Unveiling Spatial-Temporal Markers and Identifying Adverbial Associations for Different Verb Types in Hindi: Developing OntoSenseNet as a Lexical Ontological Resource

Thesis submitted in partial fulfillment
of the requirements for the degree of

*(Master of Science in **Exact Humanities** by Research*

by

Jyoti Jha
201156091
jyoti.jha@research.iiit.ac.in

International Institute of Information Technology
Hyderabad - 500 032, INDIA
June 2024

International Institute of Information Technology
Hyderabad, India

CERTIFICATE

It is certified that the work contained in this thesis, titled 'Unveiling Spatial-Temporal Markers and Identifying Adverbial Associations for Different Verb Types in Hindi: Developing OntoSenseNet as a Lexical Ontological Resource' by Jyoti Jha, has been carried out under my supervision and is not submitted elsewhere for a degree.

Date

Late Prof. Navjyoti Singh and Prof. Dipti Misra Sharma

To Prof. Navjyoti Singh and my late grandparents

Acknowledgments

Embarking on the journey of writing this dissertation has been a profoundly enriching experience. I am eternally grateful to my advisor, the late Prof. Navjyoti Singh, whose invaluable guidance extended beyond the formulation of research questions and methodology to profoundly shaping my life. This work, and all my future endeavors, are dedicated to his cherished memory.

I would also like to extend my heartfelt thanks to Prof. Dipti Misra. Her unwavering support and insightful counsel throughout my research and beyond have been indispensable. Thank you, ma'am, for being both a mentor and a friend. It has been an immense privilege to work under your tutelage. Your expertise and guidance not only helped me navigate my research but also ignited my passion for Natural Language Processing, which has since become the cornerstone of my professional pursuits.

My deepest gratitude goes to my parents and family. Their steadfast support has been the bedrock of my achievements.

I am also profoundly thankful to the entire CEH and LTRC department community. Special thanks to Tavva Rajesh Sir and Pranav Sir for their unwavering support. I am also grateful to my friends Jaya, Prateek, Vinay, and Tarun, whose constant encouragement kept me motivated.

To all these wonderful individuals, I extend my sincere appreciation for their unique contributions in helping this work reach its fruition.

Abstract

This thesis explores the pivotal role of verbs in the semantics of natural language understanding, highlighting their essential function in expressing actions or states within sentences. Verbs convey the main idea or action and are critical in expressing tense, aspect, and mood, as well as establishing subject-verb agreement. These aspects make verbs central to unfolding sentence meaning.

To computationally specify the meaning of a sentence, the research considers both ontological and linguistic aspects of verbs and other parts of speech. Ontology provides a framework for organizing and classifying the concepts and entities referred to in a sentence, aiding in the precise derivation of meaning. This thesis utilizes various theories of formal ontology, including Montague grammar, situation semantics, generative semantics, cognitive semantics, and verb semantic-based ontology, to deepen the understanding of the relationship between language and meaning.

The primary objective of this research is to develop a resource for the Hindi language using the formal ontology of language This resource is employed to analyze verb-adverb collocations in Hindi, exploring the relations of spatial and temporal markers with ontological categories.

The thesis is organized into seven chapters. Chapter 1 introduces the research problems and objectives. Chapter 2 elaborates on the theoretical framework of verb and adverb ontological classification. Chapter 3 describes the manual annotation procedures for Hindi and provides validation. Chapter 4 discusses the automatic identification of verb sense-types and adverb sense-classes. Chapter 5 presents a corpus-based analysis of verb types with spatial and temporal markers in Hindi. Chapter 6 investigates verb-adverb collocations in Hindi through a corpus-based approach, employing statistical measures like log-likelihood to assess association strength. Finally, Chapter 7 summarizes the research findings and conclusions.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Verb - as a key in semantics

Semantics has been a core component of natural language understanding, and verb plays a critical role in unfolding the meaning of a sentence [35]. It is because they express the action or state of being that is taking place in the sentence. The verb is often the most important word in a sentence because it conveys the main idea or action [12, 15, 16].

Verbs are essential for expressing tense, aspect, and mood, which are important components of meaning in a sentence. Tense indicates when the action or state occurred (past, present, or future), while aspect indicates how the action or state was carried out (simple, progressive, perfect, etc.). Mood indicates the speaker's attitude towards the described action or state (indicative, imperative, subjunctive, etc.).

In addition, verbs often take objects or complements, providing more information about the described action or state. For example, in the sentence "She gave the book to him," the verb "gave" conveys the action of giving, while the object "book" and the prepositional phrase "to him" provide additional information about the action.

Some languages have the aspect of subject-verb agreement grammatically. Verbs also play a role in establishing subject-verb agreement, which is important for correctly conveying a sentence's meaning. In English, verbs change their form depending on the subject of the sentence. For example, the verb "is" is used with a singular subject, while the verb "are" is used with a plural subject.

Overall, verbs are critical for unfolding the meaning of a sentence because they express the action or state being described, convey tense, aspect, and mood, take objects and complements, and establish subject-verb agreement.

In order to computationally specify the 'meaning' of a sentence, it's important to consider ontological and linguistic aspects of verbs and other parts of speech participating in a sentence.

The next section talks about Ontology, which further helps in unraveling the meaning of a word/sentence in a language.

1.2 Ontology

Ontology can help in deriving the meaning of a sentence by providing a framework for organizing and classifying the concepts and entities referred to in the sentence. Ontology is a branch of philosophy that deals with the study of existence, including the nature of entities, their relationships, and the categories to which they belong.

In natural language processing and computational linguistics, ontology is often used to create formal representations of concepts and entities in a particular domain. These representations can be used to help derive the meaning of sentences that refer to those concepts and entities.

For example, suppose we have an ontology that represents the concepts and relationships in the domain of biology. This ontology might include concepts such as "cell," "organism," and "evolution," and relationships such as "is a part of," "has," and "evolves into." When we encounter a sentence such as "The mitochondria are the powerhouse of the cell," we can use ontology to help derive the meaning of the sentence.

We can identify "mitochondria" as a concept in the ontology and determine that it is part of a cell. We can also identify "powerhouse" as a metaphorical expression that describes the function of the mitochondria in producing energy for the cell. By using ontology to represent and organize the concepts and relationships in the domain, we can derive a more precise and accurate meaning of the sentence.

Ontology can help derive the meaning of a sentence by providing a formal representation of the concepts and entities in a particular domain and a framework for organizing and classifying those concepts and entities.

We further talk about the Formal Ontology of Language, which helps us understand the relationship between language and meaning.

1.3 Formal Ontology of Language

There have been several theories, and others have contributed to the development of formal ontologies of language using several aspects of language and have helped to deepen our understanding of the relationship between language and meaning.

Following are a few examples of some of the theories on Ontology of Language:

1. Montague grammar [31]: Developed by Richard Montague in the 1960s, Montague grammar is a theory of natural language semantics that uses formal logic to represent the meanings of sentences. Montague grammar aims to provide a systematic way of mapping natural language sentences to their logical equivalents, using a formal language that includes syntax for expressions, a set of logical operators, and semantics that assigns truth values to expressions.

2. Situation semantics [2]: Developed by Jon Barwise and John Perry in the 1980s, situation semantics is a theory of natural language semantics that emphasizes the importance of context and situational factors in determining the meaning of sentences. Situation semantics aims to provide a formal model of meaning that takes into account the various factors that contribute to the interpretation of sentences, such as the speaker's intentions, the listener's expectations, and the physical and social context in which the sentence is uttered.

3. Generative semantics [20]: Developed by George Lakoff and others in the 1970s, generative semantics is a theory of natural language semantics that emphasizes the role of syntax in determining the meaning of sentences. Generative semantics posits that the meaning of a sentence is derived from its underlying syntactic structure, and that semantic features such as truth conditions and presuppositions are built into the syntax of the sentence.

4. Cognitive semantics [21]: Developed by Ronald Langacker and others in the 1980s, cognitive semantics is a theory of natural language semantics that emphasizes the role of conceptual structures and mental processes in determining the meaning of sentences. Cognitive semantics posits that the meaning of a sentence is derived from the conceptual structures that underlie it, and that these structures are formed through a process of categorization and conceptualization.

5. Verb semantic-based Ontology: Rajan [34] proposed an alternative verb semantic-based ontological classification. Otra [30] further extended the concept to adverbs and developed an ontological resource for English language, in order to add the intensional forms present in words to the existing information that has been used to capture the meaning of words in language in many ways.

   The Formal Ontology of Language as proposed by them [Rajan, Otra] [30, 34] suggests that the meaning of a word is continuous. This continuous medium of meaning can be thought as a dense medium. This theory accepts that there are hard and rigid points in situ in the meaning of a word whose existence and contribution to the meaning of a word can be understood. As these points are identified by the eruptive senses that shout louder among the infinite points when a meaning of a word is heard. The elusive objective is to find as many points in this continuous realm until the saturated meaning is reached. These points are organized into sense-types and sense-classes. The sense-types overlap among the words which have it and sense-classes do not overlap. It further proposes that verbs and adjectives have sense-types whereas nouns and adverbs have sense-classes. The sense-types and sense-classes along with their relations and other relations like morphological, and etymological constitute the entire framework of language. The basic motivation is to computationally manipulate language at the level of lexical meanings. Lexical mean-

ings, in the transaction of language, have discrete intrinsic forms of types and classes. These types and classes are unambiguously locatable in parts of speech through collective introspective inquiry first and then enriched with the help of computational methods of corpus study.

1.4 Objective and Contribution

The goal of this thesis is to use the Formal Ontology of language, that has been proposed by Rajan [34] and Otra [30] and develop a resource for Hindi. Furthermore, we use the resource to understand the verb-adverb collocation in Hindi. We further explore the relations of spatial and temporal markers in relation to the ontological categories provided by the Formal Ontology of Language. It helps in further corroborating the categorization of verbs and adverbs. Using the resource we also establish the maximum likelihood estimation of certain verb categories with the Hindi word 'sE' .

1.5 Thesis Summary and Organization

The thesis is divided into 8 chapters.
Chapter 1 gives an introduction about the problems that are being addressed and showcases the background behind the objective of the thesis.
Chapter 2 of the thesis explores the theory of the ontological classification of verbs and adverbs, as proposed by Rajan and Otra. The thesis utilizes this theory to expand the sub-categorization of adverb sub-classes, thereby extending its application.
Chapter 3 shows the manual annotation procedure for the Hindi language and provides validation.
Chapter 4 explains automatic sense-type identification of verbs and sense-class identification of adverbs using machine learning techniques.
Chapter 5 presents the associativity of different verb types with space and time markers in Hindi. The study constitutes a preliminary result obtained from a corpus-based analysis.
Chapter 6 investigates verb-adverb collocations in Hindi through a corpus-based approach, employing statistical measures like log-likelihood to assess association strength.
Chapter 7 presents the summary and conclusion of this research work.

Chapter 2

Extending the Formal Ontology of Language

In this chapter, we extend the Formal Ontology of Language proposed by Otra [30], by adding ontological sub-categories. The motivation for the Formal Ontology of Language stems from the overlapping verb-senses [34]. These overlapping verb-senses, in turn, are derived from the several theories of common intensional sense in verb, in various linguistic traditions [18, 29, 36]. This common intensional sense is also termed as universal verb.

We take a detailed look at the theory of universal verb in linguistic traditions, and the idea of punctum as proposed by Leibniz [23]. These two theories form the basis of the Formal Ontology of Language.

We also contribute to the Formal Ontology of Language by adding ontological attributes that are sub-classification of the adverbial categories.

2.1 Universal Verb in Linguistic Traditions

In any language, one can describe a particular state of affair using multiple verbs. For example, when describing a situation where the phenomena of 'cut' is involved, one can use any of the terms from 'divide', 'tear', 'dissect' etc. Furthermore, a verb describing a particular situation may carry overlapping meanings. For example, the verb 'walk' has 'move' as well as 'do' in its meaning form. This aspect of verbal ambiguity in language is widely present. The delineation of a minimalistic intensional sense in the verb necessitates looking at the theory of universal verb [18, 29, 36].

There have been several discussions on identifying universal verb in various linguistic traditions such as Greek and Indic. While in the Greek tradition, 'be' is considered as the universal verb [18], Indic tradition considers 'happening (bhavati)' as the universal verb, as per Nirukta verse 1.1 [29, 36]. The next section discusses 'happening (bhavati)' as a Universal Verb in Indian Linguistic Tradition.

