

BUILDING BENGALI VERB-FRAME NETWORK: A PANINIAN DEPENDENCY APPROACH

A dissertation submitted in partial fulfilment
of the requirements for the award of the degree of

Doctor of Philosophy

in

Applied Linguistics

By

Arunava Kar

17HAPH01



Supervisor:

Prof. K. Rajyarama

Centre for Applied Linguistics and Translation Studies

Centre for Applied Linguistics and Translation Studies

School of Humanities

University of Hyderabad

India

June, 202



CERTIFICATE

This is to certify that the thesis titled “**Building Bengali Verb-frame Network: A Paninian Dependency Approach**” submitted by **Arunava Kar**, bearing registration number **17HAPH01**, in partial fulfillment of the requirements for the award of Doctor of Philosophy in English is a bonafide work carried out by him under my supervision and guidance and has not been submitted in part or in full to this University or any other University or Institution for the award of any degree or diploma.

The thesis is free from plagiarism and has not been submitted previously, either in part or in full, to this or any other University or Institution for award of any degree or diploma.

Further, the student has the following publication before submission of the thesis for adjudication and has produced evidence for the same in the form of acceptance letter or the reprint in the relevant area of his research:

1. Use of Ontology in Bengali Verb Frame and its importance in Verb Sense Disambiguation, *Jadavpur Journal of Language and Linguistics*, Vol 4, IC3C Proceedings, p. 161-172, ISSN: 2581-494X (UGC-CARE)

and

has made presentations in the following papers in the conferences:

1. *Developing Bengali Verb Frame Network* at ‘The 44th International Conference of Linguistic Society of India’ held at Bhubaneswar, Odisha from 24th to 26th November, 2022
2. *The Riddle of Impersonal Subject Construction in Bengali: An Alternative Descriptive Approach* at Graduate Research Meet held at Guwahati, on 6th and 7th January, 2023
3. Positioning Bengali Passives within the Framework of Impersonal Construction: A

Verb-frame Approach at the seminar ‘All About Voice: A Crosslinguistic Perspective’ held at IIT Madras, Chennai, Tamil Nadu on 14th and 15th April, 2023.

Further, the student has completed the following courses towards the fulfillment of coursework requirement for PhD:

Code	Name	Credit	Pass/Fail
AI-801	Research Methodology	IV	Pass
AI-821	Readings in Applied Linguistics	IV	Pass
EG-825	Academic Writing for Doctoral Students	IV	Pass

Supervisor

Head of the Department

Dean of the School



DECLARATION

I, Arunava Kar, hereby declare that this dissertation titled “**Building Bengali Verb-frame Network: A Paninian Dependency Approach**” submitted to the University of Hyderabad in partial fulfilment of the requirements for the award of the degree of **Doctor of Philosophy in Applied Linguistics**, is a bonafide research work carried out by me under the guidance and supervision of **Prof. K. Rajyarama**, Centre for Applied Linguistics and Translation Studies, School of Humanities, University of Hyderabad.

I declare that this thesis has been composed solely by myself and that it has not been submitted, in whole or in part, to this or any other university or institution for the award of any degree or diploma. Except where stated by reference or acknowledgment the work presented is entirely my own.

I declare that this research work is free from plagiarism. A report on the plagiarism statistics from the Indira Gandhi Memorial Library, University of Hyderabad is enclosed. I hereby agree that my thesis can be deposited in Shodhganga/INFLIBNET.

Date- 30-06-2023
Place- Hyderabad

Arunava Kar
Reg No: 17HAPH01
CALTS

Countersigned by:

Prof. K. Rajyarama
Supervisor, CALTS
University of Hyderabad

ACKNOWLEDGEMENT

This research would not have been possible without the help of my supervisor, Prof. K. Rajyarama. I owe this research to my supervisor. It has been an honor to work under the guidance of my supervisor who has been a constant source of inspiration and invaluable moral support all through these years. Whenever I faced any problems and approached her, she was always ready to help me find a solution. I can't thank her enough for her contribution.

