Identification of Spatio-Temporal Attributes for Hindi Verbs based on Formal Ontology of Language

Abstract. The notion of space and time in a language has been recognised as a fundamental unit of cognitive primitive. Several studies have been done in order to extract spatio-temporal information using verbs in a language. This paper uses Formal Ontology of Language to study relations between verbs and their spatial, temporal attributes. Formal Ontology of Language proposes that meaning of words are *in-formed* by intrinsic and extrinsic ontological structure. A lexical resource, namely OntoSenseNet was developed for English, Hindi and Telugu in which different parts-of-speech have been annotated based on the Formal Ontology of Language. It is still a work in progress. Using OntoSenseNet, the spatial and temporal attributes for verbs have been identified based on kārak (case relation), sense-classes for adverbs and Tense-Aspect-Modality. These attributes have been identified in Hindi Literature Corpora. We show that different verb sense-types exhibit different relations with spatial and temporal attributes.

1 Introduction

Space and time play a pivotal role in human thinking and cognition. Their role in language and literature has intrigued many. Studies have been done in order to identify their conceptualization in different languages[1]. Spatial and temporal aspect in a language has also been studied using verbial aspect[2]. Linguistically, spatial and temporal features can be understood by identifying relations between verbs and other constituents in a sentence.

In this paper, we have made an attempt to study the spatial, temporal features using ontological attributes of verb. In order to identify these attributes, Formal Ontology of Language as proposed by Otra[3] has been used. In the past, several theories have been proposed to identify the meaning primitives. Conceptual Semantics[4], Semantic Primes[5], Conceptual Dependency Theory[6] and Language of Thought[7] are such attempts. The proposed Formal Ontology of Language uses intrinsic meaning of words to identify meaning primitives. In linguistic traditions of India, verbs have been regarded as the centre of a language. According to the proposed theory, the maximum overlapping sense of verbs is *happening*. In this proposal, seven primitive senses of verb have been identified. These fundamental senses of verbs give different contiguity of states. Meaning of a verb can be explicated using these primitive senses and its ontological attributes. These ontological attributes are derived from the context the verb is used. The ontological attributes are concepts like space, location, time, manner

etc. Considering verb at the center of a sentence and identfying ontological relation between different consituents, the intensional meaning of the sentence can be understood.

Verbs have been organised as sense-types. Adverb being a modifier of verb plays an important role in unfolding its meaning. Otra[3] proposes four fundamental sense-classes of adverbs and six sense-types of adjective. She has also outlined seven/eight different types of verb-noun pairings and twelve different types of noun-adjective pairing. This paper is a part of an ongoing effort that aims at developing verb-centric ontological resource, namely OntoSenseNet for Indian languages, using Formal Ontology of Language. The resource contains annotated lexicons for Hindi and Telugu.

The paper focuses on the identification of spatial and temporal ontological attributes of verb and statistically studying their frequency distribution in a corpus. Space and time play a key factor in articulating any narrative. These are strongly captured in literature and are representative of a human experience. To understand these spatial and temporal attributes, we have used Hindi literature corpora.

Section 2 of the paper talks about the proposed theory of sense-types of verbs and sense-classes of adverbs. It lists other kāraka/case relations as another kind of ontological points. Sections 3 outlines spatial and temporal attributes of verbs. Section 4 discusses corpora preparation and tool used for the analysis. Section 5 depicts the results obtained from the corpora analysis. Section 6 talks about Conclusion and Future Work.

2 Sense-Type and Sense-Class

Otra derives [3] two logical forms of intension that is *type* and *class*. In this proposal, types are derived from 'form of transference' whereas classes are derived from 'form of abstraction.' Verbs being collocative in nature, that is, many verbs can be used to describe the same situation and same sense point can be shared across different verbs have been rendered as 'sense-type'. Seven fundamental sense-types of verbs have been derived. Table 1 shows seven sense-types of verb along with their fundamental senses.

Adverbs on the other hand are used to modify verbs and do not have an overlapping structure. Hence, it has been classified as 'sense-class'. Four different sense-classes of adverbs have been delineated. Following are the four sense-classes cited with English and Hindi examples.

