Representation of Complex Predicates in Wordnet

G.Uma Maheshwar Rao
Centre For ALTS
University of Hyderabad
`guraohyd@yahoo.com`

K.Rajyarama
Centre For ALTS
University of Hyderabad
`krajyarama@gmail.com`

Abstract

Indian languages exhibit a large number of complex predicate constructions comprising of a V+V combination or an N+V combination. The present paper focuses on issues related to the representation of complex N+V constructions in wordnet. We have attempted to evolve a mechanism based on the semantic features of the nouns involved in the N+V constructions for predicting the compositionality of a complex N+V construction. Such an analysis it is felt will help in classifying complex predicates into various types and an appropriate mechanism for representing each of these types has been suggested.

Transliteration: a,A,i,I,u,U,e,E,ai,o,O,au,M,H, k,kh,g,gh,ng,c,ch,j,jh,ny,T,Th,D,Dh,N,t,th,d,dh,n, p,ph,b,bh,m,y,r,l,L,v,sh,S,s,h

1   Introduction

Telugu and almost all Indian languages possess a large number of complex word constructions which comprise of (a) verb + verb or (b) non-verb + verb. The non-verb in (b) type of constructions can be a noun, an adjective or an adverb. These constructions are variously referred to as Compound verbs (Hook, 1974), Conjunct verbs (Wallace, 1985), Composite verbs and Complex predicates (Krishnamurti, 1992; Verma, 1993). Depending on the grammatical category of the non-verb constituent they are also referred to as 'NV compounds', 'AV compounds' (cf. Kachru, 1980; Hook, 1993; Verma, 1993). In recent literature the term Complex Predicates has become more or less stabilized to refer to these constructions. Complex predicates of N+V type are discussed in this paper. Even though complex predicates structurally comprise of two words, syntactically and semantically they behave like single constituents. Such structurally complex but semantically unique constructions can raise important issues of processing and representation for lexical storage and retrieval purposes. In this paper we would like to focus on issues of the following sort:

- a. How should complex predicates be processed and represented in wordnets?

- b. What are the principles/assumptions which underlie storage and retrieval of complex predicates in wordnets?

The paper is organized as follows: In section 1 we provide a brief introduction to complex predicates. In section 2 we discuss about wordnets and their organization. Section 3 deals with complex predicates in Telugu. Section 4 discusses issues related to representing complex predicates in Wordnets. Sections 4.1 and 4.2 deal with the proposals for representing complex predicates in wordnets. Section 5 concludes the paper.

2   Introduction to Wordnet

Large scale lexical data bases like wordnets are essential prerequisites for various NLP related tasks like machine translation, information extraction and retrieval. From the theoretical point of view, identification of word senses (including multiword sequences) their representation and their disambiguation is of great concern for computational studies. In Wordnets, words are represented as related to each other by a set of semantically statable relations. In wordnets nouns, verbs, adjectives and adverbs are grouped into sets of cognitive synonyms or synsets, each expressing a distinct concept. Miller's (1993) paper spells out the objectives and the architecture of wordnet. Miller (1993), Fellbaum et al. (1993) and Fellbaum

(1993) describe the organization of nouns, adjectives and verbs in wordnet respectively.

Wordnets for major grammatical categories viz. nouns, verbs, adverbs and adjectives have been created. Nouns in wordnets are organized on the basis of lexical inheritance system and are related to each other on the basis of relations like hyponymy, hypernymy , synonymy and meronymy (Miller, 1993). The organization of the nouns in the wordnet is based on the intuition that common nouns provide a superordinate term along with their distinguishing features. This generates a lexical inheritance system in which each word inherits the distinguishing features of all its superordinate words.

