
   
   Also known as anti-causative, ergative. It affects verbs of change of state or change of position.

c. **Induced Action alternations:**
   
   *The horse jumped (over the fence)/ Sylvia jumped the horse over the fence.*
   
   With some movement verbs, directional verb is necessary.

d. **Other causative alternations:**
   
   *The baby burped. / The father burped the baby.*
   
   Basically intransitive verbs, limited range of objects (*burp the doctor*).

e. **Substance/Source alternations:**
   
   *Heat radiates from the sun. / The sun radiates heat.*
   
   Verbs of emission

### 2.2 VerbNet

VerbNet\(^3\) is a verb lexicon with syntactic and semantic information for the verbs referring to Levin’s verb classes for systematic construction of lexical entries. The fundamental assumption is that the syntactic frame of a verb as an argument taking element is a direct reflection of the underlying semantics. Therefore, verbs in the same verb class share common syntactic frames, syntactic behaviors, semantic elements and thematic roles. For example, *bang, bash, batter, beat, bump, butt, dash, drum and hammer* belonging to the class ‘hit’ have the same frames.

\(^3\) [http://www.cis.upenn.edu/~mpalmer/project_pages/VerbNet.htm](http://www.cis.upenn.edu/~mpalmer/project_pages/VerbNet.htm)
The entry for each verb in a class is associated with the most suitable WordNet sense(s), with the same verb in a different class typically receiving a different WordNet assignment. The argument list of each entry consists of thematic labels and possible selection restrictions on the arguments. The syntactic information in each verb’s entry maps the list of thematic arguments to the deep-syntactic arguments of that verb. The semantic predicates list the participants during various stages of the event described by the syntactic frame. The argument list of the lexicon consists of a set of 21 thematic roles used to map verb arguments for all classes. Each verb argument is assigned one unique thematic role within the class. The thematic roles are defined to all verbs in a class, so verbs present in more than one class may have different roles.

Examples of syntactic frames of verb ‘hit’

Agent V Patient (e.g. John hit the ball)
Agent V at Patient (e.g. John hit at the window)
Agent V Patient [+plural] together (e.g. John hit the sticks together)

3. Data and Analysis

The example sentences we have taken demonstrate the systematic differences between Hindi and English in the flexibility with which verbs can occur in syntactic frames. We divide our data into three groups:

- Semantic domains following Talmy
- Alternations following Levin
- Others

3.1. Semantic domains following Talmy:
Talmy’s (1985) study presents a range of cross linguistic data showing systematic differences between English-type languages and Spanish-type languages. Talmy suggests that the typological differences arise from differences among languages in their "lexicalization patterns." As Hindi belongs to the Spanish-type, we extend his observations on motion verbs for our purpose.
Talmy (1985, 1991) notes that languages differ in the surface structural realization of the motion event.

Following Talmy’s hypothesis, Narsimhan [1993] presented evidence showing that there are no differences between Hindi and English in patterns of lexicalization of meaning in the verb, or in the lexical aspect encoded by the verbs. In simpler words, Narsimhan refutes Talmy’s hypothesis that there are lexical gaps. Moreover, she suggests that the differences between Hindi and English can be attributed to the role that constructions play in the mapping between verbs and syntactic frames.

The semantic domains we have chosen are: *Spontaneous motion, Caused motion, Spontaneous change of state, Caused change of state, Spontaneous creation, Caused creation*. The examples taken are from Narsimhan [1993].

(i) **Spontaneous motion**

4. (a) **English**: The bird fluttered to the tree.
   Agent V Goal
   (b) **Hindi**: *pakshii peDa-ko phaDaphaDaayaa.
      Agent Goal V
      Bird tree-Dat fluttered-Sg.Masc.Perf
   (c) **Hindi**: pakshii phaDaphaDaa-kar peDa-par gayaa.
      Agent Adv (manner) Goal V
      Bird flutter-by tree-Dat go-Sg.Masc.Perf
      The bird went to the tree, fluttering.