1. Temporal - Adverbs that attributes to sense of time. e.g Now; āj (today)
2. Spatial - Adverbs that attributes to physical space. e.g Upward; abhimukh (In front of somebody)
3. Force - Adverbs that attributes to cause of happening e.g. Freely; amdhādhumdha (blindly)
4. Measure - Adverbs dealing with comparison, judgement. e.g - Few; adhik (More)

Table 1. Sense-type of verbs

Sense-Type	Primitive Sense
Means|End	Do
Before|After	Move
Know|Known	Know
Grip|Grasp	Have
Wrap|Wrapped	Cover
Part|Whole	Cut
Locus|Located	Is

Apart from the type and class relation, the proposed theory also talks about seven/eight quasi-types of thematic roles between nouns and verbs. These roles are relational context of the point between a verb and a noun. The resource(OntoSenseNet) adopts these thematic roles as ontological points. Otra [3] describes following different thematic roles.

1. Agent: The initiator of some action
2. Patient: The entity undergoing the effect of some action.
3. Theme: The entity which is moved by an action.
4. Experiencer: The entity that receives sensory or emotional input
5. Beneficiary: The entity for whose benefit the action was performed.
6. Instrument: The means by which an action is performed.
7. Location: The place in which something situated or takes place.
8. Goal: The entity towards which something moves.
9. Source: The entity from which something moves either literally or metaphorically.

Computational Paninian Grammar Framework define kāraka relations as syntactico-semantic relations between verbs and other constituents (generally verb) in a sentence[8].

We have used Hindi dependency parser[1] to extract these depedency relations. The dependency parser uses guidelines from AnnCorra [9] for annotations. It is treebanks for Indian Languages Guidelines for annotating Hindi Treebank. The annotation is stored in SSF format [10]. It identifies other dependency relations apart from kāraka relations. Some of the kāraka and other dependency relations are denoted by following notations in the Hindi Treebank.

1. *k1* - Agent(kartā)
2. *k1s* - Noun complement of kartā (vidheya kartā - kartā samanadhikarana)
3. *k2* - Object/Patient(karma)
4. *k2p* - Goal, Destination
5. *k3* - Instrument (karna)
6. *k4* - Recipient (samprādana)

---

[1] https://bitbucket.org/iscnlp/

7. *k4a* - Experiencer (anubhava kartā)
8. *k5* - Source (apadāna)
9. *k7* - Location elsewhere (vishayadhikarana)
10. *k7p* - Location in space (desadhikarana)
11. *k7t* - Location in time (kālādhikarana)

3 Spatial, Temporal Attributes of Verbs

In the previous section, we discussed different kinds of ontological attributes(sense-class, quasi-types) that help explicate meaning of a verb. There are spatio-temporal attributes in these ontological points. Broadly they can be understood as Tense-Aspect-Modality(TAM) features of verbs, Spatio-Temporal Nouns, Spatio-Temporal kāraka or case relation. Spatio-Temporal Adverbs. This paper studies the frequency distribution of these spatio-temporal attributes with different sense-types of verbs.

4 Method

4.1 Corpus Preparation

We carried out statistical study based on the frequency distribution of sense-types of verbs with different spatio-temporal attributes in Hindi Literature Corpora. It was collected from Hindi Samay website[2]. Table 2 shows different corpus collected from writings of different authors.

Table 2. Different corpus types collected for different authors

Author	Corpus-Type	Total Number
Premchand	Novels	10
Premchand	Stories	54
Agey	Novels	2
Jay-Shankar Prasad	Novels	2
Jay-Shankar Prasad	Stories	28
Sharat Chandra	Novels	2

The total number of sentences in the total corpus were 3,20,074.

---

[2] http://www.hindisamay.com/

4.2 Dependency relation and morphological feature extraction

We used morph analyzer[3] and POS Tagger[4] of Hindi to extract Tense-Aspect-Modality of verbs. In order to extract dependency relations, Hindi dependecy parser was used. After extracting morphological and dependency features, the primary sense-types of verbs were extracted from the OntoSenseNet resource. The frequency distribution of each of the verb sense-types with different spatio-temporal attributes were calculated.

5 Results and Discussion

5.1 Tense-Aspect-Modality

Tense-Aspect-Modality are important morphological features of a verb. They not only specify temporal aspect but also tells about the status or ability to perform an action. They are combinations of inflections and auxillary verbs or modals or words indicating negativity. We used morph analyzer of Hindi to extract Tense-Aspect-Modality(TAM) markers of verbs. The sense-types of these verbs were further identified from OntoSenseNet. Figure 1 presents the frequency distribution of the sense-types of verbs with their TAM markers. It is interesting to note that 'WA'(thA,was) and 'hE'(hai,is) have occured only with Locus|Locate sense-types of verbs. On the other hand 'kara' which roughly translates to "after"$<verb>$ + "ing" in English. has occurred the least with all the verb sense-types.
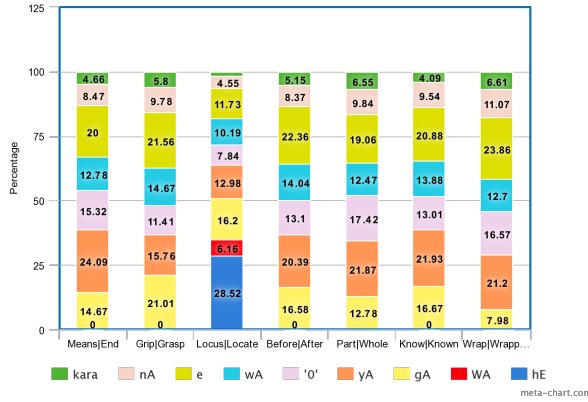

