

8.6 Causative Compound Verb Constructions: A Generative Lexicon Account

Sanjukta Ghosh and Anil Thakur

Abstract—Complex predicates are always in the focal area of research in both theoretical and computational linguistics due to their interesting syntactic and semantic behavior. One of the research questions in the literature of the complex predicate is where to account them: in the syntax or in the lexicon. Present paper provides a lexical treatment of one of the special type of complex predicates, viz., compound verbs used for causativization following Pustejovsky (1995).

Index Terms—Causative Verbs, Complex Predicate, Generative Lexicon

I. INTRODUCTION

COMPLEX predicates are always in the focal area of research in both theoretical and computational linguistics due to their interesting syntactic and semantic behavior. One of the research questions in the literature of the complex predicate is how and where to account and accommodate them: in the syntax or in the lexicon. The present paper takes up one of the types of complex predicates, viz., the compound verbs used for causativization and presents their analysis within Generative Lexicon framework proposed by Pustejovsky (1995).

A compound verb can be defined as a verbal predicate composed of two verbs where the meaning of the compound is not compositional, i.e., it cannot be obtained by computing the meaning of their constituents. Because of this specific property of the compound verb and the complex predicate in general, they have always been considered as one of the major problems in many NLP applications. The complex predicates fall under the broader category known in Computational Linguistics literature as Multi Word Expressions (MWEs) which can be roughly defined as a sequence of words acting as a single unit at some level of linguistic analysis. Building lexicon for this kind of expressions is still not as developed as it is for single word expressions. However, the necessity for developing lexicons of different kinds of MWEs taking both syntactic and semantic information is recognized by the NLP research community worldwide especially in a multilingual environment. But as a matter of fact, this kind of project is still in the conceptual level in a truly multilingual society like India mostly because of the linguistic complexity associated with these expressions and lack of formal mechanism to handle that complexity in the mainstream linguistics.

Sanjukta Ghosh and Anil Thakur are with the Department of Linguistics, Banaras Hindu University, Varanasi (e-mail: san_subh@yahoo.com, anilt@bhu.ac.in)

This paper is an attempt to formalize the lexical information of one particular kind of compound verbs for building lexical resources of Indian languages. The work can be extended for other kinds of compound verbs (where causativization is not involved) and complex predicates in general.

II. BACKGROUND OF WORK

The languages of India (at least for instance, Indo-Aryan and Dravidian, the two major language families of India) have a large number of complex predicates. From theoretical linguistics point of view there have been attempts to analyze these constructions within Chomskyan generative grammar tradition as well as in some other theoretical models like Head-driven Phrase Structure Grammar (HPSG) and Lexical Functional Grammar (LFG). All these models of grammar attempt to generate the complex predicates by some syntactic operations in the grammar itself. Though LFG is a more lexicon-oriented framework, yet the operation adopted to change the predicate status is by introduction of a lexical rule in the grammar instead of a syntactic rule.

It is only in the last decade that a completely lexicon-centric model has been introduced by James Pustejovsky (1995). It differs from all the previous frameworks by virtue of its claim that the lexicon, like syntax, can also be generative. The Generative Lexicon framework is devised to account for the polysemous nature of words as well as their creative uses. It analyzes a word in four different structural levels: argument structure, event structure, qualia structure and lexical-inheritance structure. The argument structure in this framework not only provides the logical arguments of a predicate as discussed in other literature but also other possible arguments such as shadow arguments etc. Event structure is the definition of the event type (like state, process, transition) of a lexical item or a phrase. The framework also provides mechanism to divide an event into sub-events and allows a concept of event-headedness which specifies the foregrounded sub-event in the predicate. Qualia structure is the mode of explanation composed of formal, constitutive, telic and agentive roles. The lexical-inheritance structure places a lexical word in a broader paradigm and distinguishes it from the others.

The advantage of adopting this framework is that it allows one to specify a greater internal structure of a word and provides a richer co-compositional mechanism for creating new words.

III. CAUSATIVE COMPOUND VERB CONSTRUCTIONS IN BANGLA AND HINDI

Compound verbs consist of two verbs; the first of them provides the core meaning of the verb and the second modifies its meaning in a significant way. Examples of such verbs from Hindi are /cal denaa/ 'to start', /cal paRnaa/ 'to start suddenly' etc. where the first element is /calnaa/ 'to move' and a broader category or hypernym of the compound verbs constructed. The second verb is responsible for the finer semantic nuances associated with the compound verb.

A causative construction can be defined using the following criteria:

1. It is a single event constituting of two sub-events.
2. These sub-events can be sequential or overlapping.
3. One of the sub-events causes the other to happen.
4. There must be some argument sharing between the two sub-events (argument coherence in Pustejovsky's term).
5. Logical entailment relation holds between the two events.

