Understanding Verbs based on Overlapping Verbs Senses

Kavitha Rajan
Language Technologies Research Centre
International Institute of Information Technology Hyderabad (IIIT-H)
Gachibowli, Hyderabad. 500 032.
AP. India.
kavitha@research.iiit.ac.in

Abstract

Natural language can be easily understood by everyone irrespective of their differences in age or region or qualification. The existence of a conceptual base that underlies all natural languages is an accepted claim as pointed out by Schank in his Conceptual Dependency (CD) theory. Inspired by the CD theory and theories in Indian grammatical tradition, we propose a new set of meaning primitives in this paper. We claim that this new set of primitives captures the meaning inherent in verbs and help in forming an inter-lingual and computable ontological classification of verbs. We have identified seven primitive overlapping verb senses which substantiate our claim. The percentage of coverage of these primitives is 100% for all verbs in Sanskrit and Hindi and 3750 verbs in English.

1 Introduction

Communication in natural language is simple. Looking at the ease to learn and communicate in and across natural languages, the claim of existence of interlingual conceptual base (Schank, 1972) seems plausible.

Conceptual Dependency (CD) theory tried to represent a conceptual base using a small set of meaning primitives. To achieve this goal, they put forward a proposal consisting of a small set of 12 primitive actions, a set of dependencies which connects the primitive actions with each other and with their actors, objects, instruments, etc. Their claim was that this small set of representational elements could be used to produce a canonical form for sentences in English as well as other natural languages. Representational theories like Scripts, Plans, Goals and Understanding(SPGU) representations (Schank and Abelson, 1977) were developed from the CD theory. None of the descendant theories of CD could focus on the notion of 'primitives' and the idea faded in the subsequent works.

Identification of meaning primitives is an area intensely explored and a vast number of theories have been put forward, namely, (PRO: Conceptual semantics (Jackendoff, 1976), Meaning-text theory (Mel’čuk,1981), Semantic Primes (Wierzbicka, 1996), Conceptual dependency theory (Schank, 1972) Preference Semantics (Wilks, 1975) CONTRA: Language of Thought (Fodor, 1975)). Through our work, we put forward a set of seven meaning primitives and claim that the permutation/combination of these seven meaning primitives along with ontological attributes is sufficient to develop a computational model for meaning representation across languages.

This paper looks at the Conceptual Dependency Theory created by Roger Schank (Schank, 1973; Schank, 1975) and compares it with theories in Indian grammatical tradition. We discuss these in section 2 and section 3. We then analyze if we can modify Schank's approach to define a more efficient set of primitives. We conclude by introducing the small set of meaning primitives which we have found to cover all verbs in Indian languages like Sanskrit, Hindi and almost all verbs in English.

2 Conceptual Dependency

According to Schank, linguistic and situational contexts in which a sentence is uttered is important for understanding the meaning of that sentence. The CD theory was developed to create a theory of human natural language understanding. The initial premise of the theory
is: basis of natural language is conceptual. According to the theory, during communication, to-and-fro mapping happens between linguistic structures and the conceptual base through concepts. It is due to the existence of this conceptual base and concept based mapping that a person, who is multilingual, is able to switch between languages easily.

The conceptual base consists of concepts and the relations between concepts. Therefore, it is responsible for formally representing the concepts underlying an utterance. There are three types of concepts: a) nominal; b) action and c) modifier. We will concentrate only on 'action' since our work is related to verbs.

CD's basic premise is that the ACTION is the basis of any proposition that is not descriptive of a static piece of the world. Conceptualization consists of action, actors and cases that are dependent on that action. An ACTOR is defined as an animate object and an OBJECT as any concrete physical entity. CD representations use 12 primitive ACTs out of which the meaning of verbs, abstract and complex nouns are constructed.

Primitives are elements that can be used in many varied combinations to express the meaning of what underlies a given word. In CD, primitives were arrived at by noticing structural similarities that existed when sentences were put into an actor-action-object framework. Using these acts, set of states and set of conceptual roles, it is possible to express a large amount of the meanings expressible in a natural language.