2.2 'Happening (bhavati)' as Universal Verb in Indian Linguistic Tradition

In order to find verb in any sentence, we consider the question "What are you doing?". This can be answered with any possible action, indicating 'do' as the primitive sense of the verb. Hence, one can say that 'do' can be a primitive sense that will be present in every verb. However, Patanjali mentions three verbs that cannot be an answer to the above question. These are

1. being/existence (asti)
2. presence (vidyate)
3. happening (bhaāva)

Considering these verbs, it's apparent that the sense of 'doing' is not collocated in them. Therefore, 'be' can't be a universal verb. When the phenomena of birthing is to be signified, collocation of absence is obstructed, and when death is to be signified collocation of presence is obstructed. On the other hand, the sense of happening is collocated in every verb sense. Hence, one can say that happening is a universal verb whose minimal sense is collocated in every intensional sense of verb.

According to Bhartrihari, a verb has a sense of sequence and state. Hence, it has a sense of happening. Linguistic traditions in India have long regarded verbs as the centre of language (in both syntactic and semantic terms) right from Yāska , Pānini, Patanjali and Bhartrihari [27, 33, 38]. Meaning of verbal element is seen as *bhāva* (happening) as opposed to *sattā* (being) which stands behind nominal elements [7, 29]. Later, independent of linguistic discourse, logicians brought out hundreds of *bhāva-s* (happening) with atomic transformational structure <$entity_1$ | $entity_2$>like cause|effect, part|whole, predecessor|successor, qualifier|qualified, ascribed|ascriber, locus|locate, etc. These are atomic discriminants that form elementary meanings.

In the next section, we take a look at how Leibniz's idea of 'Punctum' has been constituted in deriving the ontological categories of different parts-of-speech.

2.3 Leibniz idea of 'Punctum' as Meaning Form

In the formal ontology of language, meanings of verbs are not seen as an entity like semantic primitive [41] but as a unified discriminative structure with a form <$entity_1$|*contiguous with*|$entity_2$>, in context of <continuum>. For example, the verb 'move' has a sense <predecessor state | *contiguous with*|successor state, in the context of 'move' *continuum*>. Its meaning is a continuant feel of motion punctuated by the discriminating logical structure of predecessor

Primitives (Elementary Bhvas/puncts)	Explanation
Know:Sense of knowing (Sanskrit, **jnāna—jnāpya** bhāva) (object of 'know' / the process involved in knowing that object)	Know/Knower Conceptualize, construct or transfer information between or within an animal. E.g. "forget" - to forget something one has to know about it. Forget is a process having state change from knowing to not knowing over a period of time. So,this particular verb has "knowing" as primary sense and change of state/"move" as secondary sense.
Move:Sense of Move /change /process (Sanskrit, **pūrva—apara** bhāva) (state at the beginning of a process / state at the end of the process)	Before/After Every process has a movement in it. The movement maybe a change of state or a change in location. E.g. "fall" - change of position from a higher state to lower state physically or in abstract sense. Actions like falling of leaves do not have a sense of agency, the fall happens on its own. So the word has the sense of 'pure movement'.
Do : Sense of agency (Sanskrit, **sādhya—sādhaka** bhāva) (something to be accomplished /accomplished)	Agent/Action A process which can be accomplished only with a doer. E.g. "cook" - has a sense of someone doing cooking. The process of cooking involves change of state from raw to cooked by a doer. So, it has 'doing' sense as primary and 'move' sense secondary to it.
Have : Sense of possession or having (Sanskrit, **grāhya—grāhaka** bhāva) (something that is the object of grasping /to grasp)	Grip/Grasp Possessing, obtaining or transferring a quality or thing. E.g. "like" - To like something one must have prior 'knowledge' about it. Liking is something you "have or possess". Hence 'have' is primary sense and "knowing" is secondary sense.

Figure 2.1 Seven Senses of Verb-Part-1

and successor states. One can read in its meaning such discrimination points. Leibniz called such punctuations as actual points [23], as endeavors, as ontologically vacuous, and as different from Euclidean points. Brentano [10] also built an idea of mental continuants as punctuated with *modo recto* and *modo obliquo*. These boundaries, punctuations or points are ontological as they vacuously discriminate ontic entities or states which are perceived as continuous.

Otra [30] developed a formal ontology of lexical meaning using such punctuational boundaries and 'bhavati' (happening) sense of verbs Figure 2.1 and 2.2. The next section talks in detail about the Formal Ontology of Language.

2.4 Deriving the meaning of verb: Formal Ontology of Language

.

Let us consider the verb 'move'. Its meaning is always more than the discriminative senses we read into it. The formal ontology of language establishes seven discriminant punctuation. When we say, 'rapidly move' or 'hesitantly move', we are doing adverbial modification of the meaning of 'move' and adding new modifier|modified points in its meaning. When appending

Primitives (Elementary Bhvas/puncts)	Explanation
Be : Sense of state of being (Sanskrit, **adhāra—ādheya** bhāva) (location/attribute)	Locus/Locatee Continuously having or possessing a quality. E.g. - "confuse" (I am confused). It is a state and it is located in me. 'Be / is' is the primary state and to get confused you must know and have contradictory opinion about the object. So, sense of 'know' is secondary.
Cut : Sense of part and whole (Sanskrit, **aṁśa—aṁśi** bhāva) (part of an object or process/whole to which the part belongs)	Part/Whole Separation of a part from whole or joining of parts into a whole. Processes which causes pain. Processes which disrupt the normal state. E.g. - "break" It has a sense of a thing being divided into parts. 'Cut' sense is primary to it and breaking is 'done' by someone so has a sense of agency 'do' as the secondary sense.
Cover : Sense of ascribe and ascription (Sanskrit, **āropya—āropaka** bhāva) (to be attributed/the one to which it is attributed)	Wrap/Wrapped Processes which pertain to a specific object or category. It is like assigning a boundary. E.g. - "guarantee" - when you guarantee you are putting a kind of cover (ascription) on that object so it has 'cover' as primary sense and someone has to do it, and so has 'doing' as secondary sense.

Figure 2.2 Seven Senses of Verb-Part-1

'rapidly' we add temporal-feature-class in the adverb whereas while appending 'hesitantly' we add force-feature-class in the adverb to the meaning of 'move'. Even when we have discriminated temporal or force features, meanings of 'rapidly' and 'hesitantly' are more than their adverb sense-classes and has delineated four adverb classes of discriminant point. Verbs are also seen as contiguants of nouns in seven or eight case relations. These seven/eight classes of verb-noun pairing are further coincident boundaries in the meaning of articulation with the verb. Further, noun-noun pairs and noun-adjective pairs are more coincident points. Otra [30] also proposes twelve noun-verb types of sense-points in the ontology. The verb-centric formal ontology of meaning is based on the categories that denote the logical forms of the meaning which are termed sense-type and sense-class. Otra [30] categorises different parts-of- speech as sense-type and sense-class. The next sub-section elaborates further on 'type' and 'class'.

2.4.1 TYPE and CLASS

In philosophy, the difference between type and class has been discussed as an aspect of logic [40]. Class is a collection defined by its members, whereas a set of things that meet some requirement to be a member is a type. Hence, we observe a sense of 'extension' in the definition of class and a sense of 'intension' in the definition of type.

Otra [30] proposes two kinds of the logical form of types and classes:

1. Forms of transference - When there is a change of state involved, in a particular context, there is a further alteration in the state of happening. Considering the verb 'moving',

there is a change of state from departure to arrival or vice versa. Those categories where a point structure of transference is obtained have been termed as 'sense-type'.

2. Forms of abstraction - The universal verb 'happening' has several ontological attributes as its collocation. The meaning of the form 'happening' is disambiguated in the context of its occurrence, identified by space, location, manner, and existential nouns. These extensional attributes have been termed as 'sense-class'.

Taking Patanjali's notion of non-doing verbs, the formal ontology of language formalizes three realms, in which different parts-of-speech of language are situated. These realms form a tripartite ontology, in a sentence structure. The next sub-section talks about the tripartite ontology in detail.

2.4.2 TRIPARTITE ONTOLOGY

As discussed above (section 2.2), the three non-doing verbs are *exist*, *present*, and *happen*. Based on the above non-doing verbs, Otra [30] proposed three realms of reality.

1. Existential - Those loci that can stand on their own, e.g Nouns

2. Presential - Those loci which are collocated with another loci, e.g. Adjectives, and Adverbs

3. Happening - Entities that have a sense of transference in a set of context, e.g. Verbs

Otra [30] proposes that the different parts-of-speech of language, belong to the three realms of reality, as described above.

In a sentence, a noun (which is in the realm of existential) and a verb (which is in the realm of happening) are related through case markers, adverbs(which is in the realm of presential) are a qualifier for verbs, and adjective is a qualifier for nouns. Figure 2.3 shows the tripartite structure of the formal ontology of the language

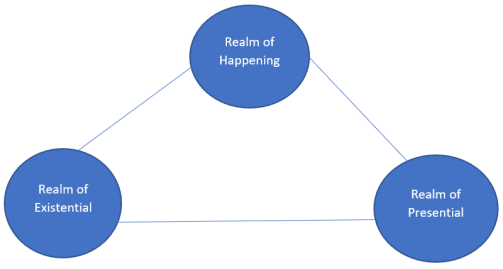

Figure 2.3 Tripartite Ontology

As is evident from the tripartite ontology, the meaning of a verb is also layered in a context, and the context has certain ontological attributes (nouns, space, location, manner, reason). Consequently, the ontological attributes of a verb can also be termed as its meaning components, which can be used to discern and distinguish its meaning from other verbs.

An interesting observation here is that even when two verbs are highly similar in their meaning, there is always an aspect that differentiates them from each other, and this aspect lies in their meaning component set. This implies a difference in their ontological attribute.

In the next subsection we take a look at the seven sense-type of verbs, and four sense-class of adverbs, as proposed by Otra [30]. In order to define the meaning of a verb exhaustively, and disambiguate its meaning from other similar verbs, we introduce sub-categories of adverb sense-types.

2.4.3 Sense-Type of Verb and Sense-Class of Adverbs

2.4.3.1 Verb

The seven sense-types of verbs have been derived by collecting the fundamental verbs used to define other verbs. These verbs are then grouped using intrinsic senses, which get designated to a particular sense-type. The seven sense-types of verbs are listed below with their primitive sense along with two Hindi and Telugu examples each.

1. Means|End - Do; khelanā (play), karanā (do); āduta (play), ceyuta (do)
2. Before|After - Move; bahanā (flow), calanā (walk); pāruta (flow), naduvuta (walk)
3. Know|Known - Know; jānanā (know), parakhanā (examine); ūhimcuta (imagine), parisīlimcuta (examine)
4. Locus|Located - Is; rahanā (stay), honā (happen); umduta (to be, stay), jaruguta (happen)
5. Part|Whole - Cut; kātanā (cut), mitānā (erase); koyuta (cut), vidipovuta (separate)
6. Wrap|Wrapped - Cover; jhāmpanā (cover), pahanānā (dress-up someone); mūyuta (cover) , ākramimcuta (contain forcefully)
7. Grip|Grasp - Have; pānā (get), lenā (take); bhayapaduta (fear), tīsukonu (take)

2.4.3.2 Adverb

Meaning of verbs can further be understood by adverbs, as they modify verbs. The sense-classes of adverbs are inspired by adverb classification in Sanskrit. Following are the identified sense-classes along with their fundamental sense, illustrated with English, Hindi and Telugu examples

1. Temporal - Adverbs that attribute to sense of time. e.g Never; sasamaya (timely); varusagā (continuously)
2. Spatial - Adverbs that attribute to physical space. e.g There; pās (near); davvu (far away)
3. Force - Adverbs that attribute to cause of happening e.g. Dearly; barbas (unwillingly); gattiga (tightly)
4. Measure - Adverbs dealing with comparison, judgement. e.g - Only,; lagbhag (approximately); gaddu (abundantly)

In the following examples, we look at different adverbs modifying verbs carrying similar meanings. It is observed that the verbs carrying the same semantic meaning takes different adverbs as its qualifier.

Let us consider the following examples: .

1. *usane rAma se dogunA hisAba lagAyA*
   उसने राम से दोगुना हिसाब लगाया।
   She solved twice the sum as that of Ram

2. *usane jaldI se samAdhAna kiyA*
   उसने जल्दी से समाधान किया।
   She resolved it quickly.