Fellbaum (1993) proposes a different set of semantic relations for verbs, since verbs differ substantially from other categories like nouns and adjectives in terms of the semantic relations that hold between them. Verbs are divided into different semantic groups based on semantic criteria. Semantic relations like polysemy, synonymy, entailment and troponymy form the basis for the organization of relations among verbs. In wordnet adjectives are divided into two types viz. descriptive and relational. According to Fellbaum et al. (1993), "Descriptive adjectives ascribe to their head nouns values of (typically) bipolar attributes and consequently are organized in terms of binary oppositions (antonymy) and similarity of meaning (synonymy)". Relational adjectives are classified to be stylistic variants of the nouns they modify and so are cross referenced to the noun files. (Fellbaum et al., 1993).

In addition to the single lexemic representations in terms of their grammatical categories, it is also necessary to represent morphological relations expressed by constructions like complex predicates in wordnets.

3 Complex Predicates in Telugu

In this section, we focus on the complex predicates of the N+V type in Telugu. The high frequency and productivity of N+V can be attributed to the fact that Telugu is a verb final language and hence does not require any special efforts to incorporate a noun in the complex predicate. More over this appears to be a very handy mechanism to accommodate into Telugu loan words from the domains of science and technology.

The following dicussion involves verbs, both transitive and intransitive, which occur as constituents of NV constructions.

3.1 Transitive Verbs in Telugu

Rajyarama (1998) identifies an exhaustive list of 22 transitive verbs which occur as heads in N+V constructions. Verbs like *ceyyi* 'to do', *peTTu* 'to keep', *koTTu* 'to beat', *paTTu* 'to hold', *tiyyi* 'to take', *veyyi* 'to throw' are a few examples of this category. Out of these *ceyyi* 'to do', *veyyi* 'to throw' are very frequently used (both transitively and intransitively). A few examples illustrate the use of the verbs mentioned above:

- nidra paTTu 'fall asleep' 'sleep hold',

- Ita koTTu 'to swim' 'swimming hit

- guraka peTTu 'to snore' 'snoring keep'.

Consider a verb like *ceyyi* 'to do' as in 1. *snAnaM ceyyi* 'to bathe' which has a literal meaning 'do/make bath' and in 2. *jalubu ceyyi* 'to catch cold' which literally means 'to do cold' as illustrated below:

- 1. sIwa snAnaM cEsiMdi.
  Sita bath-do -pst-3p-sg-f
  'Sita bathed'

- 2. sIwaki jalubu cEsiMdi
  Sita -dat. cold do-pst-3p-sg-f
  'Sita caught cold'

3.2 Intransitive verbs in Telugu

Nearly around 11 verbs occur as constituents of N+V constructions. Some of them are *avvu* 'to become', *paDu* 'to fall', *kalugu* 'to occur', *puTTu* 'to be born'. Consider the following examples:

- 3. bAdha paDu 'to be pained' 'pain fall'

- 4. Akali avvu 'to be hungry' 'hunger become'

- 5. digulu paDu 'to worry' 'worry fall'

It is important to note that in none of the examples cited above the meaning of the N+V sequences is compositional. The verb loses its lexical meaning

in this particular context.

Most of the earlier works on complex predicates (Hook, 1974, 1993; Dasgupta 1977; Verma, 1993; Krishnamurti, 1992; Uma Maheshwara Rao, 1995; Rajyarama, 1998) have applied a number of morphological and syntactic tests in order to test their compositionality. Some of these include tests like scrambling, coordination, modification and interrogation. These tests help in distinguishing between N+V constructions which are combinations of direct object+verb and N+V constructions which are instances of an incorporated noun and the verb. The outcome of these tests is that complex predicates with non-compositional semantics and which do not yield themselves to any of the tests mentioned above should be treated as 'wholes' and not as independent units.

4 Representing Complex Predicates in Word net

Complex predicates comprising of a noun and a verb involve a number of processes which can be morphologically stated and are used basically to represent newer verbal concepts. Although the verbs which are constituents of these constructions are a finite set, the non-head constituent nouns are not. Therefore it is important to evolve a mechanism which can predict the meaning of the complex N+V constructions from the meaning of these non-head constituents in the context of the head.