3 Indian grammatical tradition

The Nirukta\(^1\) (Sarup, 1920; Kunjunni et al., 1990) statement "Verbs have operation as its predominant element" proposes that "process" is the most important element in a verb. As all words can be derived from verbal roots, we can say that words in a natural language are either activities (verbs) or derived from some activity (nouns). For example:

- rājā (king) is derived from (the root) rāj (to shine)
- vṛkṣa (tree) is derived from (the root) vṛ (to cover) kṣa (the earth)

Verb is called kriyā in Sanskrit. kriyā stands for action or activity. Verbs consists of both action and state verbs. Sage Kātyāyana (3rd century BC) put forward the bhāvo-based definition to define all types of verbs. According to Nirukta verse 1.1 (Sarup, 1920) the characteristic that defines a verb form is its verb having bhāva as its principal meaning. In Sanskrit, bhāva is a morphological form of bhavati and bhavati means 'happening'. So structure of bhāva can be defined as structure of happening which is explained in section 4.1.

According to sage Vāryāyaṇi, Nirukta verse 1.2 (Sarup, 1920), there are 6 variants of bhāva or verb which, we believe, can be compared to 6 fundamental processes. That is, a process ‘verb’ consists of six stages. They are:

- coming into being - jāyate
- existing - asti 'is'
- changing - viparītamāte
- increasing - vardhate
- diminishing - apaśkṣyate 'diminishes'
- ceasing to be - vinaśyati 'perishes'

4 Our Approach

We are trying to use existing theories in the traditional school of Sanskrit language, namely, Navya-Nyāya for identification and formal representation of primitive actions. We work within the formal framework of Neo- Vaiśeṣika Formal Ontology (NVFO)\(^2\).

4.1 Form of verb

Happening is formally conceived as punctuation between two discrete states in a context. Since every happening consists of minimally two different states, there is an atomic sense of movement in it. Movement means whenever an action takes place two states come into existence. The initial state, at the beginning of an action and a final state, after the completion of the action. The two states can be same or different. Time is an inseparable part of this structure because between initial and final states there can be \( n \) number of intermediate states which are sequential.

Happening (Sanskrit, bhavati) is the change of state from one to another in a context. According to Bhartṛhari (5th century CE) every verb has

---

\(^1\) Nirukta (Kunjunni et al., page-88).

\(^2\) Vaiśeṣika ontology, due to Kaṇṭhā (Rensink, 2004), Prasārapāda (Hutton, 2010) and Udayana (Kaṇṭhā, 1986) has been formalized by Navjyoti (Tavva and Singh, 2010).
'sense of sequence' and 'state' in it. Hence, every verb projects a 'sense of happening', making this sense omnipresent in all verbs. Therefore, bhavati is a 'universal verb'. In the nominalization of bhavati, 'bhāva' has a formal structure and has been named 'punct'. The formal representation of bhāva is shown in Figure1.

The structure of (universal verb) 'punct' is:

\[
\langle \text{state1} / \text{state2}, \text{(Context)} \rangle \text{ Feature Space }
\]

From Sanskritist tradition, we have adopted the concept of universal verb. Our original contribution is that we have defined an ontological structure (see Figure1) to represent 'universal verb' and have used it to represent the seven primary verb senses (primitives) which we have identified. All verbs in a language can be represented formally using this structure.

### 4.2 Identifying Overlapping Verbal senses

Can we have a few number of primitive meaning senses whose permutation / combination will enable us to explain all meanings in a language? Primitive verb senses in language were identified using an approach similar to Lesk's method (Lesk, 1986) of finding meaning overlaps for solving Word Sense Disambiguation problem.

All verbs and definitions of all senses of each verb in Sanskrit (2500) and 3750 verbs in English were collected. The verb senses were collected from various on-line dictionaries in both the languages. From these definitions, verbs which are used to explicate defined verbs were identified. The procedure followed for identifying frequent verbs is explained using a sample verb 'fall'.