In the above sentences, the verb 'hisAba lagAyA' means to solve a mathematical question.

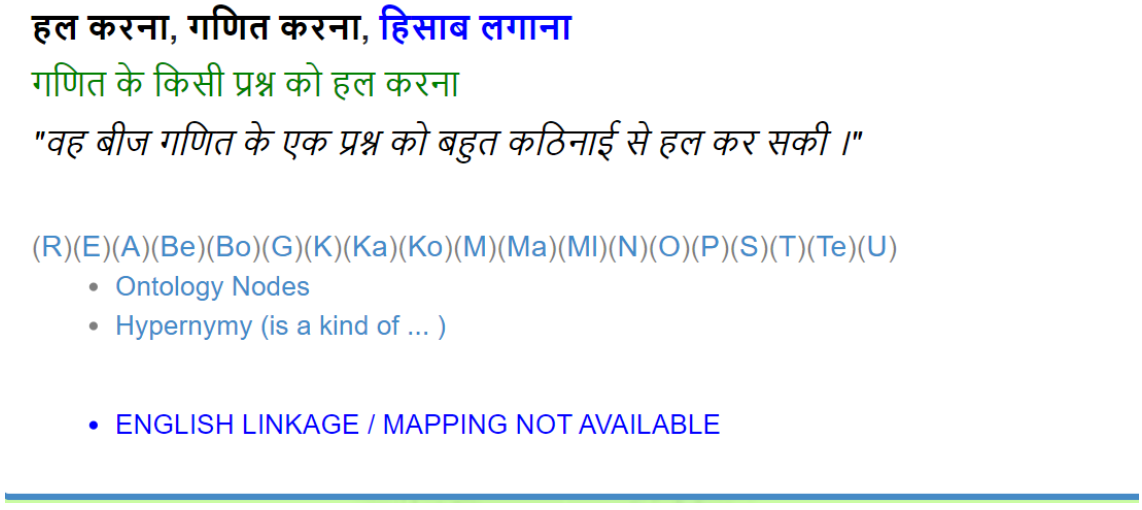

Figure 2.4 hisAba lagAyA Synset in Hindi WordNet

and
'samAdhAna kiyA' means to resolve a situation (Figure-2.5)

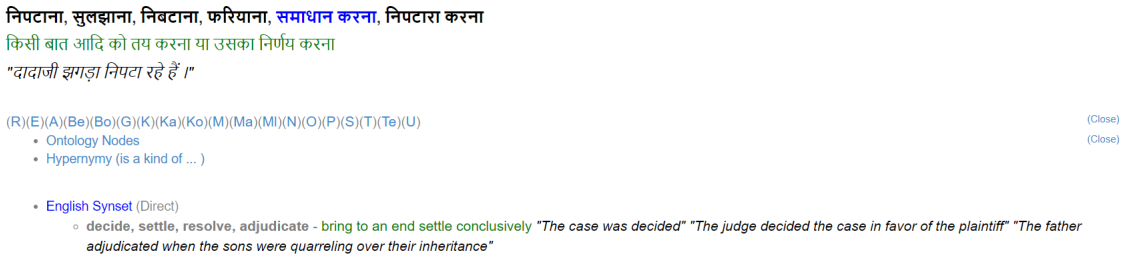

Figure 2.5 samAdhAna kiyA Synset in Hindi WordNet

It is observed that both the verbs have *solve* as their meaning at a deeper semantic level. Now, if we observe the adverbs ('dogunA' and 'jaldI) that are used here to modify these verbs (hisAba lagAyA and samAdhAna kiyA), respectively, they belong to the same sense-class i.e. 'Measure'

Let's now swap the adverbs used above to modify the verbs (hisAba lagAyA and samAdhAna kiyA) .

1. *usane rAma se jaldI hisAba lagAyA*
   उसने राम से जल्दी हिसाब लगाया।
   She solved the sum quicker than Ram

2. *usane rAma se dogunA samAdhAna kiyA*
   उसने दोगुना समाधान किया।
   She resolved the situation twice as Ram .

   It is clear that the above sentence does not carry any meaningful semantics

It is noted that if the adverb 'dogunA' is used to modify the verb 'samAdhAna kiyA', the sentence will not carry any meaningful semantics. Hence, there must be a difference in the ontological attributes of these two verbs that differentiate their meaning. We propose that there must be a difference at the attribute level of the sense-class of adverb. This creates a need to sub-classify the adverb sense-class.

Our Contribution: We propose further sub-classification of 'measure' and 'temporal' classes of adverbs. The next sub-section further elucidates upon the need to sub-classify the sense-classes of adverbs.

2.4.3.3 Adding sub-class for Measure and Temporal Adverbs

Measure-Adverb subclassification

Measure class of adverbs has been further subdivided into 5 sub-classes. The sub-classification is listed below:-

- Judgement - Adverbs that have attributes of assessment, evaluation. e.g (sāf-sā)Clearly
- Comparative - Adverbs that have attributes of polarity. e.g. (adhiktar)Mostly
- Conditional - Adverbs that have relational/conditional attribute. e.g agar (If)
- Absolute - Adverbs that have n dependencies. e.g.darshanārth (For visiting).
- Size- Adverbs that are related to quantities. e.g dugunā(doubly)

Temporal sub-classification We derived seven sub-classification of temporal adverbs.

1. Pace: An adverb denoting the speed of an action e.g. atishIghra (Fastly), vilaMbapUrvaka (Delay)
2. Starting time: An adverb denoting the start of an event, e.g.: sarvaprathama (firstly), pahale-pahala (starting)
3. Interruption: An adverb denoting impeding or ceasing of an action or event e.g. achAnak (Suddenly), anAyAs (Suddenly)
4. Build up: An adverb stressing on the intensity of the event or action: lagAtAra(continuously), niraMtara(regularly)
5. Repition: Adverb denoting the repeated occurrent of an event or action, dobArA(again), phira (again)
6. Ending time: An adverb signifying the end of an action or event: e.g. antataH (At the end), AkhirakAra (At the end)
7. Duration: Adverb denoting the total time lapsed for an activity or an event, e.g. :dinabhara (all the time), dinarAta (all the time)

2.4.3.4 Deriving the adverbial attributes to represent the semantic meaning of similar verbs

In the examples given in the section 2.4.3.3, adverb 'jaldI' belongs to 'Comparative' sub-class and 'dogunA' belongs to 'Size' sub-class of 'Measure' sense-class. Hence, we are able to represent the ontological attributes that differentiates the meaning of similar verb.

This distinction necessitates populating the database of a language with the sense- type of verbs and sense-class of adverbs. Therefore, in the next chapter, we focus on manually annotating Hindi verbs and adverbs with these attributes.

Chapter 3

Sense-Type and Sense-Class Identification for Hindi verbs, and adverbs

3.1 Background

In the previous chapter, we discuss the ontological categories of verb-types and adverb-classes, proposed by Rajan [34] and Otra [30], respectively. Using the categories, ontological resources for English and Telugu have also been developed by them. The ontological resources contain annotation for verbs and adverbs for these languages. It is observed that no such resource has been developed for Hindi language. Hence, an effort has been made to develop the ontological resource for Hindi verbs and adverbs, and it is aimed at covering majority of the Indian Languages.

3.2 Procuring the Dataset

To populate the ontological resource for Hindi, we collect the data for different word classes related to verbs and adverbs. Hindi verbs and adverbs from different resources like dictionary and WordNet were collected. For our purpose, we use the Hindi Shabdsagar dictionary (Benares: Nagaripracarini Sabha, 1965-75. 11 vols. 2nd edn) [1] and Hindi Wordnet [2] In the following sections, each of the resources has been discussed and the overall statistics of the populated data have been presented.

3.2.1 Shabdsagar Dictionary

Hindi Shabdsagar is considered to be the standard dictionary for Hindi. It was developed by Shyam Sunder Das and published by Kashi Nagari Pracharani Sabha. It contains around 100,000 Hindi roots. Verbs are categorized as sakarmak(transitive) and akarmak(intransitive)

[1] https://ia601603.us.archive.org/20/items/in.ernet.dli.2015.348711/2015.348711.Hindi-Shabdasagar.pdf
[2] http://www.cfilt.iitb.ac.in/wordnet/webhwn/

Parts-of-Speech	Total number
Transitive Verbs	2821
Intransitive Verbs	1245
Adverb	3000

Table 3.1 Distinct number of verbs, adverbs and adjectives in Hindi Shabdsagar dictionary

Parts-of-Speech	Total number
Verbs	6778
Adverb	2114

Table 3.2 Distinct number of verbs, adverbs and adjectives in Hindi WordNet

kriyA. It does not contain phrasal verbs. Besides containing Hindi roots, it also contains words from different dialects of Hindi. Table 3.1 summarises the distinct number of verbs and adverbs present in the dictionary.

3.2.2 Hindi WordNet

Hindi WordNet [28] is inspired by the work of English WordNet [26] and has been developed by Bhattacharya [8]. HWN's structure is comparable to that of Princeton WordNet. It's made up of synsets and semantic connections. A synset is a collection of synonyms for the same idea. Hypernymy, hyponymy, meronymy, holonymy, troponymy, and other fundamental semantic relations are associated with synsets. In contrast to Princeton WordNet, HWN includes additional relations such as gradation, causative, compounds, conjunction, and so on. Table 3.2 summarises the overall statistics for verbs, adverbs, and adjectives in the Hindi WordNet 1.5 version.

In the next section, we look at the sense-type categories of verbs, and sense-class categories of adverbs, as developed by Otra [30]. Furthermore, we look at the categories of the sub-classification of 'Measure' and 'Temporal' class of adverbs. After briefly defining each of the verb-types and the adverb-class, we demonstrate the manual process to identify the sense-class and sense-types of adverbs and verbs, respectively in the subsequent sections.

Our Contribution: We extend the previous work to create the resource for Hindi language, and also add annotations for the sub-classes of 'Measure' class of adverbs.

3.3 Ontology Enrichment

After collecting verbs and adverbs from the dictionary and the WordNet, we identified sense-types and sense-classes for verbs, adverbs. The sense-types for around 3678 Hindi verbs, and sense-classes for around 1516 Hindi adverbs have been manually annotated. Following sections talk about sense-identification for verbs and adverbs.

3.3.1 Verb

As discussed in the previous chapter, different verbs can be used for describing the same situation, and are collocative in nature. In a single verb many verbal sense points can be present and different verbs may share same verbal sense points. For example "walking", "running" entails a sense of 'move'. Verb like "studying" entails a sense of 'know' and 'do'. "Eating" entails a sense of 'do' and 'have'. As detailed in the previous chapter, verbs are organized as sense-types. Otra [30] has shown the existence of seven primitive sense-types of verbs. These seven sense-types of verbs have been derived by collecting the fundamental verbs used to define other verbs. These verbs were then grouped using intrinsic senses, which were designated to a particular sense-type. The seven sense-types of verbs are listed below with their primitive sense along with two Hindi and Telugu examples each.

1. Means|End - Do; khelanā (play), karanā (do); āduta (play), ceyuta (do)
2. Before|After - Move; bahanā (flow), calanā (walk); pāruta (flow), naduvuta (walk)
3. Know|Known - Know; jānanā (know), parakhanā (examine); ūhimcuta (imagine), parisīlimcuta (examine)
4. Locus|Located - Is; rahanā (stay), honā (happen); umduta (to be, stay), jaruguta (happen)
5. Part|Whole - Cut; kātanā (cut), mitānā (erase); koyuta (cut), vidipovuta (separate)
6. Wrap|Wrapped - Cover; jhāmpanā (cover), pahanānā (dress-up someone); mūyuta (cover) , ākramimcuta (contain forcefully)
7. Grip|Grasp - Have; pānā (get), lenā (take); bhayapaduta (fear), tīsukonu (take)

Each of the verbs can have all the seven dimensions of sense-types. The degree depends on the usage/popularity of a sense in a language. In our resource, we have identified two sense-types of each verb, i.e. primary and secondary. Consider the verb 'dance' in the sentence "Madhuri is dancing gracefully". Here 'dance' involves a sense of movement which a doer does. Thus Before|After is a primary sense and Means|End is a secondary sense. For polysemous verbs, sense-type identification was done for each of their different meanings. For example, the verb "rap" has three meanings. Thus rap1, rap2, rap3 have been added in the resource along with its meaning sense-types.

1. rap1- Criticizing someone, Means|End and Know|Known.