I thank my research advisory committee member Prof. S. Arulmozi and Dr. K. Parameswari for their support and motivation to continue and complete my research program. Their valuable feedback and suggestions helped me to complete the thesis.

I am indebted to Prof. Niladri Sekhar Dash for his constant support throughout the course of my Ph.D. Without his crucial support, it would not be possible to develop Bengali Verb-frame Network. I will use this opportunity to express gratitude towards both Ministry of Electronics and Information Technology and Prof. Niladri Sekhar Dash for the permission to use TDIL corpus for the development of verb-frames.

I would also like to thank Monalisha, junior research scholar and friend, who thoroughly checked and corrected the draft of this dissertation. Her contribution cannot be overstated. Her effort was instrumental in bringing this dissertation in its present state.

Arunava Kar

June, 2023

CONTENTS

Certificate	i
Declaration	iii
Acknowledgement	iv
Contents	v
List of Figures and Tables	x
List of Verb-frames	xi
List of Abbreviations	xiii
Transliteration	xiv

Chapter 1

Introduction	1
1.1 Bengali Verb-frame Network: An Introduction	1
1.2 Verb-frame in The Present Context	2
1.3 Dependency Grammars vis-a-vis Constituency Grammar	3
1.3.1. Dependency Formalism in the Context of Verb-frame	4
1.4 Framework: Paninian Dependency Model	5
1.5 The Objectives	7
1.6 The Research Question	8
1.7 Problem Statement and Scope	8
1.8 Justification and Likely Benefits	9
1.9 Organisation of the Thesis	10

Chapter 2

Background	13
2.1. Theoretical Approaches	13
2.1.1. Early Initiatives in Generative Semantics	14
2.1.2. Jackendoff's Lexical Conceptual Structure	16
2.1.3. Fillmore's Semantic Frame	19
2.1.4. Generative Lexicon	22

2.1.5.	Verb in the Indian Grammatical Tradition	26
2.2.	NLP Resources	29
2.2.1.	Levin's Verb Classification	29
2.2.2.	VerbNet Project	31
2.2.3.	FrameNet Project	32
2.2.4.	PropBank	33
2.2.5.	Hindi Dependency Treebank	34
2.3	Ontology	35

Chapter 3

Methodology	38
3.1 Lexical Ambiguity and Verb-frame	38
3.1.1 Homonymy and Polysemy	38
3.2 Event Semantics	39
3.2.1 Event Structure and Verb Sense	39
3.2.2. The Approach to Resolve Lexical Ambiguity by Using Verb-frame	41
3.3 Computational Paninian Grammar	41
3.4 Motivation	44
3.5 Modus Operandi	44

Chapter 4

Architecture of the Verb-frame	48
4.1 Dependency Relation	48
4.2 Lexical Type	55
4.3 Dependency Marker	56
4.4 Ontology	60

4.4.1. Defining Ontological Terms	63
4.4.2. Importance and Issues	67
4.5 Full Verb-frame of <i>Daakaa_1</i> and <i>Daakaa_2</i>	71

Chapter 5

Analysis and Development	74
5.1. Organisation of the Chapter	74
5.2. Some Important Features of the Bengali Verbal System	74
5.3. Verb Senses: Number and Statistics	76
5.4. Some Basic Concepts	80
5.4.1 Essential and Non-essential Roles	80
5.4.2 Core Verb-frame	81
5.4.3 Case Syncretism	82
5.5. Verb-frames	82
5.5.1 Verb-frames of the Verb <i>balaa</i>	83
5.5.2 Some Other Verb-frames	89
5.6. Verbs with Multiple Verb-Frames: Theory and Statistics	100
5.7. Verbs and Lexical Types	102
5.7.1. Verbs and the Clausal Argument	102
5.7.2. Verbs and the Infinitival Argument	104
5.7.3. Verbs and the Adverbial Argument	106
5.7.4 Verbs and the Adjectival Argument	107
5.8 Frequency count of Dependency Relations	108
5.9 Verb-frames and Ontology	110
5.10 Some Special Class of Verbs	113