Definitions of different verb senses of 'fall' from two different sources are given below:

Source 1 (Dictionary.com):
- (to drop or descend under the force of gravity, as to a lower place through loss or lack of support), (to come or drop down suddenly to a lower position, especially to leave a standing or erect position suddenly, whether voluntarily or not), (to become less or lower; become of a lower level, degree, amount, quality, value, number, etc.; decline)

Source 2 (WordNet):
- (descend in free fall under the influence of gravity), (decrease in size, extent, or range), (move downward and lower, but not necessarily all the way), (move in a specified direction), (lose an upright position suddenly), (drop oneself to a lower or less erect position) are few senses.

All words in bold represent 'movement' in a negative manner. Since movement is the most common concept, 'move' is taken as an overlapping primitive verb sense. Other primitives like know, do, is, have, cut, and cover were obtained by similar procedure.

In dictionaries, overlapping verb senses used to explicate meaning of defined verbs, show the relatedness of two verbs. The phenomenon known as 'Dictionary circularity' (Wierzbicka, 1996) confirms the existence of this claim.

In WordNet, the existence of most frequently used verbs is represented through 8 'common verbs' (Miller et. al, 1990): have/ has, be, make, run, set, go, take and get. State is dealt with separately in WordNet. We have modified the 'common verbs' concept of WordNet to include the concept of verbality – the ability to denote a process developing in time (Lyudmila, 2010).

To analyze the phenomena of overlapping meanings of verbs, we studied verbs from a database of 3750 verbs and two other lexical resources:WordNet, Webster English Dictionary. From the word frequencies of the verbs in these three resources, we calculated the percentages of overlapping verb senses used to explicate meaning of defined verbs. The results are shown in Table 1. Total verbs (unique word forms) in the three resources –

---

3 Bhāva is defined by Patañjali as (1)existence, (2) something that comes into being, and (3) something that is brought into being.

4 The formalization in NVFO is based on the idea of an ontological form which is recursive. This form is called 'punct'. Using punct's categories of Vaśesika ontology can be derived.

5 Percentage is calculated taking the frequency of a verb w.r.t the total verbs in the particular source.
When verbs and their definitions in English language were analyzed it was found that basic verb senses like 'know', 'do', 'have', 'move', 'is', 'cut', and 'cover' have higher frequency. The occurrence of higher frequencies of some verbs indicated that those were the verbs with maximum meaning sense overlap with other verbs.

### 4.3 The Seven Puncts

In order to handle similarities and overlaps in meaning we have developed the concept of overlapping verbal sense or 'punct'. These primitive verbal senses are intended to be building blocks out of which meaning of verbs can be constructed. We have identified seven 'puncts'. Two works WordNet (8 common verbs) and Nirukta (6 fundamental processes) were influential in restricting the number of overlapping verb senses to 7. We have modified the 8 common verbs in WordNet (have, be, get, set, make, do, run, take) in a way that each primitive meaning sense can be represented as a combination of 'state' and 'change'. Concepts like exist and un-exist, join and un-join, know and un-know, do and un-do, ascribing some actions to some objects and un-ascribe, movement / change and possess and un-possess are the basic meaning senses we have identified. 'un' stand for opposite here. Each primitive meaning sense consists of a sense and its negation. We have seen that verbs across languages can be classified using this seven primitives. Percentage of coverage of these seven primitives in Sanskrit and English are given in Table 2.

| Puncts | Percentage in English Verbs | Percentage in Sanskrit Verbs |
|--------|----------------------------|----------------------------|
| Know   | 4.96                       | 4.27                       |
| Move   | 17.69                      | 12.41                      |
| Do     | 58.90                      | 56.99                      |
| Have   | 4.12                       | 7.79                       |
| Is     | 6.36                       | 7.41                       |
| Cut    | 3.22                       | 7.06                       |
| Cover  | 4.75                       | 4.07                       |

Table 2. Percentage of coverage of the seven verb senses (puncts) in English & Sanskrit

Using this set of 7 'puncts' it is possible to express meaning inherent in verbs in a language and also to link the related verbs across languages. We will explain this by a deeper analysis of the seven 'puncts' (see Table 3).