From a lexicographic point of view these constructions pose problems of representation and storage. In this paper we seek to address and resolve some of the vital problems faced by a lexicographer while handling such constructions.

We propose that complex N+V constructions like *bhaya paDu* 'to fear' can be represented in Wordnet by adopting one of the following approaches:
(a) List based approach
(b) Rule based approach
In the following sections, we discuss the basic assumptions and the consequences of each of these approaches:

4.1 List based approach

Under this approach we assume that it is possible to list all the N+V sequences in the lexicon. Such a listing begins with identifying the verbs that can occur as constituents of the complex predicates. In any language only a finite set of verbs can occur as complex predicate constituents.

Once the verbs are identified, the nouns that can co-occur with each of these verbs will also have to be exhaustively listed.

Following this all possible N+V combinations in a language will be listed. Theoretically such an approach may appear feasible especially taking into consideration the huge memory available with the computer and also the speed and accuracy with which the machines operate. However, this method of merely listing the N+V combinations may be constrained by various limitations like:

- 1. Thousands of forms (N+V combinations) will have to be listed which can be quite cumbersome.

- 2. There is always a possibility of a new N+V combination entering a language.

- 3. There is also the problem of unnecessary redundancy which can be avoided.

In order to overcome the limitations cited above we propose an alternate approach which takes its cues from human language processing mechanisms. We strongly argue for an approach which simulates how humans process, store and retrieve complex predicates.

4.2 Rule based Approach

The basic assumption is that it is possible to process and store complex predicates in a way similar to how human beings recognize and store these constructions in their mental repertoire. Human beings do not list and store each and every complex predicate form whenever they come across one. Rather, human behavior seems to be rule governed. It seems that humans formulate rules based on the abstractions derived from the semantic features of the words involved in such constructions.

In this approach we argue for a method which is rule based. The rules are based on the generalizations drawn form the semantic and syntactic features of the words involved in a complex predicate. This approach is based on the assumption that it is possible to predict the nature of a complex N+V construction on the basis of the semantic features of the nouns involved. This is evident from the behavior of some of the most frequently used verbs like *ceyyi* 'to do', *veyyi* 'to throw', *peTTu* 'to keep' and *koTTu* 'to

beat/hit' when they function as heads of complex N+V constructions. In the following section we illustrate the semantic behavior of each one of these verbs mentioned above:

Let us consider the verb *ceyyi* 'to do'. The following examples demonstrate the types of nouns that this verb can combine with:

- 6. vaMTa ceyyi 'to do cooking'
  cooking   do

- 7. kUra ceyyi 'to cook curry'
  curry   do

Therefore it is possible to say that the verb *ceyyi* 'to do' when it combines with any noun which is characterized by the features [+artifact, +edible] the complex verb denotes 'to make'.
ceyyi 'to do' on the other hand when it combines with nouns which are [+artifact, -edible] as in:

- 8. bomma ceyyi 'to make a toy'
  toy   do

- 9. kuMDa ceyyi 'to make a pot'
  pot   do

the verb *ceyyi* 'to do' functions as a resultative verb bringing about change in the state denoted by the N. Consider the nature and function of the verb *ceyyi* 'to do' when it combines with nouns which denote states both physiological and psychological, as in the examples (10-13):

- 10. jalubu ceyyi 'to catch cold'
  cold   do

- 11. jabbu ceyyi 'to fall sick'
  sickness   do

- 12. mOsaM ceyyi 'to cheat'
  cheating   do

- 13. alavAtu ceyyi 'to create habit'
  habit   do

- 14. snAnaM ceyyi 'to bathe'
  bath   do

- 15. type ceyyi 'to type'
  type   do

- 16. print ceyyi 'to print'
  print   do

Here the verb is completely bleached of its lexical content and functions as a verbalizer. Consequently, the verb *ceyyi* 'to do' functions differently in each case depending on the semantic nature of the noun that is incorporated. The same can be represented as follows:

- 6-7. N   ceyyi
  [+af.,+edi.]