2. rap2- To perform rap music, Means|End and Before|After.

3. rap3- To hit or say something suddenly and forcefully, Means|End and Part|Whole.

Considering the attempt made by Kachru [17] to classify the verbs based on the inherent properties of verbs having syntactic consequences, we observe that the sense-type of verbs in our resource exhibit some similarities. Following are a few examples.

1. **Stative**- Stative verbs indicate state of the subject. They are composed of an adjective or past participle and the verb 'be'. khulaa honaa 'to be open' is an example of stative verb. In OntoSenseNet, we mark the primary sense-types of such verbs as 'Locus|Locate'

2. **Inchoative** verbs indicate change of state. They are either a simple verb or a complex verb. The complex verbs are composed of a nominal and a verb having the meaning of 'become' or 'come'. khulanaa 'to become open' and yaad aanaa 'to remember' are examples of inchoative verbs. In our resource, we mark the primary sense of 'khulanaa' as Before|After, and 'yaad aanaa' as Grip|Grasp

3. **Active** verbs indicate actions. They are either causal verbs which are morphologically derived from the intransitive verbs or conjunct verbs composed of a nominal and the verb 'do'. kholanaa 'to open' and yaad karanaa 'to recall' are examples of active verbs.

   The primary sense-types of such verbs have been tagged as 'Agent|Action'

4. **Volitional** verbs denote deliberate actions. Such as 'chalwAnA', 'dhulwAnA'. We mark the primary sense-types of such verbs as 'Agent|Action'

Sense-types for 3,152 Hindi were manually identified. Previous work on Telugu [32] and English [30]

Figure 3.1 shows the sense-type distribution for English, Hindi and Telugu verbs in OntoSenseNet.

3.3.2 Adverb

Meaning of verbs can further be understood by adverbs, as they modify verbs. The sense-classes of adverbs are inspired from adverb classification in Sanskrit. Following are the identified sense-classes along with their fundamental sense, illustrated with English, Hindi and Telugu examples

1. Temporal - Adverbs that attributes to sense of time. e.g Never; sasamaya (timely); varusagā (continuously)

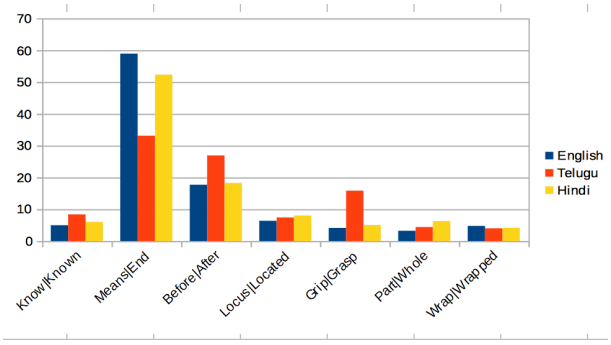

Figure 3.1 Verb Sense-type distribution

Table 3.3 Adverb Sense-Class Distribution

Sense-Class	English	Hindi	Telugu
Temporal	5.5	24.3	28.7
Spatial	2.7	13.5	12.8
Measure	39.4	32.2	31.6
Force	52.2	30	26.7

2. Spatial - Adverbs that attributes to physical space. e.g There; pās (near); davvu (far away)
3. Force - Adverbs that attributes to cause of happening e.g. Dearly; barbas (unwillingly); gattiga (tightly)
4. Measure - Adverbs dealing with comparison, judgement. e.g - Only,; lagbhag (approximately); gaddu (abundantly)

Sense-classes for 2,214 Hindi and 101 Telugu adverbs have been manually identified. Table 3.3 shows sense-class distribution of adverbs for English, Hindi and Telugu.

3.3.3 Measure-Adverb sub-classification

Measure class of adverbs have been further subdivided into 5 sub-classes. 690 Measure adverbs were identfied for Hindi. The subclassification is listed below:-

- Judgement - Around 142 adverbs had attributes of assessment, evaluation. e.g (sāf-sā)Clearly
- Comparative - Around 354 adverbs had attributes of polarity. e.g. (adhiktar)Mostly
- Conditional - Around 78 adverbs had relational attribute. e.g agar (If)

- Absolute - Around 72 adverbs had attributes that had no dependencies. e.g.darshanārth (For visiting).
- Size- Around 44 adverbs had attributes related to quantities. e.g dugunā(doubly)

3.4 Resource Validation

To show the reliability of the resource, Cohen's Kappa [19] was used to measure inter coder-agreement. The annotation was done by one human expert and it was cross-checked by another annotator who was equally trained. Verbs and adverbs were randomly selected from our resource for the evaluation sample. The inter coder agreement for 500 Hindi verbs and 1,000 adverbs were 0.70 and 0.91 respectively. Validation of the resources shows high agreement [20].

The manual annotation is a time-consuming and laborious process. Hence, we automate the sense-identification of verbs and adverbs.

In the next chapter, we look at the methods to automate the sense identification of Hindi verbs and adverbs.

Chapter 4

Automatic sense-type and sense-class identification

4.1 Word Embeddings

Word Embedding is a method that maps words and phrases to vectors of real numbers. It is a vector space model that has been widely used in information retrieval.

Word2Vec is one of the kinds of Word Embedding method. It is a neural network based approach that uses two algorithms; Continuous Bag of Words(CBoW) and Skip-Gram [24]. It takes large amount of text and creates high-dimensional representation of words present in the corpus. These vectors seems to capture linguistic regularities between words.

4.1.1 Continuous bag of Words

In the continuous bag of words model, for a given target word multiple words are represented as its context. Fig 4.1 shows the architecture of a CBoW method. It takes C inputs (contexts), where C denotes total window size of the context. The architecture consists of one hidden network and one output layer. It finally returns probability of a target word given some contexts. These probabilities signify relationship of the target word with different context words.

4.1.2 Skip-Gram

Skip-Gram takes target word as input and outputs probabilities for different contexts. Fig 4.2 demonstrates Skip-Gram architecture. It has an input vector(the target word) and the hidden layer consists of N neurons. The final layer consists of C different neurons, which is the context. The probabilities of the last layer gives the relationship of the input target word with different context words. In a nutshell, the model uses the current word to predict the surrounding window of context words. It gives more weightage to the nearby context words than more distant context words.

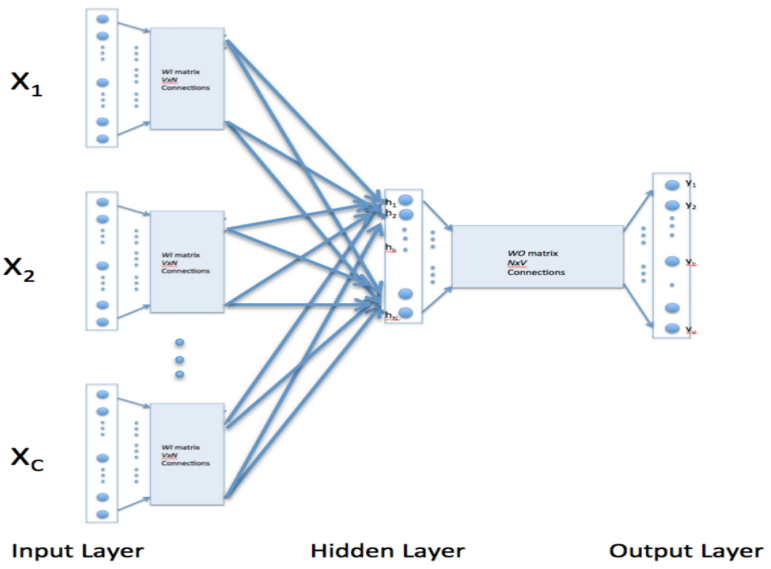

Figure 4.1 CBoW Architecture

CBOW is computationally faster than Skip-gram.[1] The latter performs better on infrequent words.

Word Embeddings have been widely used for extracting similar words [22]. Previous study has shown that word embedding has significant improvement over WordNet based measures [37]. We used this property to assign sense-type of verbs and sense-class of adverbs. This was done in order to facilitate the annotation task. However, this was further verified manually. The following section describes the sense identification method for English and Hindi.

4.2 Method

4.2.1 Corpus Collection

Hindi: Hindi corpus was collected from Leipzig [2], Hindi wiki-dump [3] and Lindat [1]. It contained 3,73,45,049 sentences and 75,31,64,082 words. To get verbs, adverbs and verb-adverb pair, this corpus was fully parsed using iscnlp tagger[4]

. **English**: English corpus was collected from wikipedia and health domain. It contained 5,82,27,633 sentences and 68,58,53,091 words. The corpus was fully parsed using stanford-

[1]https://code.google.com/archive/p/word2vec/
[2] http://corpora2.informatik.uni-leipzig.de/download.html
[3]http://kperisetla.blogspot.in/2013/01/wikipediahi-offline-wikipedia-in-hindi.html
[4]https://github.com/iscnlp/iscnlp

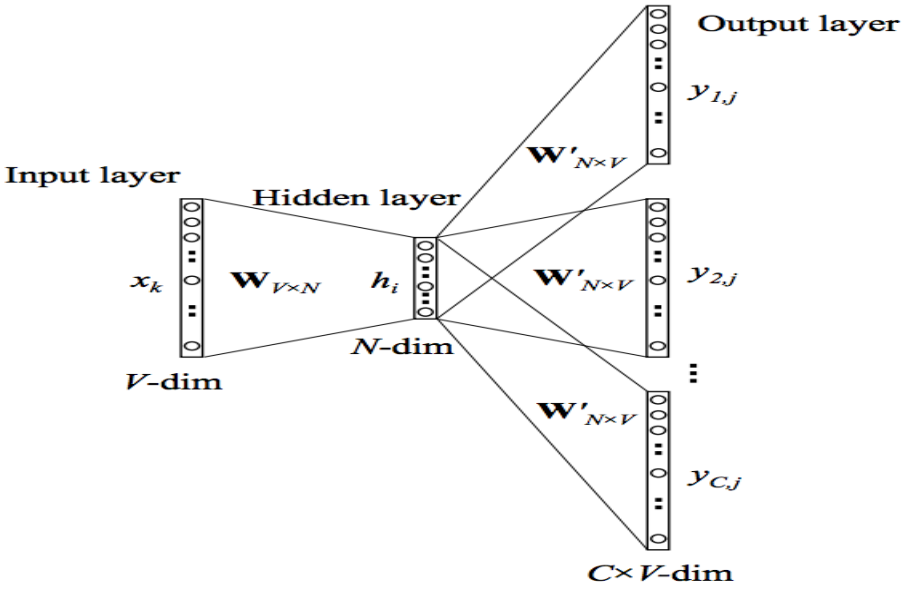

Figure 4.2 Skip-gram Architecture

dependency-parser [11] in order to extract verbs, adverbs and verb-adverb pairs.

4.2.2 Parameter Tuning

Word2vec [25] model was trained using CBOW and Skip-Gram technique on English and Hindi corpus. The vector dimensions used were 50, 100, 200, 300.

4.2.3 Similarity Clusters

For both English and Hindi, verbs and adverbs from the OntoSenseNet database were used to get the similarity clusters. The words were there in their root form. Using the different Word2Vec model, the similarity clusters were obtained. The cosine similarity was calculated for each verbs and adverbs and 0.8 was chosen to be the minimum similarity score. Recall that in the OntoSenseNet database every verb has been assigned two sense-types i.e. Primary and Secondary, whereas the adverbs have been assigned just one sense-class. We just considered Primary sense for the evaluation for verbs. From the similarity clusters, the most frequent sense-type/class was extracted; the actual sense-type/class of the word was obtained from OntoSenseNet. There can be polysemy in verbs and adverbs and our corpus does not contain annotated disambiguated senses of verbs and adverbs. For our purpose we considered all the meanings of a veb/adverb and their primary sense-type/class. All the combinations of the sense-type/class were obtained from the cluster. The accuracy was measured in terms of how

A	B	C	D
50	142	108	76.05%
100	142	118	83.09%
200	142	120	84.50%
300	142	119	83.80%

Table 4.1 Statistics for the sense-identification by Word2Vec

many of these clusters correctly predicted the sense-type/class. Accuracy increases if any of the maximum occurring sense from each of the combinations matches with the sense-type/class of the word whose sense-type/class is being predicted.

Following is an example to elucidate it further.