The 'punct' can be used for identifying similarities between verbs like 'fall', 'plummet', 'flow' all of which have 'move' as primary sense and they can be used for finding out different senses of the same verb like 'break'. Thus 'break' can have primary sense of 'cut' and secondary sense of 'do' when the meaning is 'to destroy or stop or interrupt or cause something to separate something'. Similarly, 'break' can also have 'move' as primary sense and 'is' as secondary sense when the meaning is 'voice change of a person or day or dawn break or breaking news'. Though a verb can have two to all seven verbal senses, we are grouping verbs looking at just the primary and secondary verb senses. A verb can be in more than one group. Once they are classified according to their primary and secondary meanings we put verbs in groups, say all verbs having 'move' as primary sense and 'do' as secondary sense will be in a group.

| Punct (Elementary Bhāva-s) | Explanation |
|---------------------------|-------------|
| Know: Sense of knowing    | Know / Knower Conceptualize, construct or transfer information between or |

---

*A verb can be explicated by more than one verb (overlapping meaning component) hence the total of the percentages of the verbs, which have been identified as the overlapping components is not 100.*
5 Comparison of primitives

A comparison of primitives of CD theory and our approach is given in Table 4. Corresponding to each ACT of CD theory the explanation and Puncts in order of priority of meaning senses is given.

| ACT     | Explanation about ACT                                                                 | PUNCTS in order of meaning sense |
|---------|---------------------------------------------------------------------------------------|----------------------------------|
| ATRANS  | Transfer of an abstract relationship such as possession or control (give)              | Do / Have / Cut                  |
| PTRANS  | Transfer of the physical location of an object (go)                                     | Do / Move / Cut                  |
| PROPEL  | Application of a physical force to an object (push)                                    | Do / Move / Cut                  |
| MOVE    | Movement of a body part of an animal by that animal (kick)                             | Do / Move                        |
| GRASP   | Grasping of an object by an actor (grasp)                                              | Do / Have / Cut                  |
| INGEST  | Taking in of an object by an animal to the inside of that animal (eat)                  | Do / Have / Move / Cut           |
| EXPEL   | Expulsion of an object from the object of an animal into the physical world (cry)       | Move / Do / Is                   |
| MTRANS  | Transfer of mental information between animals or within an animal (tell)              | Do / Know / Move                 |
| MBUILD  | Construction by an animal of new information of old information (decide)                | Know / Do / Cover / Move          |
| CONC    | Conceptualize or think about an idea (think)                                           | Know / Do / Move                 |
| SPEAK   | Actions of producing sounds                                                            | Do / Move                        |

Table3. Puncts

We believe that every word is distinct. 'There are no real synonyms and that no two words have exactly the same meaning' (Palmer, 1986 page-89). If all words are distinct how can we show its distinctness? We have observed that there is at least one ontological attribute which makes each word different from the other. They are called ontological attributes as they are concepts like space, time, manner, reason and sub-features like direction-linear, source, destination, effect etc. which can be represented inter-lingually. We have named the set of attributes as ‘feature set’. Feature set is a part of the context C defined in the structure of ‘punct’. Verbs with same feature set across languages can be cross-linked. For example, if we want to represent verb ‘breathe’ in another language, we just have to map the attributes identified for ‘breathe’ which are – breathe1) move, instrument-lungs, object-air, manner-into and out of breathe2) say, object-something, manner- very quietly breathe3) open, object-wine bottle, duration- short time, purpose-improve flavor.
ATTEND  Action of attending or focusing a sense organ towards a stimulus (listen)  Know / Do

| (say) |
|-------|
| ATTEND |

Table 4. Comparison of ACT and Punct.

6 Issue and Solution

The uniform identification of verb sense means identifying the most general sense attached to a verb, as done by an ordinary person. One can see that more than one verb can be used to explicate the meaning of a verb and there is an order in which the verbs are used. This order helps in finding the primary, secondary and tertiary meaning senses. The order is found by nominalizing verbs in a simple sentence. This method helps in resolving inconsistencies, if any, while identifying meaning senses. For example:

- you confuse me -> you create {confusion in me}
- You create {{confused (state of Knowledge) about something (object of knowledge)} in me}
- {You do creation of} {{‘Confused (state of Knowledge) about something (object of knowledge)} in me}.

