SE367A Project Report Complex Predicates in Hindi

By:

Sachet Chavan (Dept. of HSS) Pranav Kumar (Dept. of Electrical Engineering) Guide: Prof. Amitabh Mukherjee

Abstract:

Complex predicates are found in South Asian languages more than European languages. We will analyze different complex predicates in Hindi and try to gauge their acceptability rates. Thus the complex predicate found to be acceptable to majority of people might go on to get stored in WordNet or a dictionary.

Introduction:

Complex predicate is a multi-word compound that functions as a single verb. Complex predicates can be formed by two type of combination:

Noun + Verb Verb + Verb राम किताब <u>पढ रहा ह</u>ै |

For this project we have considered Verb + Verb type combination only. This type of combination is formed by combining a Heavy verb (HV) with light verb:

Heavy Verb + Light Verb

निकल गया, हस पड़ा

Heavy verb:

Heavy Verb contains most of the meaning of the compound.

E.g. in निकल गया, निकल is the heavy verb.

Light Verb:

Light Verb contributes to finer aspect of temporal meaning.

Verb + Verb type combination can also be of two types depending on which occurs first HV or LV.

Standard aspectual complex predicate construction: These are formed by combining HV + LV (in order)

```
रामने श्याम को तमाचा मार दिया
```

Reverse aspectual complex predicate construction: These are formed by combining HV + LV (in order)

```
रामने श्याम को तमाचा दे मारा
```

Standard Aspectual complex Predicates are more common than Reverse Aspectual.

Except for a few exception all Verb + Verb Type complex predicates can be replaced by some inflection of HV. E.g. निकल गया can be replaced by निकला

Also all HV + LV does not give a complex predicate. E.g. निकल डाला is not found acceptable by most people.

Related Works:

The HSS department of IIT Bombay also conducted some research in this subject. Their work is motivated primarily by the need to automatically augment lexical networks such as the Princeton Wordnet. Another paper from IITB presents their experience in the construction of lexical knowledge bases for Indian languages with special attention to Hindi. The question of storing or deriving complex predicates has been dealt with linguistically and computationally in their work.

Our Experiment:

Complex predicates are lexical compound verbs. Now as they are a part of our day-to-day language we are very much accustomed to them. But if a person has no knowledge of Hindi, he won't be able to decipher the meaning of these compound verbs. And that's because individually these heavy verbs and light verbs mean different but when put together as a complex predicate, the meaning they try to convey changes. Also not each and every heavy verb and light verb combination can be said to be a complex predicate. There are very limited number of combinations which are prevalent enough to be accepted as a part of general vocabulary.

Our experiment used two methods to obtain the same kind of results. Those two methods were gaze tracking-based model and survey based study. The first part of our experiment was that of gaze tracking in which we made our subjects read a text comprising of sentences which contained standard aspectual complex predicates, reverse aspectual complex predicates and plain verbs. For our survey based part, we circulated a survey containing a grid like structure comprising 5 heavy verbs and 5 light verbs which made 25 different combinations and asked the participants to vote the ones they find appropriate.

Following are the sentences from the reading text we used in the gaze tracking part:

(We gave standard aspectual, reverse standard aspectual and inflection of same HV)

लीला ने श्याम को चिठ्ठी लिखी । शीला ने राम को चिठ्ठी मार लिखी। घनश्याम ने विवेक को खत लिख मारा। राम गाना गा पड़ा। राम ने गाना गाया। कुत्ते को देख कर शिवम भाग पड़ा। बन्दर को देख कर भीम भागा। बाज़ परिंदे पे झपट गया। चूहे को देख बिल्ली उसपे झपटी। ग्स्सा आने पर सचिन चीख पड़ा। अपनी माँ की डाट स्न कर बच्चा रो डाला। मिठाई ना मिलने पर बच्चा रोया। घडी देख कर रोहित निकल उठा । चुटकुला सुनकर दिनेश हस चला। विजय ने घर को दे बेचा। गीली फर्श पे नरेश फिसल पडा। शाम के सात बजते ही रोहन निकल चला। सुबह के आठ बज गए और वो चल निकला। Our main aim was to track the eye movement of the subjects while they read these sentences.

For the survey part we circulated a google doc in our hostel. As you know hostel inmates consist of students from different parts of the country, so this prevented our survey from getting dialect biased and give us a more general result. So as you can see in the doc below, we created a grid-like structure asking them to vote for the combinations they found to be acceptable in their vocabulary. There were given 5 heavy verbs, namely गया, चला, पड़ा, डाला, उठा, and 5 light verbs, namely निकल, कह, रो, हस, बोल. So there were 25 different choices and they were allowed to vote as many as they wanted.

survey

Following are some possible combination of verb . if you think they are used in general then check box corresponding to them

eg . if निकल + गया seems a valid combination as used in "राम घर से निकल गया।" . then select the box corresponding to them.

	गया
निकल	O
कह	Ø
रो	Ø
हस	O
बोल	O

पड़ा
. ©
©
©
O
©

	चला
निकल	©
कह	ø
रो	Ø
हस	0
बोल	O

	डाला
निकल	Ø
कह	O
रो	©
हस	©
बोल	O

	उठा
निकल	O
कह	O
रो	©
हस	0
बोल	O

Results:





कह







हस



The above graphs show the number of votes each light verb got for its respective combination with heavy verb.

Following are the results of the gaze tracking experiments for three different subjects: (Radius of circle is proportional to Saccade duration).





Conclusion:

Thus from result of survey we can assume that the combinations with higher number of votes can definitely be a part of the WordNet and has a high acceptability rate. The votes received are subjective to the participants but we do get a rough idea of which combinations are prevalent and which are not. like निकल डाला is not found acceptable by any of the subject .Also from Gaze tracking result we verify as HV (निकल) contains most of the meaning circle of saccade is much longer around them than their LV counterpart (चला).

Future Work:

(a)Survey on larger population so that we have more reliable data and can study demographic relation (expected) for acceptability of HV +LV combination.

(b)Analyzing data of saccade time for standard aspectual, reverse standard aspectual and inflection of same HV complex predicates to get ratio of saccade time for those case to get idea of relative use of these form in day to day usage.

References:

- D.Chakravarti H.Mandalia R.Priya V.Sharma P.Bhattacharya; 2008: Hindi Compound Verbs and their Automatic Extraction(IITB)
- Shakthi Poornima, Jean-Pierre Koenig; 2009: Hindi Aspectual Complex Predicates (State University of New York at Buffalo)