Word - *nipatanA*

It has broadly four meanings and they are as follows:-

1. kisi cheez ka khtm ho jana - Part|Whole

2. paraast hona - Part|Whole

3. tay hona - Before|After

4. koi kaam karne se nivritt hona - Before|After

The list of different primary sense is ('Part|Whole', 'Before|After') The words in the similarity clusters are:- ('jUJanA', 'nibatanA', 'ubaranA')

Different combinations of the primary senses of each of the words are:- (('Agent|Action', 'Part|Whole', 'Before|After'), ('Agent|Action', 'Before|After', 'Before|After')]

From the second group of the cluster the maximum occurring sense-type is *'Before/After'*. This matches with one of the sense-types of the word *nipatanA*

As the coverage was different for different word2vec models, we selected the words that were common in all the word2vec models.

4.2.3.1 Hindi-Verb

Table 4.1 and 4.2 summarizes the statistics for Hindi using CBoW and Skip-Gram learning model respectively.

A	B	C	D
50	437	270	61.78%
100	437	277	63.38%
200	437	271	62.01%
300	437	278	63.61%

Table 4.2 Statistics for the sense-identification by Word2Vec

	A	B	C	D
CBoW-200	82	64	78.04%
Skip-gram-100	82	63	76.82%

Table 4.3 Accuracy for CBoW and Skip-Gram

Dimensions 200 and 100 performed the best for CBoW and Skip-Gram respectively. Hence, the comparison was made using these two methods to get the best performing models out of these two. Verbs that were commonly covered by both the learning algorithms were chosen for the comparison.

4.3 shows that CBoW performs marginally better than the Skip-Gram model. Hence, for Hindi verbs CBoW with dimension 200 was used for automatic primary sense identification.

4.2.3.2 Hindi-Adverb

Table 4.4 and 4.5 that for Hindi-adverbs, the dimension 200 outperformed others for both the learning algorithm (CBoW and Skip-Gram)

Table 4.6 shows comparison between skip-gram and cbow learning algorithm for the dimension 200.

A	B	C	D
50	388	273	70.36%
100	388	293	75.51%
200	388	296	76.28%
300	388	293	75.51%

Table 4.4 Statistics for the sense-identification of Adverbs using Word2Vec(Skip-Gram)

A	B	C	D
50	236	220	93.22%
100	236	221	93.64%
200	236	224	94.91%
300	236	223	94.49%

Table 4.5 Statistics for the sense-identification of Adverbs using Word2Vec(CBoW)

	A	B	C	D
CBoW-200	201	189	94.02%
Skip-gram-200	201	188	93.53%

Table 4.6 Accuracy for CBoW and Skip-Gram

A	B	C	D
50	51	23	45.09%
100	51	34	66.66%
200	51	31	60.78%
300	51	29	56.86%

Table 4.7 Statistics for the sense-identification of Adverbs using Word2Vec(CBoW)

A	B	C	D
50	214	112	52.33%
100	214	114	53.27%
200	214	109	50.93%
300	214	124	57.94%

Table 4.8 Statistics for the sense-identification of Adverbs using Word2Vec(Skip-gram)

Table 4.9 Statistics for the sense-identification by Word2Vec

A	B	C	D	E	F
Verb	1,485	1,182	303	185	61.056%
Adverb	1,054	832	222	220	99.09%

Table 4.10 Similarity Cluster and the maximum occurring sense-type

Verb	Verb-Clusters	Maximum occurring sense-type
cīranā	nocanā, ghisanā, chedanā, khuracanā, pīsanā,phulānā	Part|Whole
jānanā	batānā, kahanā, lenā-denā, mālūma, mānanā, pūchanā	Know|Known

4.2.3.3 English-Verb

Column A is part-of-speech. Column B shows number of words in that part-of-speech that were randomly sampled from the corpus. Column C shows number of words for which sense were already present in the resource. Column D shows number of words for which sense identification was carried out. Column E contains number of words whose sense were correctly identified by Word2Vec. Column F shows accuracy in percentage.

Table 4.10 and Table 4.11 shows the verb and adverb clusters, respectively. In each of the tables the similarity with the words in column-1 is above 0.7. Column 3 shows the maximum occurring sense-type/sense-class

Table 4.11 Similarity Cluster and the maximum occurring sense-class

Adverb	Adverb-Clusters	Maximum occurring sense-class
tigunā	dogunā,dugunā,caugunā	Measure
yakāyaka	sahasā,ekāeka,acānaka	Temporal

Chapter 5

Spatial and Temporal Dynamics in Hindi Verbs: A Corpus-Based Analysis

5.1 Introduction

The concept of space and time play a crucial role in human thinking and cognition, and their significance in language and literature has been a topic of intrigue. Various studies have explored their conceptualization in different languages [15]. Natural language manifests the concept of space in diverse ways, such as through verbal suffixes, prosodic markers, and others. Multiple multifaceted approaches have been pursued to illuminate the intricate relationship between space-time constructs and human cognition.

This chapter explores the relationship between the verb-types and spatial and temporal markers in a corpus-based setting. The Formal Ontology of Language is employed as a theoretical framework to investigate the interconnections between verbs and their spatial and temporal attributes. The OntoSenseNet resource serves as a valuable tool for identifying the specific spatial and temporal characteristics of verbs, which include their association with kārak relations, adverbs, and Tense-Aspect-Modality features.

To further quantify and validate these associations, we extended our study by employing two well-established statistical methods: Log-Likelihood and Bayesian inference.

Previous studies in the field of linguistics have demonstrated the effectiveness of Log-Likelihood and Bayesian inference in deriving reliable associativity measures [19,39]. By leveraging these statistical tests, we aim to provide a rigorous and data-driven assessment of the alignment between the fundamental meanings of each verb-type and the spatial-temporal markers in the Hindi language. The statistical analyses will enable us to ascertain whether the observed patterns are statistically significant or merely coincidental and ultimately enhance our understanding of the linguistic representation of space and time in Hindi.

In the next section, we discuss various studies on the linguistic expression of time, space, and spatial concepts. Furthermore, it underscores the multidimensional nature of linguistic analysis concerning space, time, and semantics.

5.2 Background

5.2.1 Linguistic Expression of Space and Time

Studies have shown how time is conveyed and understood in both literal and non-literal language. Behavioral research has looked into how time is expressed in metaphor and grammatical terms. [9] investigate the relationships between the 'under' and 'over' time and space domains in the three languages. They investigate what types of lexicalized temporal concepts are permitted in the temporal constructions under investigation using corpus and Internet data. [13] demonstrated that the spatial concepts of distance and proximity are at the heart of the social deictics of addressee honorifics and their non-honorific counterparts.

5.2.2 Statistical Measure for Association Measure

In the realm of Natural Language Processing (NLP) and corpus linguistics, the analysis of linguistic associations is a fundamental endeavor, revealing crucial insights into language structures, relationships, and patterns. To quantitatively evaluate the strength and significance of these associations, various statistical measures have been developed. These measures play a pivotal role in discerning the interplay between linguistic elements, which is essential for tasks such as collocation analysis, topic modeling, and understanding semantic relationships.

5.2.2.1 Log-Likelihood

One prevalent statistical measure employed in this context is the Log Likelihood (LL) ratio. LL assesses the likelihood of co-occurrences between words or phrases in a corpus compared to what would be expected by chance. Its formula,

$$\text{Log Likelihood} = 2 \times \left(\sum_{i,j} O_{ij} \times \log \left(\frac{O_{ij}}{E_{ij}} \right) \right)$$

encapsulates the observed and expected frequencies of co-occurrences, aiding in the identification of significant word associations. Notable research by Dunning [14] introduced and elucidated the log likelihood ratio, serving as a cornerstone in statistical linguistic analysis.

5.2.2.2 Bayesian Inference

Bayesian inference emerges as a powerful tool to quantify and refine the understanding of association measures. In the context of linguistics, where we seek to discern connections between variables such as verb categories and space-time markers, Bayesian inference offers a principled approach to estimate and update the strength of their associations. This methodology accounts for our prior beliefs and adapts as new evidence in the form of observed data is introduced.

Bayesian inference operates within the framework of Bayes' theorem, which expresses the posterior probability of an event given observed data. For our linguistic investigation, let us denote the variables of interest as Verb Categories (V) and Space-Time Markers (M). We aim to ascertain the extent to which specific verb categories and space-time markers co-occur, thus revealing the strength of their association.

In mathematical terms, Bayes' theorem takes the form:

$$P(\text{Association Measure}|D) \propto P(D|\text{Association Measure}) \times P(\text{Association Measure})$$

Here:

- P(Association Measure D) denotes the posterior distribution of the association measure after considering the observed data (D)
- P(D Association Measure) represents the likelihood of the observed data given a particular association measure. This involves analyzing the frequency and patterns of co-occurrence between verb categories and space-time markers.
- P(Association Measure) encapsulates the prior distribution of the association measure, encompassing our initial beliefs or assumptions about their relationship.

In the subsequent sections, we employ the outlined statistical measures to extract the association measures for spatial-temporal markers and verb-types in Hindi.

5.3 Spatial and Temporal Markers

Broadly the space and time marker can be understood as Tense-Aspect-Modality(TAM) features of verbs, Spatio-Temporal Nouns, Spatio- Temporal kāraka or case relation. Spatio-Temporal Adverbs.

5.3.1 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.

5.3.2 NST Nouns

Indian languages contain content words that denote time and space. They can be present as a spatial and temporal argument of a verb along with an appropriate case marker. These nouns are marked as NST [3] 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.

5.3.3 Spatial and temporal kāraka or case markers

Some nouns 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.

5.3.4 Spatial, Temporal Adverbs

Based on the Formal Ontology of Language, adverbs whose sense class is spatial/temporal, can be used as another space-time marker for the analysis

We extract these space-time markers and their frequency counts for their association with different verb-types from Hindi corpus.

5.4 Data

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 . Table 5.1 shows different corpus collected from writings of different authors.

5.4.1 Dependency Relation and Morphological feature extraction

We used Hindi dependency parser[1] to extract these depedency relations. The dependency parser uses guidelines from AnnCorra [5] for annotations. It is treebanks for Indian Languages Guidelines for annotating Hindi Treebank. The annotation is stored in SSF format [4]. It

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

Table 5.1 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

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)
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)

5.5 Frequency Distribution

5.5.1 Tense-Aspect-Modality

Figure 5.1 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.