In the last sentence: ‘do’ is tertiary sense, ‘know’ is secondary sense and ‘is {state of knowledge}’ is the primary sense of verb ‘confuse’.

The seven verb senses thus identified are the building blocks out of which meanings of verbs are constructed. The primary and secondary senses of all verbs in English and Sanskrit were identified. For English verbs, the entire verb list (3750) enlisted by Levin (Levin, 1993) including extensions (Dang et. al, 1998; Kipper et. al, 2006; Korhonen and Briscoe, 2004) was classified according to the new classification. For Sanskrit verbs, data (more than 3000 verbs (Sanskrit धातु) including variations in accentuation) was collected from various resources (Palsule, 1955; Palsule, 1961; Liebich, 1922; Varma, 1953; Kale, 1961; Apte, 1998; Williams, 2008; Capeller, 1891). The meanings of English verbs were obtained from various dictionaries (on-line English dictionaries) and the senses were identified based on intuition. The annotation process was to identify the primary and secondary meaning senses of all verbs and ontological attributes of verbs in 7 groups (all verbs with the same primary verb senses formed one group). The annotation of verbs was done for four languages: Sanskrit, English, Hindi and Telugu. Verbs in Sanskrit and English were compiled and annotated by one trained annotator and cross-checked by an equally trained second annotator. The differences in annotation, around 10%, were resolved by discussion. Annotation in Hindi and Telugu was done by 9 and 25 annotators respectively. The annotators were humans and native speakers of their languages, having an idea of the new approach. The average ratio of correctness was 64%. The classification was done manually.

Based on this classification the verb groups formed have exhibited similarity in syntactic and semantic behavior. The pattern of Stanford dependency relations formed among verbs of same groups showed a similarity of 60%. This similarity in relations were used to form WSD rules which helped in increasing the accuracy of English to Hindi Anusaaraka8 Machine Translation system output by 36.04%.

7 Conclusion

Conceptual Dependency theory was based on two assumptions:
1. If two sentences have same meaning they must have similar representation regardless of the words used.
2. Information implicitly stated in the sentence should be stated explicitly.

Our approach is based on two assumptions:
1. There is a conceptual base underlying all natural languages.
2. All content words are derived from verb root. ‘Punct’ is a mathematical representation of conceptual base in terms of state and change which can be used for computational purpose. Identification of overlapping verbal sense enables a classification based on meaning. Verbal sense identification along with feature space which includes ontological attributes can give a better classification and understanding of verbs and their behavior. Application of the concept of ‘punct’ in NLP applications like

---

8 http://anusaaraka.iiit.ac.in
machine translation has shown to increase its performance by 36.04%.

8 References

Anna Korhonen and Ted Briscoe. 2004. *Extended Lexical-Semantic Classification of English Verbs*. Proc. of the 42nd Meeting of the ACL, Workshop on Computational Lexical Semantics.

Anna Wierzbicka. 1996. *Semantics: Primes and universals*. Oxford: Oxford University Press.

Arend Rensink. 2004. *GROOVE: GRaphs for Object-Oriented VERification*. http://groove.cs.utwente.nl/.

Beth Levin. 1993. *English Verb Classes and Alternation, A Preliminary Investigation*. The University of Chicago Press.

Bruno Liebich. 1922. *Materialien zum Dhatu patha*, Heidelberg. Carl Winter’s University.

Capeller. 1891. *Sanskrit-English Online Dictionary*. Retrieved from http://www.sanskrit-lexicon.uni-koeln.de/scans/MWScan/tamil/index.html

Frank Robert Palmer. 1976. *Semantics*. Cambridge: CUP.

Hoa Trang Dang, Karin Kipper, Martha Palmer, and Joseph Rosenzweig. 1998. *Investigating regular sense extensions based on intersective Levin classes*. Proc. of the 36th Meeting of the ACL and the 17th COLING.

Gajanan Balkrishna Palsule. 1955. *A Concordance of Sanskrit Dhatupath*, Deccan College Dissertation Series, Bhandarkar Oriental Research Institute, Poona.

Gajanan Balkrishna Palsule. 1961. *The Sanskrit Dhatupathas*, University of Poona.