(ii) **Caused motion**

5. (a) **English**: I shook the powder out of the box.
   Agent V Theme Source
   (b) **Hindi**: *mai-ne pauDar-ko Dibbe-me-se hilaayaa.
      Agent Theme Source V
      I-Erg powder-Acc box-Source-Abl shook

6. (a) **English**: I got the powder out of the box by shaking.
   Agent Theme Loc Adv (manner)
   (b) **Hindi**: mai-ne pauDar-ko Dibbe-me-se hilaakar giraayaa.
      I-Erg lid-Acc box-Loc Abl shake-by removed
      I removed the powder from the box by shaking.

(iii) **Spontaneous change of state**

7. (a) **English**: The door slid shut.
   Theme V Adj.
   (b) **Hindi**: *darvaazaa band sarakaak
      Theme Adj V
8. (a) English: The door got shut by sliding.
(b) Hindi: darvaaza sarak-kar band huaa.
   Agent Adv (manner) V1 V2
Door slide-by shut became.
The door became shut by sliding.

(iv) Caused change of state
9. (a) English: I beat the dogs silent.
   Agent V Theme Adj.
(b) Hindi: *mai-ne kuttoN-ko khaamosh maaraa.
   Agent Patient Adj. V
   *I-Erg dogs-Acc silent beat.
10. (a) English: I silenced the dogs by beating them.
    Agent V Theme Adv (manner)
(b) Hindi: mai-ne kuttoN-ko maar-kar khaamosh kiyaa.
    Agent Theme Adv (manner) Adj V
    I-Erg dogs-Acc beat-by silent did.
    I made the dogs silent by beating them

(v) Spontaneous creation (example from Talmy, 1985)
11. (a) English: A hole burned in his shirt.
    Result V Location
(b) Hindi: *usa-ke kamiiz-me ek ched jalaa.
    Location Result V
    *He-Gen shirt-Loc one hole burned.
12. (a) English: A hole formed in his shirt by burning.
    (b) Hindi: usa-ke kamiiz-me jaane-se ek ched ho gayaa.
    He-Gen shirt-Loc burning-cause one hole happened
    A hole formed in his shirt by burning.

(vi) Caused creation
13. (a) English: Raju ironed a crease in his shirt.
    Agent V Theme Location
(b) Hindi: *raaju-ne apnii kamiiz-me chunat istrii kii.
    Agent Location Theme V
    Raju-Erg self shirt-Loc crease iron did.
14. (a) English: Raju created a crease in his shirt by ironing.
    (b) Hindi: raaju-ne apnii kamiiz-me istrii kar-ke chunat banaaii.
    Raaju-Erg self shirt-Loc iron do-by crease made.
    Raju made a crease in his shirt by ironing (it).

(vii) Communication (example from Levin & Rapport, 1988)
15. (a) English: She smiled a welcome.
    Agent V Theme
(b) Hindi: *usa-ne swaagat muskaraayaa.
    Agent Theme V
    *She-Erg welcome smiled.
16. (a) English: She expressed welcome by smiling.
In Hindi, the so-called manner verb (“flutter”, “slide”, “iron”, “smile”) becomes an adverb and we can only insert a vanilla verb (such as “go”, “become/get”, “create”, “convey”) as the syntactic head of the sentence.

In languages such as Hindi, there is a restriction on the use of telic path phrases with manner verbs. Verbs of manner of motion can occur with atelic path phrases but not with telic path phrases:

The above examples from different semantic domains demonstrate the systematic differences between Hindi and English in the flexibility with which verbs can occur in syntactic frames.

### 3.2. Alternations following Levin

We also find that Hindi verbs are much more restricted in their ability to appear in some of Levin’s alternations also. These alternations, which involve an extension of their basic meaning and argument structure, vary from their English counterparts.

We will discuss three such cases in this section, the swarm alternation (Levin, 1993:54), the material product alternation (Levin, 1993:56) and Understood body-part object alternation [Levin, 1993:34 ].

We found that, as expected, Hindi verbs are much more restricted in their ability to appear in these constructions, which involve an extension of their basic meaning and argument structure, than their English counterparts.