Fig. 1. Frequency Distribution of TAM features for Verb sense-types

---

[3] https://ltrc.iiit.ac.in/showfile.php?filename=onlineServices/morph/index.htm

[4] http://ltrc.iiit.ac.in/analyzer/hindi/

5.2 Kāraka relations of Verb sense-types with Spatial and Temporal Nouns

Indian languages contain content words that denote time and space. They can be present as a spatial and temporal arguments of a verb along with an appropiate case marker. These nouns are marked as NST [11] which are spatial and temporal nouns.

To understand relation between verbs and NST nouns, kāraka relations have been used.

We extracted the kāraka relations between the spatial-temporal nouns and verbs by using full dependency parser for Hindi. The frequency distribution of sense-types of each verb with the kāraka relations with spatial/temporal nouns were tabulated. Figure-2 shows the frequency distribution for the same. From
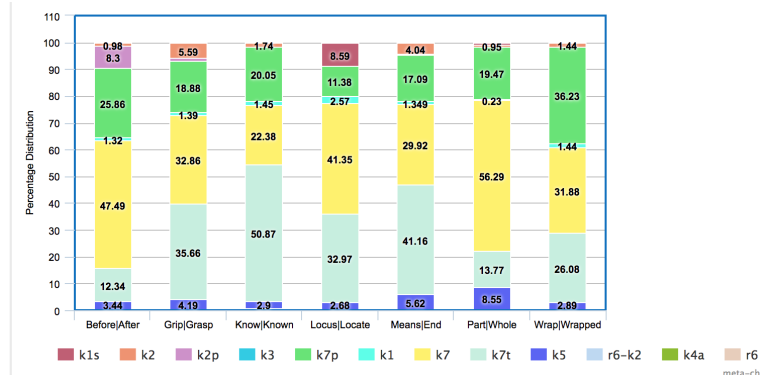

Fig. 2. Frquency Distribution of kāraka relation of Verb sense-types with NST nouns

the figure it can be seen that only verb of sense-type Locus|Locate has occurred with NST nouns with 'k1s' ((vidheya karta - karta samanadhikarana, 'noun complement of karta')) relation. 'samanadhikarana' indicates having the same locus. Hence, 'karta samanadhikarana' indicates having the same locus as 'karta' (doer, agent).

Before|After sense-types are the only verbs that have occured with NST nouns with 'k2p' kāraka relation. 'k2p' case marker are defined to be the goal or destination where the action of motion ends and are mostly the objects of motion verbs. This is clearly evident even at the fundamental ontological form.

5.3 Spatial and temporal kāraka or case markers

There are nouns that carry k7p (location in space) and k7t (location in time) kāraka relations with verbs. These were also incorporated to study the distribution of verb sense-types in this paper. Figure-3 demonstrates the frequency
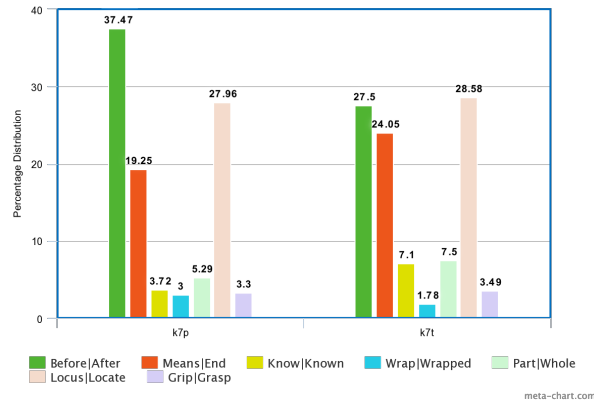
Fig. 3. Frequency Distribution of Verb sense-types with spatial(k7t) and temporal(k7p) kārakasfor common nouns