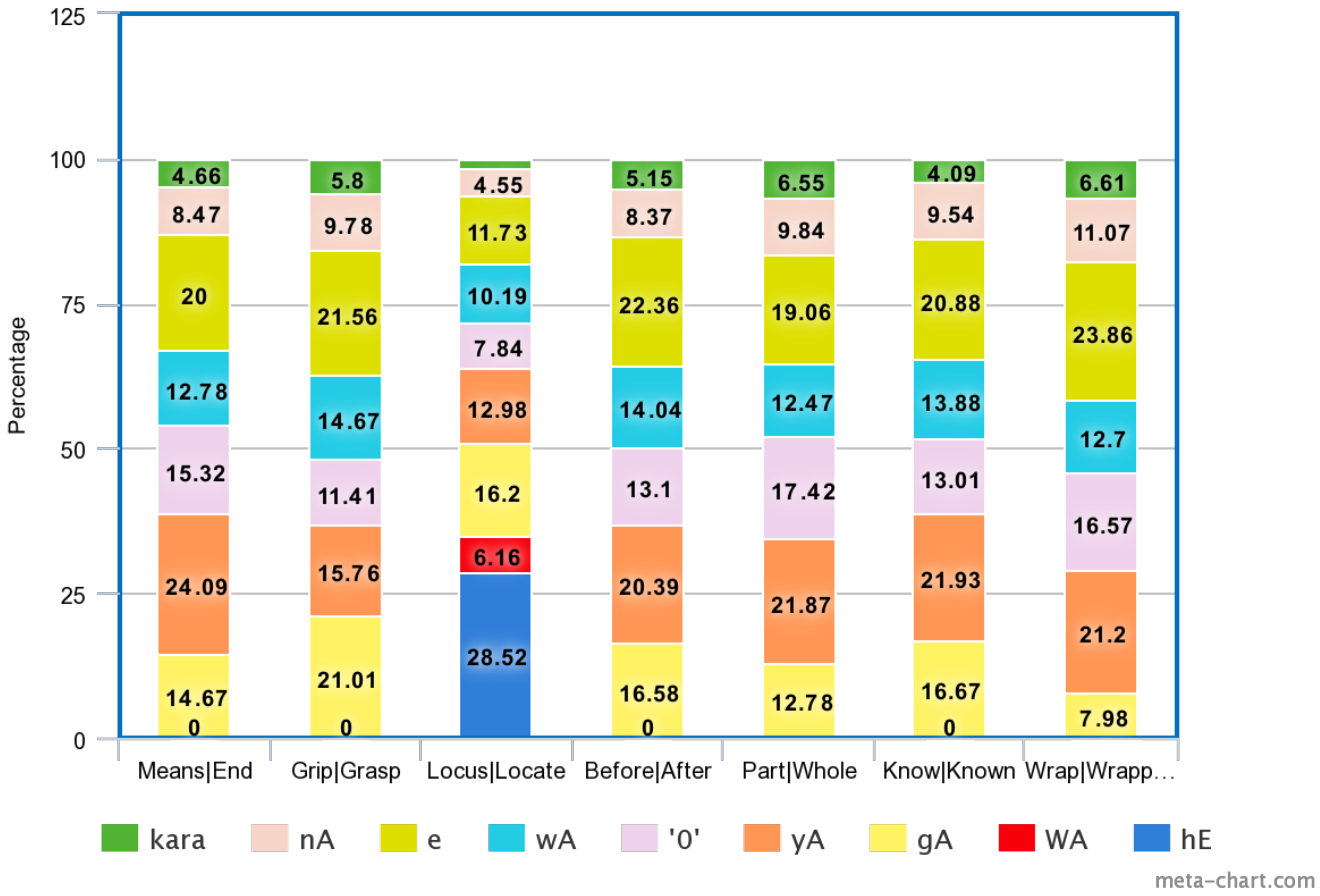

Figure 5.1 Frequency Distribution of TAM features for Verb sense-types

5.5.2 NST Nouns

The frequency distribution of sense-types of each verb with the kāraka relations with spatial/temporal nouns were tabulated. Figure 5.2 shows the frequency distribution for the same. From 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.

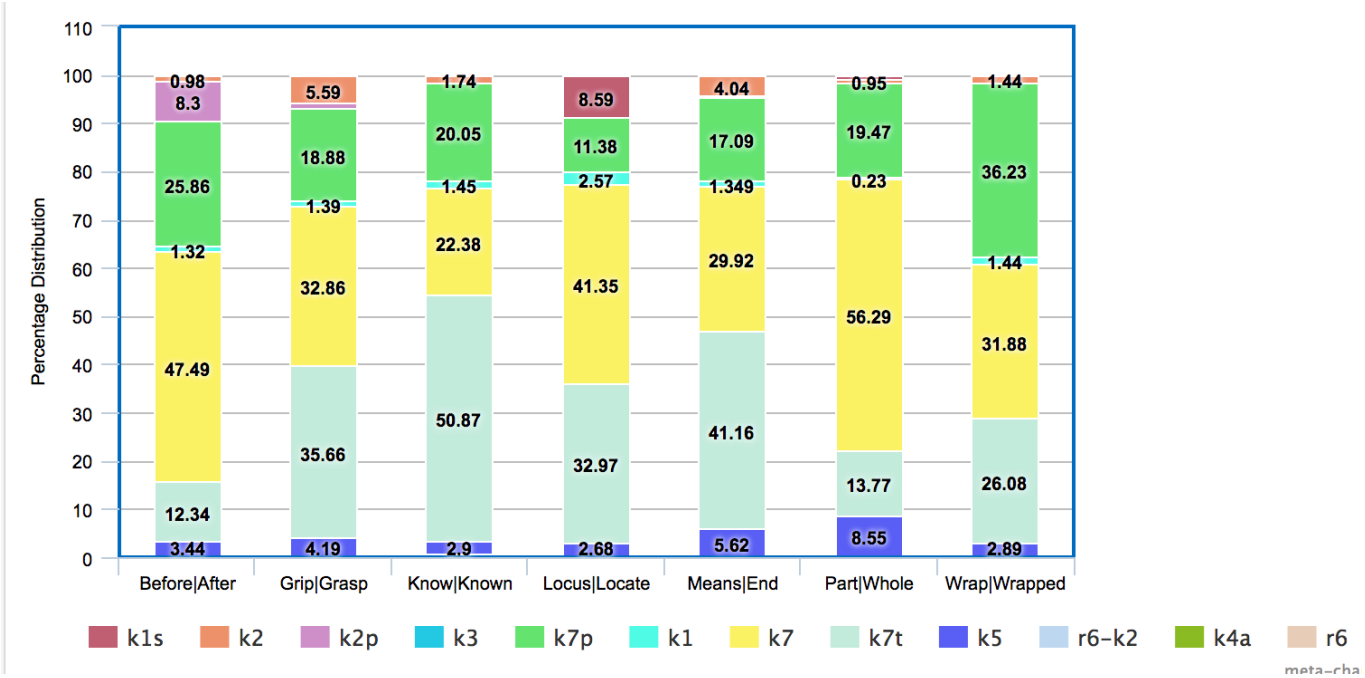

Figure 5.2 Frquency Distribution of kāraka relation of Verb sense-types with NST nouns

5.5.3 Spatial and temporal kāraka or case markers

Figure 5.3, shows association of spatial and temporal kāraka 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.

5.5.4 Spatial, Temporal Adverbs

Figure 5.4 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.

As observed from the preceding discussions, it becomes apparent that the verbs categorized as Before|After, Locus|Locate, and Means|End exhibit a more pronounced affinity with spatio-temporal attributes. This observation aligns harmoniously with their fundamental semantic connotations. These particular verbs, rooted in concepts of movement, action, and existence, naturally necessitate the presence of space and time modifiers to elucidate their meanings fully.

In the next section, we establish the statistical test to quantify these association measures.

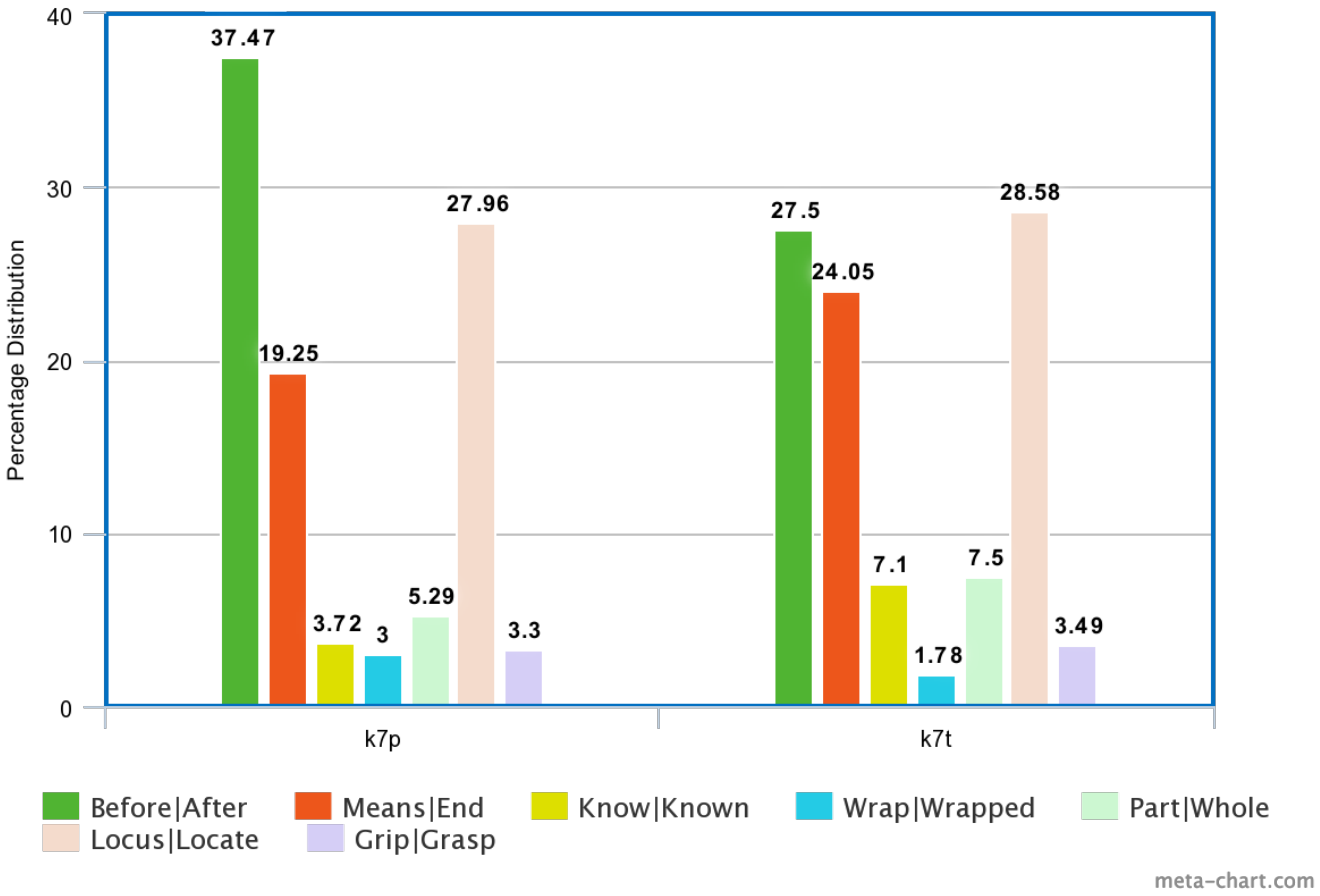

Figure 5.3 Frequency Distribution of Verb sense-types with spatial(k7t) and temporal(k7p) kārakasfor common nouns

5.6 Statistical Measure

5.6.1 Log-Likelihood

We present log-likelihood for the associations of verb-types with different space-time marker. Table 5.2 contains log-likelihood for all the seven types of verb-types with TAM marker.

From the analysis of the association measures presented in Table 5.2 and Table 5.3, a notable synergy emerges between the Log-Likelihood scores and Bayesian Posterior Probabilities. This synergy underscores the validity and significance of the identified associations between verb types and their corresponding TAM markers.

One interesting observation is drawn from the Locus|Locate type of verb. This verb, fundamentally implying existence, exhibits a prominent co-occurrence with the 'WA' TAM marker. The 'WA' marker, with its fundamental meaning of 'existed,' aligns exceptionally well with the core essence of the Locus|Locate verb type. This remarkable synergy between the two measures

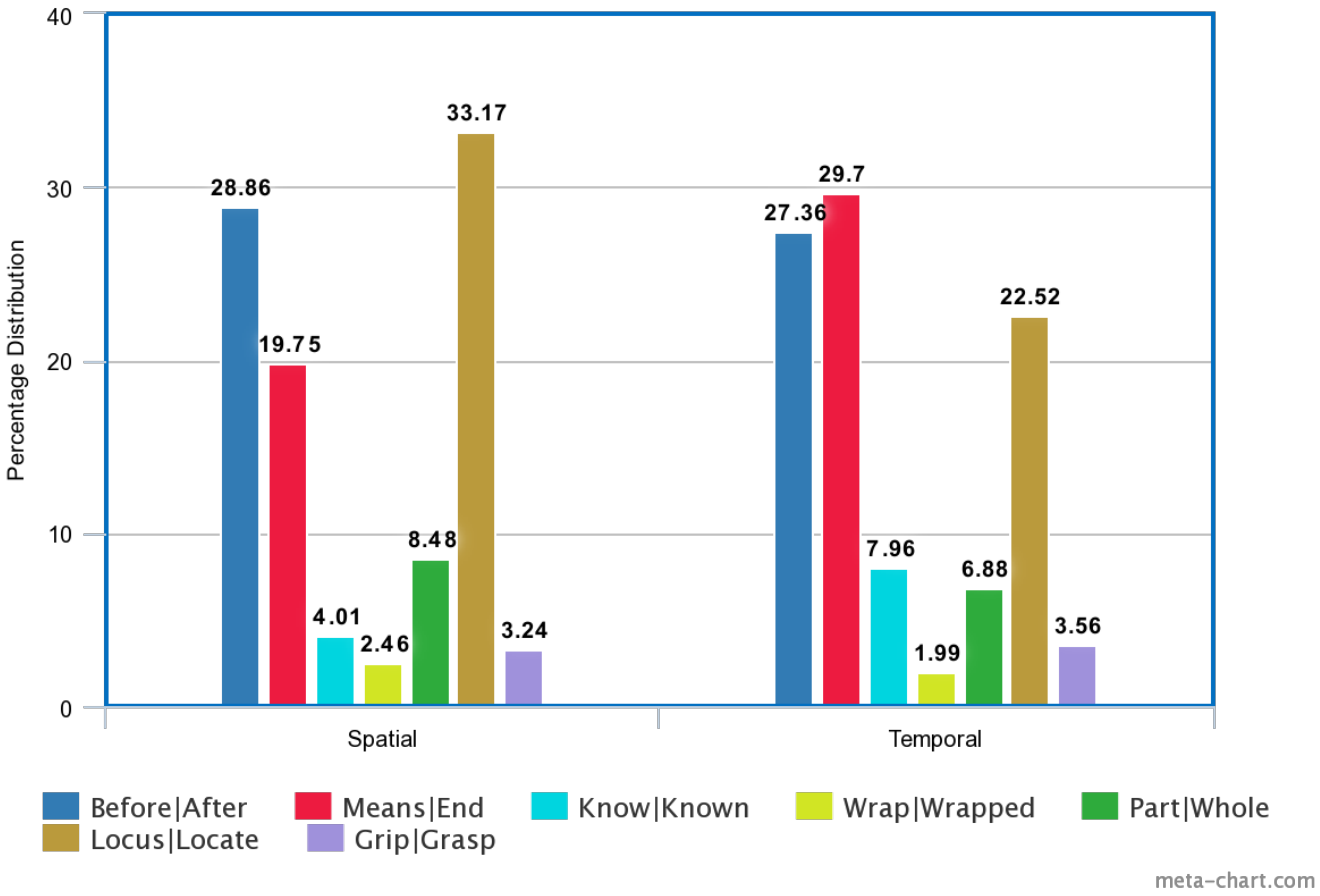

Figure 5.4 Frequency Distribution of Verb sense-types and spatial, temporal adverbs