A causative construction is generally expressed in Hindi by some morphological markers, viz., adding /aa/ and /waa/ to the verb root. The traditional grammar calls these as first and second causatives respectively. Examples of such causatives are from the root /khaanaa/ 'to eat', /khilaanaa/ and /khillwaanaa/ 'to feed' and 'to cause to eat (by somebody)'. In Bangla also there is a morphological affix to mark this causativization process, viz., /aa/, as from /bOl/ 'to speak' the derived form is /bOlaa/ 'to make to speak'. However, if somebody makes one to speak with the help of a third person, which is expressed by Hindi /waa/ causative (/bulwaanaa/ in this case), the choice in Bangla is to use a compound verb like /bolie neoa/ or /bolie deoa/ where the first verb is in conjunctive participial form of the verb /bOlaa/. The second verb denotes whether the action is for oneself (in case of /lena/) or for other (in case of /denaa/). Bangla exhibits such kind of causative constructions in large numbers. Sometimes they are the only form to express causality. Even in Hindi, such kind of compound verbs are found which express only causality like /samjhaa denaa/ 'to convince'. The following section takes up some of the representative verbs of Bangla and Hindi of this kind and compares them with the corresponding pure morphological causative forms. These compound verb forms have not been mentioned before formally in any work on lexical resources. The present paper suggests a formal model of representation for their meaning differences.

IV. LEXICAL RESOURCE MODEL FOR CAUSATIVE COMPOUND VERBS

This section takes up four major categories of verbal predicates following Vendler (1967) and analyzes one example of the verbs of each category in a lexical resource model adopted in this paper. The four main categories are event, state, accomplishment and achievement predicates. From the first type activity predicate the representative example chosen is Bangla /douRono/ or Hindi /dauRnaa/ 'to run'. Bangla does not have a morphological causative form for this, it uses a complex predicate form of N-V construction /dauR korano/ 'run(n) do-causative-inf' or 'make somebody run' with the causative form of the second verb /kOra/. For Hindi /waa/ form of this verb, i.e., /dauRwaanaa/ Bangla uses one more verb to make it a more complex predicate. The entry in the lexical model of the verb is illustrated in table I below.

Table I

<p>H. <i>dauRwaanaa</i> B. <i>douR korie neoa/douR kOrano</i> EVNTSTR = E1 = e1: process E2 = e2: process RESTR = < α HEAD = e1 < = sequential ARGSTR = Arg1 of e1 = animate (x) Arg2 of e2 = animate (y) Arg3 of e2 = animate (y)</p> <p>Qualia = cause-lcp Formal = dauRnaa (e2,y) Agentive = dauRwaanaa_act (e1,x)</p>	<p>H. <i>dauRaanaa</i> B. <i>dauR kOrano</i> EVNTSTR = E1 = e1: process ARGSTR = Arg1 = animate (x) Arg2 = animate (y)</p> <p>Qualia = process-lcp Agentive = dauRaanaa_act (e1,x)</p>
--	---

The examples below illustrate the difference of the two forms in terms of entailment.

3a. *apni Sobaike douR kOraben manei to ar Sobai doUrobena.*

run do-fut

3b. *aap sabko dauRaeyeMge, iskaa matlab yah nahiiM ki sabhii daureMge hii.*

'If you run everybody, it does not mean that all would run.'

4a. **apni amake die procur dour korie niechen kintu ami douRoi ni.*

run do-caus-perf take-pr-prf

4b. **aapne mujhe bahat dauRwaayaa lekin maiN dauRii nahiiM.*

'You made me run a lot but I did not run.'

The next category is the state class of verbs and the representative example taken here is Hindi /samajhnaa/ 'to understand' and its counterpart in Bangla. For this verb, the second causative form in both the languages is a compound verb. The difference in the semantics between the first and the second causative is also clear from the lexical model given below.

Table II

<p>H. <i>samajhaa denaa</i> B. <i>bujhie deoa</i> 'to cause someone understand'</p> <p>EVENTSTR = E1 = e1: process E2 = e2: state RESTR = < α HEAD = e1 < = sequential ARGSTR = Arg1 of e1 = animate (x) Arg2 of e1 = concept (z) Arg3 of e2 = animate (y)</p> <p>Qualia = cause-lcp Formal = samajh (e2,y) Agentive = samajhaa-dena_act (e1,x)</p>	<p>H. <i>samjhaanaa</i> B. <i>bojhano</i> 'to make understand'</p> <p>EVENTSTR = E1 = e1: process</p> <p>ARGSTR = Arg1 = animate (x) Arg2 = animate (y) Arg3 = concept (z)</p> <p>Qualia = process-lcp Agentive = samjhaanaa_act (e1,x) STATE = samajhaa_state (e2, y)</p>
---	--

The sentences below capture the difference of the two verbs in both Bangla and Hindi.

- 5a. H. *maiNne usko samjhaayaa lekin wah nahiiN samjhaa.*
tried to make understand understood
- 5b. B. *ami oke bojhalam kintu o bujhlo na.*
tried to make understand understood
'I tried to make her understand but she did not understand.'
- 6a. *H. *maiNne usko samjhaa diyaa lekin wah nahiiN samjhii.*
understand gave understood
- 6b. *B. *ami oke bujhie diechilam kintu o bojhe*
understand-causative-conjunctive participle gave understood
'I made her understand but she didn't understand.'

The compound verb of the example 6 expects a state of understanding of the affected argument as a result of a telic event, whereas the verb in 5 does not expect so. It denotes an atelic event.