distribution of verb sense-types having spatial and temporal kāraka relation with common nouns. Figure-3, shows association of spatial and temporal karaka relation with nouns for different verb sense-types. It is evident that majorily Before|After sense-type occur with nouns in a spatial/temporal relation.
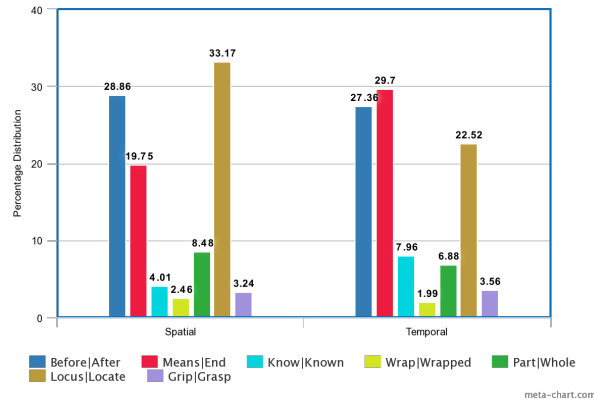

Fig. 4. Frequency Distribution of Verb sense-types and spatial, temporal adverbs

5.4 Relation of Spatial, Temporal Adverbs with different different verb sense-types

Based on the Formal Ontology of Language, adverbs whose sense-class was spatial/temporal were extracted from the corpus. The sense-types of those verbs

which were modified by these adverbs were identified. Figure-4 demonstrates this association. It is interesting to note that Spatial adverbs have mostly modified Locus|Locate types of verbs, whereas Temporal adverbs have modified Means|End types of verbs. Before|After types of verbs have been almost equally modified by Spatial and Temporal adverbs. Wrap|Wrapped types of verbs have been the least modified by Spatial and Temporal adverbs.

From the above discussions, we see that Before|After, Locus|Locate and Means|End types of verbs have more associativity with spatio-temporal attributes. It is evident from there fundamental meanings as well. These verbs have move, do, exist as there primitive sense and thuse would need space and time modifiers to further unfold their meanings.

6 Conclusion and Future Work

In this paper we tried to extract spatial and temporal ontological attributes of verb and studied their associativity. Different verb sense-types exhibits different relations with spatial and temporal nouns, adverbs. It is seen that mostly verbs of sense-types Before|After, Locus|Locate and Means|End have been modified by the spatio-temporal attributes whereas Wrap|Wrapped is the least modified verb type by spatio-temporal attributes. These results provide useful insights as it shows interaction of the verb sense-types with different spatial and temporal components, depicting framework and feature of a language. The spatial and temporal components needs to be sub-classified in order to explicate further meanings of verb. It would be interesting to see the frequency distributions of spatio-temporal attributes of verbs in different languages. Future works involve styding and comparing role of spatio-temporal attributes in different corpus types across different languages. This is an on-going project that can be utilized further for NLP tasks like word sense disamguation, cultural analysis, etc.

References

1. Hendriks, H., Hickmann, M., Demagny, A.C.: How adult english learners of french express caused motion: A comparison with english and french natives. Acquisition et interaction en langue étrangère (2008) 15–41
2. SANKARAN, L.: Chapter four verifying the distributional bias hypothesis: An analysis of tamil lavanya sankaran. Space and Time in Language and Literature (2009) 87
3. Otra, S.: Towards Building a Lexical Ontology Resource Based on Intrinsic Senses of Words. PhD thesis, International Institute of Information Technology Hyderabad (2015)
4. Jackendoff, R.: Conceptual semantics. (1988)
5. Wierzbicka, A.: Semantics: Primes and universals: Primes and universals. Oxford University Press, UK (1996)
6. Schank, R.C.: Conceptual dependency: A theory of natural language understanding. Cognitive psychology **3** (1972) 552–631
7. Fodor, J.A.: The language of thought. Volume 5. Harvard University Press (1975)

8. Bharati, A., Sangal, R.: Computational paninian grammar framework. Supertagging: Using Complex Lexical Descriptions in Natural Language Processing (2007) 355

9. Bharati, A., Sharma, D.M., Husain, S., Bai, L., Begum, R., Sangal, R.: Anncorra: Treebanks for indian languages, guidelines for annotating hindi treebank (version–2.0). LTRC, IIIT Hyderabad, India (2009)

10. Bharati, A., Sangal, R., Sharma, D.M.: Ssf: Shakti standard format guide. Language Technologies Research Centre, International Institute of Information Technology, Hyderabad, India (2007) 1–25

11. Bharati, A.: Anncorra: Annotating corpora guidelines for pos and chunk annotation for indian languages. (2006)