further supports the linguistic intuition that the 'WA' TAM marker is closely linked to the concept of existence within the context of the Locus|Locate verb."

5.7 Conclusion

The analysis revealed that Before|After, Locus|Locate, and Means|End verb sense-types exhibit more associativity with spatio-temporal attributes. This observation is consistent with their fundamental meanings, which often involve motion, action, and existence in relation to space and time. These findings provide valuable insights into the interaction between verb sense-types and various spatial and temporal components, elucidating language-specific frameworks and features. Further sub-classification of spatial and temporal attributes may deepen our understanding of verb meanings. Additionally, comparing the frequency distributions of spatio-temporal attributes across different languages and corpus types could be an interesting avenue for future research.

Verb Type	**e**	**nA**	**0**	**kara**	**yA**	**gA**	**wA**	**WA**	**hE**
Wrap|Wrapped	91.493	62.174	61.658	44.120	4.625	-119.281	-8.423	-3.678	-6.636
Before|After	181.079	-20.894	-68.626	17.237	-74.250	121.737	83.972	-6.042	-9.001
Part|Whole	-50.590	54.886	164.685	81.720	40.516	-97.336	-25.781	-4.989	-7.947
Locus|Locate	-174.040	-80.311	-125.427	-49.721	-169.870	30.179	-68.716	409.166	1936.598
Know|Known	21.231	28.523	-34.172	-25.499	30.978	51.268	27.085	-4.329	-7.287
Means|End	-22.609	-23.437	137.054	-30.735	359.055	-49.347	-29.235	-6.737	-9.695
Grip|Grasp	29.936	25.270	-53.672	19.873	-99.974	151.507	37.458	-3.571	-6.529

Table 5.2 Log Likelihood Ratios for Verb-Type- TAM Associations

Verb-Type	**e**	**nA**	**0**	**kara**	**yA**	**gA**	**wA**	**WA**	**hE**
Wrap|Wrapped	0.2375	0.1102	0.1649	0.0664	0.2108	0.0800	0.1267	0.0017	0.0017
Before|After	0.2246	0.0841	0.1316	0.0519	0.1995	0.1663	0.1409	0.0005	0.0005
Part|Whole	0.1901	0.0985	0.1737	0.0657	0.2181	0.1276	0.1246	0.0009	0.0009
Locus|Locate	0.1173	0.0460	0.0786	0.0189	0.1295	0.1617	0.1020	0.0618	0.2841
Know|Known	0.2080	0.0954	0.1296	0.0412	0.2186	0.1662	0.1384	0.0013	0.0013
Means|End	0.1998	0.0847	0.1530	0.0468	0.2406	0.1466	0.1278	0.0004	0.0004
Grip|Grasp	0.2144	0.0978	0.1139	0.0582	0.1569	0.2090	0.1462	0.0018	0.0018

Table 5.3 Posterior Probabilities for Verb-Modifier Associations

Verb Type	**k7p**	**k7t**	**k7**	**k1s**	**k5**	**k1**	**k2**	**k2p**
Locus|Locate	565.6809	1783.3810	-449.5441	0.0000	0.0000	-11.7627	0.0000	0.0000
Grip|Grasp	310.1471	0.0000	-27.2316	0.0000	0.0000	-0.9980	0.0000	0.3318
Means|End	0.0000	0.0000	205.5868	0.0000	0.0000	8.7500	109.7253	0.0000
Before|After	0.0000	0.0000	1177.3990	1.3495	66.2228	20.5165	0.0000	144.5017
Know|Known	279.6721	0.0000	-38.4790	0.0000	0.0000	0.2781	-0.2836	0.0000
Part|Whole	0.0000	0.0000	193.0888	17.0527	40.6236	-0.7931	0.0000	0.0000
Wrap|Wrapped	154.3280	10.6090	-25.2718	0.0000	0.0000	0.0000	0.0000	0.0000

Table 5.4 Log Likelihood Ratios for Verb-Type Karaka Modifier with NST Associations

Chapter 6

Unraveling the Intricacies of Verb-Adverb Associativity: A Corpus-Based Analysis in Hindi

6.1 Introduction

Collocation extraction is a crucial task in linguistics and natural language processing (NLP) as it helps reveal meaningful word associations that frequently co-occur in language usage. One prominent concept employed in collocation extraction is the use of association measures, with log-likelihood being a widely utilized measure. These measures help quantify the strength of association between words, aiding in the identification of significant collocations. There have been numerous studies that have employed log-likelihood and related association measures to derive the strength of different collocation types e.g. verb-adjective, verb-noun, and verb-adverb collocations [Sabine Schulte im Walde (2006), Yusuke Takagi and Danushka Bollegala (2012), Pascale Fung and Lo Yuen Yee (1998)].

Language, as a rich tapestry of expression, weaves together various elements to communicate nuanced meanings and convey emotions. One fundamental aspect that contributes to the fluency and naturalness of language is the collocation of words, where certain words have an innate tendency to co-occur with others. This chapter delves into the captivating realm of verb-adverb associativity in the Hindi language. With a focus on uncovering specific verb-adverb collocations, we aim to shed light on the intricate patterns that govern their combination, thereby contributing to a deeper understanding of Hindi language usage. Drawing upon a comprehensive corpus of written and spoken Hindi texts, we employ a corpus-based approach to analyze the frequency, distribution, and semantic associations of these collocations. By elucidating the unique characteristics of verb-adverb associativity in Hindi, we seek to enrich our language comprehension and provide valuable insights.

Table 6.1 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

6.2 Analyzing Verb-Adverb Associativity

:

Word frequency comparison is a fundamental approach in corpus linguistics, commonly used for various tasks like hypothesis generation and providing a basis for further investigation. In this study, we specifically concentrate on evaluating the statistical significance of disparities in the occurrence frequencies of verb-adverb collocations in our Hindi corpus. Our primary objective is to address the inquiry, "Does verb-type X exhibit higher frequency with adverb-class Y?"

To begin our exploration of verb-adverb associativity in the Hindi language, we use the ontological categories of verb and adverb as defined in Chapter 2. These semantic categories serve as essential building blocks to identify and analyze the collocations in our corpus.

6.2.1 Data

Table 6.1 shows different corpus collected from writings of different authors.

In addition to the above data, we also used Hindi Treebank Data [6]

6.2.2 Dependency Relation

We used Hindi dependency parser[1] to extract these depedency relations. The dependency parser uses guidelines from AnnCorra [5] for annotations. It is treebanks for Indian Languages Guidelines for annotating Hindi Treebank. The annotation is stored in SSF format [4].

Verbs, and their modifying adverbs were extracted from the dependency parser output. These verbs and adverbs were further annotated using the ontological resource that we have created.

Figure 6.1 shows verb-adverb type-class collocation extraction

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

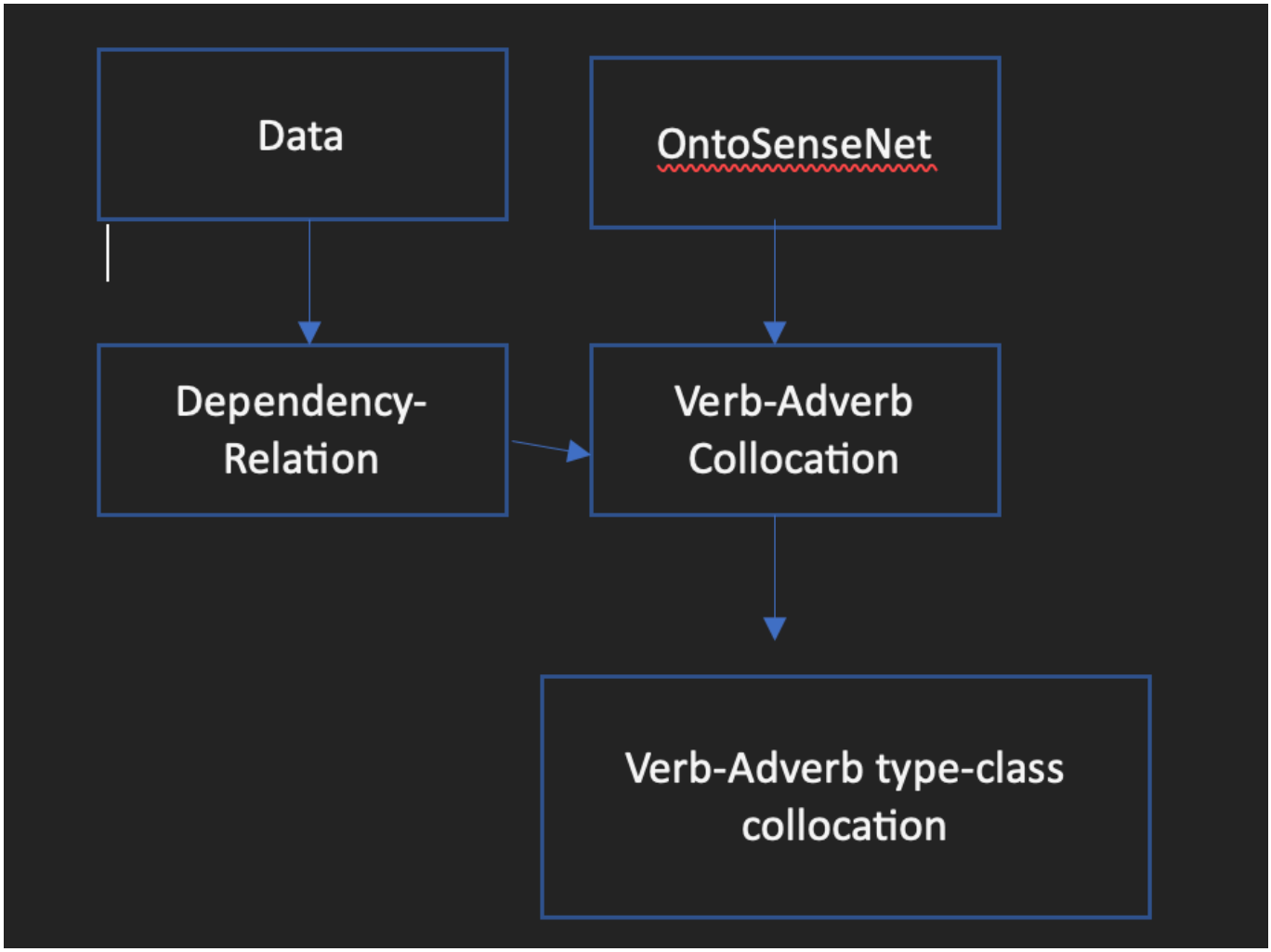

Figure 6.1 Design Diagram of Verb-Adverb Collocation extraction

6.3 Statistical Measures for Associativity

To measure the strength of associativity between verb and adverb pairs, we utilize statistical measures, prominently focusing on the log-likelihood measure. The log-likelihood ratio allows us to quantify the degree to which a verb-adverb collocation deviates from the expected frequency of their co-occurrence, indicating stronger associations.