Graham Hutton. 2010. *Introduction to Categories* Lecture notes. University of Birmingham, 23-27 April.

George A. Miller, Richard Beckwith, Christiane Fellbaum, Derek Gross, and Katherine Miller

. 1990. *Introduction to WordNet: An On-Line Lexical Database*, Int'l J. Lexicography, vol. 3, no. 4, pp. 235-244.

Igor A Mel’čuk. (1981). "Meaning-Text Models: A recent trend in Soviet linguistics". *Annual Review of Anthropology* 10: 27–62.

James R. Hurford. 2007. *The Origins of Meaning: Language in the Light of Evolution*, Oxford University Press

Jerry A Fodor. 1975. *The Language Of Thought*. Crowell Press. pp 214.

John W M Verhaar,1966. *The Verb 'Be' and Its Synonyms*, Foundation of Language Supplementary Series. Springer.

Kaṇāda. 1986. *The Vaśesika sutras of Kaṇāda with the commentary of S amkara Mis'ra and extracts from the gloss of Jayanārāyaṇa*. Translation in English by Nandalal Sinha. Allahabad (1911); Delhi (1986).

Karin Kipper Schuler. 2005. *VerbNet: A Broad-coverage, Comprehensive Verb Lexicon*. PhDDissertation, University of Pennsylvania.

Kunjumni. K. Raja and Harold G. Coward. 1990. *Encyclopedia of Indian Philosophies: The philosophy of the grammarians*, Volume 5. New Delhi, India: Motilal Banarsidass. p. 324.

Lakshman Sarup. 1920. *The Nighantu and Nirukta*. Motilal Banarasidass, Delhi.

Lyudmila Osinovskaya. 2010. *Verb Classification*. Tyumen State University.

Michael E. Lesk. 1986. *Automatic sense disambiguation using machine readable dictionaries: how to tell a pine cone from an ice cream cone*. In SIGDOC ’86: Proceedings of the 5th annual international conference on Systems documentation, pages 24-26, New York, NY, USA. ACM.

Monier Williams. 2008. *‘Sanskrit English online Dictionary’*, retrieved from http://www.sanskritlexicon.uni-koeln.de/monier/

Moreshvar Ramchandra Kale 1962. *A higher Sanskrit grammar, for the use of schools and colleges*. Online. Retrieved from http://ia700307.us.archive.org/35/items/highersanskritgr00kaleuoft/highersanskritgr00kaleuoft.pdf.

Philip D. Morehead. 2001. *The New American Webster Handy College Dictionary*. Signet Book. Fourth Edition.

Rajesh Tavva and Navjoti Singh. 2010. *Generative Graph Grammar of Neo-Vai esika Formal Ontology*. In G.N.Jha, editor, *Sanskrit Computational Linguistics*, pages 91–105. Springer.
Ray Jackendoff. 1976. Toward an explanatory semantic representation. *Linguistic Inquiry* 7 (1): 89–150.

Roger Schank. 1972 *Conceptual Dependency: A Theory of Natural Language Understanding*. Cognitive Psychology 3, pp. 552-631.

Roger Schank. 1973. *Conceptualizations underlying natural language*. In *Computer Models of Thought and Language*, R. Schank & K. Colby, eds. San Francisco: W.H. Freeman.

Roger Schank. 1975. *The Primitive ACTs of Conceptual Dependency*. Yale University. New Haven CT. TINLAP’75. Proceedings of the 1975 workshop on Theoretical issues in natural language processing. Pages 34-37.

Roger Schank and Robert Paul Abelson. 1977. *Scripts, Plans, Goals, and Understanding*. Lawrence Erlbaum Associates. Hilldale NJ.

Siddheshwar Varma. 1953. *The Etymologies of Yaska*. Vishveshvaranand Institute Publications.

Vaman Shivram Apte, 1998. *Apte Sanskrit Dictionary*, Retrieved from http://www.aa.tufs.ac.jp/~tj/un/skt dic/.

Yorick Wilks, 1975 An intelligent analyzer and understander of English. *Comm. Assn. Comp. Mach.* 18, 264-274.