---

4 The concept of ‘vanilla verb’ is taken from Greek. Each Greek verb comes in two flavors - vanilla and strawberry. With the vanilla flavor of the verb one can form certain tenses: present, imperfect, etc.; while with the strawberry flavor the rest of the tenses: past, future, etc.

5 Telic means depicting direction
(i) The swarm alternation:

The swarm alternation, sometimes characterized as the intransitive counterpart of the locative alternation, is illustrated by the following examples:

17.(a) The jewels glittered (on her neck). \( \textit{Locative variant} \)

(b) Her neck glittered (with jewels). \( \textit{With variant} \)

Verbs such as “glitter” and “bloom” which we assume to be intransitive can occur as locative variants in English, optionally taking a locative phrase. The “with” variant is different in transitivity from the locative variant since both locative (on her neck in 17a) and theme (her neck in 17b) arguments are obligatory.

In Hindi, verbs such as “camaknaa” (‘glitter”) are intransitive and appear with an optional locative phrase:

18. (a) hiire usa-ke gardan-par camak rahe haiN.
Diamonds she-Gen neck-Loc shine PROG be.
The diamonds are glittering on her neck.

However, in the \textit{with variant}, which involves an extension of its canonical argument structure, the verb “camaknaa” is ungrammatical as shown below.

(b) *usa-kaa gardan hiiroN-se camak rahaa hai.
*She-Gen neck diamonds-Inst shine PROG be.
Her neck is glittering with diamonds. \[ \text{[Bhattacharya 2002]} \]

Other verbs which show a similar restriction in Hindi but not in English are \textit{Kilanaa} “bloom”, \textit{naacnaa} “dance”, \textit{rengnaa} “crawl”, \textit{bajnaa} “sound”, \textit{Tapaknaa} “drip”.

(ii) The material product alternation:

We shall only be concerned with the types of verbs which can take the “material” in the object position and the “product” in the oblique PP phrase. A variety of verbs in English such as “grind” can occur with an oblique PP to imply the creation of a product from a material (19b).
19a. I ground the peas.

   b. I ground the peas into a paste.  *Product variant*

As in English, the Hindi equivalent of the verb “grind” (“piisnaa”), entails a change of state in the entity referred to by the direct object argument (20a). However, it still cannot occur with an oblique phrase which expresses the product created by the action described by the verb (20b).

20. (a) mai-ne maTar-ko piisaa.
   Agent    Patient  V  I-Erg peas-Acc ground.
   I ground the peas.

(b) *mai-ne maTar-ko leii-meN piisaa.
   Agent    Patient        Product       V  I-Erg peas-Acc paste-Loc ground.
   I ground the peas into a paste.

Rather, the creation sense can only be expressed when a vanilla verb entailing the sense of transformation is used in this construction:

21.(a) saadhuujii-ne duudh-ko paanii-meN badalaa.
      The hermit changed the milk into water.

(b) saadhuujii-Erg milk-Acc water-Loc changed.

A variety of verbs show a similar lack of flexibility in this construction in Hindi but not in English: “luDaknaa” ‘roll’, “Toknaa” ‘pound’, “siinaa” ‘sew’, “pakaanaa” ‘cook’.

(iii) Understood body-part object alternation:

This alternation is found with verbs that describe certain gestures or signs made with a particular part of the body. In transitive use, the verbs take this body part as their direct object and in intransitive use it is not taken but understood.

22.(a) Jennifer craned her neck.
   (b) Jennifer craned.

Both these sentences translate in Hindi as:
3.3. Others

We have extended Bhuvana’s study to other types of verbs like instrument verbs or noun verbalizer class, resultative verbs etc. We present below some interesting examples that represent the differences in the syntactic structures of English and Hindi of such verbs.

(i) (a) English: John speared the buffalo.
Agent V Theme
(b) Hindi: john-ne bhaalaa fenka-kar bhainsa-ko maaraa.
Agent Inst Adv (manner) Theme V
john-erg spear throwing-by buffalo.Acc kill-Sg.Msc.Perf

In English, instrument can become a verb but in Hindi it is not possible. We need a verb (‘maara’ here) to complete the action.