- 8-9. N   ceyyi
  [+ af.,-edi.]

- 10-13. N   + ceyyi
  [+ abs.]

Another verb *peTTu* 'to keep' behaves in a similar way. Let us consider the contexts in which *peTTu* 'to keep' occurs:

- 17. guraka peTTu 'to snore'
  snore   keep

- 18. parugu peTTu 'to run'
  pain   keep

- 19. maData peTTu 'to fold'
  fold   keep

- 20. kharcu peTTu 'to spend'
  spend   keep

- 21. lekka peTTu 'to count'
  count   keep

In the examples (17-21) the nouns are characterized by the feature [+action]. In this context the verb functions as only a verbalizing element. Similarly in the following complex predicates (22-25) nouns belong to the category of [+psychological state] and the output of the predicate is 'to cause' the state denoted by the noun.

- 22. bAdha peTTu 'to hurt'
  hurt   keep

- 23. bhaya peTTu 'to scare'
  fear   keep

- 24. sukha peTTu 'to comfort s'one'
  comfort   keep

- 25. kaSTa peTTu   'to cause hardship to s'one'
  hardship   keep

- 26. annaM peTTu 'to serve rice'
  rice   keep

- 27. bhojanaM peTTu 'to serve food'
  food   keep

- 28. kUra peTTu 'to serve curry'
  curry   keep

The complex predicates in examples (26-28) involve nouns which are characterized by the feature [+af.,+edi]. Here also the verb *peTTu* does not retain its lexical meaning of 'to keep something somewhere' but functions as a verbaliser.

Another verb *koTTu* combines with nouns and exhibits similar behavior. Consider the following examples:

- 29. Ita koTTu 'to swim'
  swim   hit

- 30. jebulu koTTu 'to pick pocket'
  pockets   hit

- 31. dArulu koTTu 'to way lay'
  ways   hit

- 32. mamdu koTTu 'to booze'
  liquor   hit

- 33. kampu koTTu 'to stink'
  stink   hit

- 34. vasana koTTu 'to smell'
  smell   hit

The noun in the example(29) is characterized by the feature [+action] as in *Ita* 'swimming' or [+concrete,+count] as in *jEbulu* 'pockets', *dArulu* 'ways', *maMdu* 'medicine' (examples; 30-32) and [+physiological state] as in *vAsana* 'smell' and *kaMpu* 'stink' (examples:33-34).

In the following, yet another verb *veyyi* 'to throw' combines with nouns of different types as in the examples (35-41):

- 35. Akali veyyi 'to be hungry'
  hunger-throw

- 36. bAdha veyyi 'to be pained'
  pain-throw

- 37. bhayam veyyi 'to be afraid'
  fear-throw

- 38. dAham veyyi 'to feel thirsty'
  thirst-throw

- 39. cali veyyi 'to feel cold'
  cold-throw

- 40. siggu veyyi 'to be ashamed'
  shame throw

- 41. ettu veyyi 'to tricks one'
  trick throw

In the above cases, complex predicates involving nouns and the verb *veyyi* 'to throw' denote psychosomatic states and these nouns are characterized by the semantic features [+psychological state] and [+physiological state].

Similarly consider the set of nouns that enter into complex predicate involving *veyyi* 'to throw':

- 42. tALaM veyyi 'to lock'
  lock   throw

- 43. muggu veyyi 'to draw designs'
  designs   throw

- 44. gaDiya veyyi 'to bolt'
  bolt   throw

- 45. ceyyi veyyi 'to lay hand'
  hand   throw

Nouns in examples (42-45) are characterized by the semantic feature [+concrete]. Consider another set of concrete nouns which are either self benefactive or other benefactive with which the verb *veyyi* 'to throw' combines:

- 46. baTTalu vEsuko
  'to wear[self] clothes'

- 47. ceppulu vesuko
  'to wear[self] chappals'

- 48. nagalu vesuko
  'to wear[self] jewellery'

Also Consider another verb *paDu* 'to fall'. This verb is found to be co-occurring with nouns characterized by the semantic feature [+psychological state] and even in this context the verb functions as a verbalizer. Consider the examples (49-53) which illustrate this observation:

- 49. bhaya paDu 'to fear'
  fear   fall

- 50. bAdha paDu 'to suffer'
  suffer   fall

- 51. kaSTAlu paDu 'to face hardships'
  hardships   fall

- 52. siggu paDu 'to be ashamed'
  shame   fall

- 53. niMda paDu 'to be blamed'
  blame   fall

The examples sited in 6 to 59 demonstrate the approach we would like to adopt. As illustrated above specification of a noun for its semantic feature(s) will help in determining the syntactic and semantic behavior of the verb. This may also help us in drawing generalizations like: When a N with a specific semantic feature combines with a given verb then the outcome is either a compositional or a non-compositional complex predicate.

Nouns identified with specific semantic features enable one to predict the meaning of the complex predicate.

5 Conclusion

We may draw the following conclusions based on the dicussion in section 4.We need to adopt a hybrid approach which combines both List-based and Rule-based approaches, while handling complex N+V constructions. In the case of synthetic constructions the semantics of the noun and the verb undergoes a complete change and these constructions become Idioms. For example: *kannu koTTu* 'to wink' which literally means 'eye hit'. Such compounds have to be listed in the lexicon. However these constructions may not be very large in any language.

Complex N+V constructions which are regularly and productively formed like *bhaya paDu* 'to be scared' can be represented in the lexicon using the rule-based approach.

If wordnets have words populated with semantic features it would be redundant to list exhaustively all such complex predicates with their meanings. Hence, the semantic properties of the head (verb) and the non-head (noun) may be listed in the lexicon and the appropriate sense of the complex predicate may be realized automatically from the rules.

References

Beckwith, R., Miller,G. A. 1993. *Design and Implementation of the WordNet Lexical Database and Searching Software.* [ftp.cogsci.princeton.edu/pub/wordnet].

Fellbaum, C., Gross, D., and Miller, K. 1993. *Adjectives in WordNet.* [ftp.cogsci.princeton.edu/pub/wordnet].

Fellbaum, C. 1993. *English Verbs as a Semantic Net.* [ftp.cogsci.princeton.edu/pub/wordnet].

Hook, P.E. 1974. *The Compound Verb in Hindi.* Ann Arbor University of Michigan, USA.

Hook, P.E. 1993. *Aspecto Genesis And The Compound Verb In Indo-Aryan. In Verma.* M.K. (ed.) Complex Predicates In South Asian languages. pp. 97-113.

Krishnamurti, Bh. 1992. *Complex Predicates in Telugu.* In Bulletin of The Deccan College Post Graduate and Research Institute. S.M. Katre felicitation volume. pp. 314-327.

Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D., and Miller, K. 1993. *Introduction to WordNet: An On-line Lexical Database.* [ftp.cogsci.princeton.edu/pub/wordnet].

Miller, G. 1993. *Nouns in WordNet: A Lexical Inheritance System.* [ftp.cogsci.princeton.edu/pub/wordnet].

Rajyarama. 1998. *A Study On Some Aspects Of Derivational Morphology In Telugu With Special Reference To Compounds.* Ph.d. Thesis submitted to University of Hyderabad.

Uma Maheshwar Rao, G. 1994. *On certain Aspects Of Word Formation Processes : A Case Study Of Telugu.* Paper presented in National Seminar on Word Formation in Indian Languages. Osmania University, Hyderabad,February 9-10,1994.

Verma, M. K. 1993. *(ed.) Complex Predicates in South Asian languages.* Manohar Publishers and Distributors. New Delhi.