The next class is the accomplishment verbs where a process event is followed by a state event. The representative example of this class is Hindi /banaanaa/ 'to make' and its Bangla counterpart /gODano/. The second causative is again a compound verb in Bangla; Hindi also alternatively uses a compound with the morphological /waa/ causative. The following table illustrates the lexical model for this type of verbs:

Table III

<p>H. <i>banwaanaa/banwaa denaa</i> B. <i>goDie deoa/neoa</i> 'to cause some make/create'</p> <p>EVENTSTR = E1 = e1: process E2 = e2: result RESTR = < α HEAD = e2 < α = sequential ARGSTR = Arg1 of e1 = animate (x) Arg2 of e1 = material (z) Arg3 of e2 = artistic-creation (y) CONST = z Formal = physical object</p> <p>Qualia = cause-lcp Formal = banaa_huaa (e2,y) Agentive = banaa-dena_act (e1,x,z)</p>	<p>H. <i>banaanaa</i> B. <i>gODano</i> 'to make create'</p> <p>EVENTSTR = E1 = e1: process E2 = e2: result RESTR = < α HEAD = e1</p> <p>ARGSTR = Arg1 = animate (x) Arg2 = material (y)</p> <p>Qualia = process-lcp Agentive = banaanaa_act (e1, x, y) FORMAL = creation (e1,y)</p>
---	--

The difference between *banaanaa* and *banwaanaa* and their Bangla counterparts lies in their event-headedness. *Banaanaa* foregrounds its process sub-event whereas *banwaanaa* foregrounds its result sub-event. e.g.

- 7a. *wah muurti banaataa hai.*
He model make-3P-sg
'He makes (clay) models.'
- 7b. *usne ek din meM yah muurti bannayii.*
He-erg one day-loc this model made
'He made this model in one day.'

The last category of verb in Vendler is the achievement verbs and the representative example here is Hindi /pahucanaa/ 'to reach'. This verb also preferably uses a causative form in both Bangla and Hindi. Hindi also has a marginal use of a morphological /waa/ causative. In Bangla, there is no difference between the first and the second causative in formal structure. The lexical model showing the details of the semantic feature of the forms is illustrated in table IV.

Table IV

<p>H. <i>?pahuMcaanaa/pahuMcaa denaa</i> B. <i>pouMche dea</i> 'to cause something reach'</p> <p>EVENTSTR = E1 = e1: process E2 = e2: state RESTR = < α HEAD = e2 < α = sequential ARGSTR = Arg1 of e1 = animate (x) Arg2 of e1 = material (y) Arg3 of e2 = pahucaa-huaa (y) Formal = physical_object/individual</p> <p>Qualia = cause-lcp Formal = pahuMcaa_huaa (e2,y) Agentive = banaa-dena_act (e1,x,x)</p>	<p>H. <i>pahuMcaanaa</i> B. <i>pouMche dea</i> 'to make reach'</p> <p>EVENTSTR = E1 = e1: process E2 = e2: state RESTR = < α HEAD = e1</p> <p>ARGSTR = Arg1 = animate (x) Arg2 = physical_object (y)</p> <p>Qualia = process-lcp Agentive = paucaanaa_act (e1, x, y) FORMAL = pahucaa_huaa (e2, y)</p>
---	---

The basic architecture of this lexical model can be extended to other types of verbs as well including causative and non-causative, simple and complex predicates.

V. CONCLUSION

Forming compound verbs to express causativization is a productive strategy in Bangla. We have also shown that though Hindi is believed to have morphological causative only, in many cases they parallel with the compound verbs and in a few cases compounding is the only way of expressing a causative predicate. So far no lexicon has listed the compound verbs with all their semantic nuances and syntactic specifications in the lexicon so as to use this information for further computational operations. In fact, lexicon of this special type can be a rich source for researches in theoretical and computational linguistics. Works in Indian languages within the theoretical framework of Generative Lexicon has begun only in recent years (Raina 2005) and have shown promise for a better analysis of certain linguistic phenomena in Indian languages.

REFERENCES

- [1] Pustojvsky, J. 1995. *Generative Lexicon*. MIT Press.
- [2] Achla M. Raina 2005. "Complex Predicates in the Generative Lexicon". Bouillon, Pierrette and Kanzaki, Kyoko (eds.), *Proceedings of GL'2005, Third International Workshop on Generative Approaches to the Lexicon*, School of Translation and Interpretation, University of Geneva, Switzerland, 210-221.
- [3] Vendler, Zeno 1967. "Verbs and Times". In Z. Vendler, *Linguistics in Philosophy*, pp. 97-121. Ithaca: Cornell University Press.

This paper was presented at LRIL-2007: National Seminar on Creation of Lexical Resources for Indian Language Computing and Processing at C-DAC Mumbai (26th to 28th March 2007), jointly organized by the Commission for Scientific and Technical Terminology (CSTT), New Delhi, MHRD, Govt. of India and the Centre for Development of Advanced Computing (C-DAC), Mumbai, Department of Information Technology, MC&IT, Govt. of India.