Log likelihood is a statistical measure used to determine the strength of association between two events. In the context of collocation extraction, it helps to assess whether two words (i and j) occur together more frequently than expected by chance. The formula for log likelihood is as follows:

$$\text{Log Likelihood} = 2 \times \left(\sum_{i,j} O_{ij} \times \log \left(\frac{O_{ij}}{E_{ij}} \right) \right)$$

The log likelihood value tells us how much the observed frequency differs from the expected frequency. If the value is positive, it indicates that words A and B occur together more often than expected (positive association). If the value is negative, it suggests that they occur together less frequently than expected (negative association). A higher absolute log-likelihood value indicates a stronger association between the two words.

6.4 Type-Class Associativity

Figure 6.2 shows verb-adverb type-class log-likelihood

6.5 Discussion

Verb "Know|Known" and Adverb "Measure": The verb "Know|Known" is associated with the adverb "Measure" with the highest log-likelihood value. Semantically, "Know|Known" represents the action of attaining information or knowledge. It indicates the process of gaining awareness or understanding about something. The adverb "Measure" quantifies this action, indicating the extent or degree of information gained by the subject.

For instance, consider the sentence: **"आज वह बहुत पढ़ा।"** He studied a lot today.

Here, the adverb **"बहुत"** (a lot) belongs to the "Measure" class and modifies the verb **"पढ़ा"** (studied), which belongs to the "Know|Known" verb type. The adverb **"बहुत"** quantifies the action of studying, expressing that the subject studied to a significant extent or to a great degree on that particular day.

Verb Type "Locus|Locate" and Adverb Class "Force":

On the other hand, the verb type "Locus|Locate" exhibits the least associativity with the adverb class "Force." The "Locus|Locate" verbs are primarily associated with stative verbs, which describe a state or condition of existence. They refer to verbs that indicate the location or positioning of something without necessarily conveying an action.

The adverb class "Force" typically includes adverbs that express the intensity, strength, or exertion of an action. Adverbs in this class often modify dynamic verbs that imply actions or activities. Since "Locus|Locate" verbs are more closely linked to stative verbs, their association with adverbs denoting forceful actions is comparatively lower.

6.6 Conclusion

In conclusion, this chapter delves into the fascinating realm of verb-adverb associativity in the Hindi language. By employing a corpus-based approach and statistical measures, we have examined the frequency, distribution, and semantic connections of verb-adverb collocations in our comprehensive Hindi corpus.

The log-likelihood measure has proven to be a valuable tool in assessing the strength of associations between verbs and adverbs. Our analysis revealed that certain verb-adverb combinations, such as "Know|Known" with "Measure," display a high log-likelihood value, indicating a robust and meaningful association. Semantically, "Know|Known" verbs denote the acquisition of knowledge, and when paired with the "Measure" adverb, the extent of information gained by the subject is quantified.

Moreover, our exploration has shed light on the intriguing linguistic patterns governing verb-adverb collocations. We observed that "Locus|Locate" verb types show the least associativity with the "Force" class of adverbs. As "Locus|Locate" verbs are often linked to stative verbs indicating existence or location, their connection with adverbs expressing forceful actions is relatively low.

These findings have significant implications for language learning, computational linguistics, and natural language processing applications. Understanding verb-adverb associations in Hindi enhances language comprehension and can lead to the development of more sophisticated language processing models.

Means|End-Spatial	8897	-3.45
Means|End-Temporal	29057	-87.28
Means|End-Measure	17388	+8.63
Means|End-Force	24611	+8.08
Before|After-Spatial	1012	-5.60
Before|After-Temporal	3486	-12.61
Before|After-Measure	2270	+0.54
Before|After-Force	3739	-6.94
Know|Known-Spatial	1740	-1.67
Know|Known-Temporal	6074	+3.62
Know|Known-Measure	5265	+113.14
Know|Known-Force	5772	+ 41.53
Grip|Grasp-Spatial	1504	-4.56
Grip|Grasp-Temporal	2868	-9.97
Grip|Grasp-Force	3119	+2.87
Grip|Grasp-Measure	2776	+3.38
Part|Whole-Spatial	111	+0.02
Part|Whole-Temporal	252	+0.28
Part|Whole-Measure	292	+1.30
Part|Whole-Force	372	-0.01
Locus|Locate-Spatial	3340	+0.05
Locus|Locate-Temporal	7649	-17.44
Locus|Locate-Measure	4485	+17.34
Locus|Locate-Force	6570	+0.0
Wrap|Wrapped-Spatial	987	-4.35
Wrap|Wrapped-Temporal	621	+0.0
Wrap|Wrapped-Measure	1187	-0.02
Wrap|Wrapped-Force	944	+0.79

Figure 6.2 Frequency Distribution and log-likelihood

Chapter 7

Conclusion and Future Work

The exploration into the association between verb types and spatio-temporal attributes in Hindi has yielded significant insights into the structure and semantics of the language. Through statistical measures such as log-likelihood and Bayesian inference, we've discerned meaningful patterns in verb-adverb and verb-noun collocations, shedding light on how space and time are encoded in Hindi discourse. In conclusion, our findings underscore the nuanced interplay between verbs and their spatio-temporal contexts. The prominence of certain verb sense-types, such as Before|After, Locus|Locate, and Means|End, in association with spatio-temporal markers highlights the intrinsic relationship between linguistic expression and concepts of motion, action, and existence within a temporal and spatial framework. These insights contribute not only to our understanding of Hindi language structures but also to broader discussions in linguistics and natural language processing.

As for future work, refining the analysis by delving deeper into specific verb types and their associations with spatial and temporal markers could provide a more detailed understanding of verb semantics in Hindi. Furthermore, extending this study to compare similar associations across different languages or dialects could offer valuable cross-linguistic insights. Additionally, exploring how these linguistic patterns manifest in different genres or registers of discourse could provide further depth to our understanding of language usage in varied contexts.

Related Publications

1. Jha, J., Parupalli, S., Singh, N. (2018, March). OntoSenseNet: a verb-centric ontological resource for indian languages. In International Conference on Computational Linguistics and Intelligent Text Processing (pp. 32-45). Cham: Springer Nature Switzerland.
2. Developing OntoSenseNet: A Verb-Centric Ontological Resource for Indian Languages and Analysing Stylometric Difference in Male and Female Writing using the Resource; Jyoti Jha and Navjyoti Singh; Research Proposal ACL-SRW (2019)

Bibliography

[1] LAC hindi corpus, 2014. LINDAT/CLARIN digital library at the Institute of Formal and Applied Linguistics, Charles University.

[2] J. Barwise and J. Perry. Situation semantics. *MIT Press*, 183:184, 1983.

[3] A. Bharati. Anncorra: Annotating corpora guidelines for pos and chunk annotation for indian languages. 2006.

[4] A. Bharati, R. Sangal, and D. M. Sharma. Ssf: Shakti standard format guide. *Language Technologies Research Centre, International Institute of Information Technology, Hyderabad, India*, pages 1–25, 2007.

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

[6] R. A. Bhat, R. Bhatt, A. Farudi, P. Klassen, B. Narasimhan, M. Palmer, O. Rambow, D. M. Sharma, A. Vaidya, S. R. Vishnu, et al. The hindi/urdu treebank project. In *Handbook of Linguistic Annotation*. Springer Press.

[7] B. Bhattacharya. *Yāska's Nirukta and the Science of Etymology: An Historical and Critical Survey*. Firma KL Mukhopadhyay, 1958.

[8] P. Bhattacharyya. IndoWordNet. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, Valletta, Malta, May 2010. European Language Resources Association (ELRA).

[9] M. Björklund and J. Viimaranta. Russian, swedish, and finnish adpositions meaning 'under' and 'over' in temporal constructions. *Time and Language*, page 13.

[10] F. Brentano. *Philosophical Investigations on Time, Space and the Continuum (Routledge Revivals)*. Routledge, 2009.

[11] D. Chen and C. Manning. A fast and accurate dependency parser using neural networks. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 740–750, 2014.

[12] N. Chomsky. Syntactic structures. In *Syntactic Structures*. De Gruyter Mouton, 2009.

[13] H. M. Cook. The japanese verbal suffixes as indicators of distance and proximity. *The construal of space in language and thought*, pages 3–27, 1996.

[14] T. Dunning. Accurate methods for the statistics of surprise and coincidence. *Computational linguistics*, 19(1):61–74, 1994.

[15] J. Grimshaw. *Argument structure.* the MIT Press, 1990.

[16] R. S. Jackendoff. Semantic interpretation in generative grammar. 1972.

[17] Y. Kachru and R. Pandharipande. Towards a typology of compound verbs in south asian languages. *Studies in the Linguistic Sciences*, 10(1):113–124, 1980.

[18] C. H. Kahn. The verb" be" in ancient greek. 1973.

[19] A. Kilgarriff. Language is never, ever, ever, random. 2005.

[20] G. Lakoff. On generative semantics. *Semantics: An interdisciplinary reader in philosophy, linguistics and psychology*, 232:296, 1971.

[21] R. W. Langacker. 10 the contextual basis of cognitive semantics. *Language and conceptualization*, 1:229, 1999.

[22] A. Leeuwenberg, M. Vela, J. Dehdari, and J. van Genabith. A minimally supervised approach for synonym extraction with word embeddings. *The Prague Bulletin of Mathematical Linguistics*, 105(1):111–142, 2016.

[23] G. W. Leibniz. The labyrinth of the continuum: Writings on the continuum problem, 1672-1686. 2001.

[24] T. Mikolov, K. Chen, G. Corrado, and J. Dean. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*, 2013.

[25] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, pages 3111–3119, 2013.

[26] G. A. Miller. *WordNet: An electronic lexical database*. MIT press, 1998.

[27] Y. Mimānsak. *Sanskrit Vyakaran Shastra ka Itihas*. Yudhishthir Mimānsak, 1985.

[28] D. Narayan, D. Chakrabarti, P. Pande, and P. Bhattacharyya. An experience in building the indo wordnet-a wordnet for hindi. In *First international conference on global WordNet, Mysore, India*, volume 24, 2002.

[29] H. Ogawa. Process & language: A study of the mahabha, sya ad a1. 3.1 bhuvadayo dhatavah. *Motilal Banarsidass, Delhi,*, 2005.

[30] S. Otra. *Towards Building a Lexical Ontology Resource Based on Intrinsic Senses Of Words*. PhD thesis, International Institute of Information Technology Hyderabad, 2015.

[31] B. H. Partee. *Montague grammar*. Elsevier, 2014.

[32] S. Parupalli. *Towards Developing a Lexical Ontology Resource and Augmenting Novel Approaches for Sentiment Analysis Task through Enrichment of Available Resources in Telugu*. PhD thesis, International Institute of Information Technology, 2018.

[33] K. H. Potter. *The Encyclopedia of Indian Philosophies, Volume 3: Advaita Vedanta Up to Samkara and His Pupils*, volume 3. Princeton University Press, 2014.

[34] K. Rajan. Ontological classification of verbs based on overlapping verb senses. *International Institute of Information Technology*, 2015.

[35] G. Ramchand. *Verb meaning and the lexicon: A first-phase syntax*, volume 116. Cambridge University Press Cambridge, 2008.

[36] L. Sarup. *The Nighantu and the Nirukta: The oldest Indian treatise on etymology, philology and semantics*. Motilal Banarsidass Publ., 1998.

[37] S. B. R. P. D. Singh and P. Bhattacharyya. Merging verb senses of hindi wordnet using word embeddings. In *11th International Conference on Natural Language Processing*, page 344, 2014.

[38] J. F. Staal and F. Staal. *A reader on the Sanskrit grammarians*. MIT Press Cambridge, MA, 1972.

[39] T. Uchihara, M. Eguchi, J. Clenton, K. Kyle, and K. Saito. To what extent is collocation knowledge associated with oral proficiency? a corpus-based approach to word association. *Language and Speech*, 65(2):311–336, 2022.

[40] A. N. Whitehead and B. Russell. *Principia mathematica to* 56*, volume 2. Cambridge University Press, 1997.

[41] A. Wierzbicka. Semantic primitives. 1972.