(ii) (a) English: He broke the lock open.
Agent V Patient Adj.
(b) Hindi: *usa-ne taalaa khulaa toDaa.
Agent Patient Adj V
He-Erg lock open break-Sg.Msc.Perf
(c) Hindi: usa-ne taalaa toDa-kara khoala.
Agent Patient V1 V2
He-Erg lock by-breaking open-Sg.Msc.Perf

In the above sentence adjective becomes main verb in Hindi, while with some verbs like ‘paint’ this does not happen as shown in the following example:

(iii) (a) English: John painted the house yellow.
Agent V Patient Adj
(b) Hindi: john-ne ghara-ko piilaa rangaa
Agent Patient Adj V
John-Erg house yellow paint.Sg.Msc.Perf

Here, adjective remains as adjective even in Hindi.

(iv) (a) English: John chaired the meeting.
(b) English: The bill was tabled in the parliament.

In Hindi similar constructions are not possible. English allows verbalization of nouns like ‘table’ and ‘chair’ while Hindi does not.
The differences arise because of the semantic properties of verbs as well as its class. For instance, the following are possible where the instrument is verbalized:

**Knife:** John knifed the burglar.
**Axed:** John axed the wild animal.
**Gun:** John gunned the burglar down.
**Spear:** John speared the wild animal.

But again, not all instruments can be verbalized.

**Pistol:** John pistolled the burglar down.

The differences in the argument structure of English and Hindi is attributed to lexical gaps, generally. But the above examples reflect that the difference is due to the complexity in the verbal semantics which is not just across languages but also within the language when we consider the class of verbs, as seen in the verbs derived from instruments, furniture etc. To support this point of view, we present some more examples.

(v) (a) Hindi: siita roti bel rahi haiN.
(b) English: *Sita is rolling roti
(c) English: Sita is rolling the dough into roti

In Hindi, it is possible to ‘roll’ the end-product, while in English it is not. But in the case of the verb ‘fry’, the end-product can be acted upon in both the languages.

(vi) (a) Hindi: Sita puri tal rahi haiN.
(b) English: Sita is frying puri.

Sometimes, to depict one action, English needs two verbs while Hindi shows the same with the help of TAMsup 6  inflection.

(vii) (a) English: He continued swimming.
(b) Hindi: vaha tairataa rahaa.

---

\(^6\) TAM = tense, aspect and modality
4. Conclusion:

In the present preliminary work, attempts have been made to deal with the verbs in detail with respect to their syntax and semantics. To capture the syntactic nature of verbs, this sort of analysis is necessary.

We have presented data showing differences in Hindi and English in the mapping of similar verbs onto syntactic frames. In order to account for this cross-linguistic difference, we argue that while the default mapping strategy is verb-driven, in some languages such as English, lexical specifications of the verbs are overridden by constructional semantics. In languages such as Hindi, the lexical specifications of the verbs and TAM will always prevail.

Furthermore, other types of verb alternations which involve an extension of the meaning (and argument structure) of the verb are also absent in Hindi. We suggest that this argues for differences between Hindi and English in the role that constructions and TAM play in the mapping between verbs and syntactic frames. We tried to account for gradients in the acceptability of verbs in a particular construction across languages, as well as across constructions within a language. Such a gradient can be represented as an (implicational) hierarchy of verb-construction mapping compatibilities.

Further research is required to investigate more fully the interaction of verb semantics, argument semantics, and construction type in constraining the mapping of verbs onto syntactic templates across languages.

The classification of verbs and placing them in the syntactic frames according to their selection preference and other semantic properties seem essential in most text information processing tasks like machine translation, information extraction etc. This study will help in machine translation of English sentences with the types of verbs we have taken into consideration into Hindi. It will also help in building concept based dictionaries.
5. Acknowledgement

The authors sincerely thank Prof. B. Lakshmi Bai and Dr. Dipti Misra Sharma for the discussion and useful feedback.

6. References

Bhattacharyya, Pushpak and Debasi Chakrabarti, Syntactic Alternations of Hindi Verbs with Reference to the Morphological Paradigm, Language Engineering Conference (LEC’02), 2002.


http://www.mpi.nl/world/persons/private/bhuvana/lingua2_papnew.doc