5.10.1	Causative Verbs	114
5.10.2	Impersonal Verbs	118
5.10.3	Verbs of Human Interaction	122
5.10.4	Conjunct Verbs	124
5.11	Verb Classification	126
5.11.1	Verb Class 1	128
5.11.2	Verb Class 2	130
5.11.3	Verb Class 3	135
5.11.4	Verb Class 6	137
5.11.5	Verb Class 7	138
5.11.6	Verb Class 11 and 17	139
5.11.7	Verb Class 8	141
5.11.8	Verb Class 16	142
5.11.9	Verb Class 5	143
5.11.10	Verb Class 30	143
5.11.11	Significance of Verb Classification	144
5.12	Corpus Observations	145
5.13	Final Remarks	149

Chapter 6

Conclusion and Future Research		151
6.1	Accomplishment of the Research	151
6.2	Aims and Objectives	152
6.3	Scope of Future Research	152
	6.3.1 Possible Practical Applications	153

6.4	Final Words	154
	Appendix I	156
	Appendix II	164
	References	168

LIST OF FIGURES AND TABLES

List of Figures

SR No.	Heading	Page
1.1	Tree Structure of Constituency and Dependency Grammars	4
2.1	Conceptual Structure of Sentence (1)	17
2.2	Specialised Formation Rules for Conceptual Structures	18
2.3	Derivation of Sentence in Paninian Grammar	28
3.1	AnnCorra Dependency Scheme	43
4.1	Ontological Hierarchy	64
5.1	Relative Frequency of Verb Senses Per Verb (Histogram)	79
5.2	Relative Frequency of Verb Senses per Verb (Pie Chart)	80
5.3	Verbs with Multiple Verb-frames (Pie Chart)	102

List of Tables

SR No.	Heading	Page
2.1	Verb Frame for class ‘Hit’ in VerbNet project	32
4.1	Dependency Relations and Dependency Markers	58
5.1	Non-finite Verb Forms in Bengali	75
5.2	List of Verb Senses	76
5.3	Frequency of Dependency Relation in BVFN	109
5.4	Verb Classification	127
5.5	Frequency of Different Lexical Types in Corpus and Dictionary	146

LIST OF VERB-FRAMES

SR No.	Verb-frame	Page
1	<i>Daakaa_1</i> ‘to call’	72
2	<i>Daakaa_2</i> ‘to make a noise’ (animal)	73
3	<i>balaa_1</i> ‘to tell’	83
4	<i>balaa_2</i> ‘to chant’	86
5	<i>balaa_3</i> ‘to be called/ pronounced as’	87
6	<i>balaa_4</i> ‘to speak as representative	88
7	<i>kaaTaa_1</i> ‘to cut’	90
8	<i>kaaTaa_2</i> ‘to deduct’ (amount of money)	91
9	<i>kaaTaa_3</i> ‘to dig’	92
10	<i>kaaTaa_4</i> ‘to recover or be okay’	93
11	<i>kaaTaa_5</i> ‘to mark with a cross or a line’	94
12	<i>kaaTaa_6</i> ‘to pass the time’	95
13	<i>naamaa_1</i> ‘to get down’	96
14	<i>naamaa_2</i> ‘to start or be involved in some activity’	97
15	<i>naamaa_3</i> ‘to reduce or decrease’	98
16	<i>saaraa_1</i> ‘to recover’	99
17	<i>saaraa_2</i> ‘to execute’	100
18	<i>bojhaa_1</i> ‘to understand’	104
19	<i>paaraa_1</i> ‘to be able to’	106
20	<i>pheraa_2</i> ‘to turn around’	107
21	<i>bakaa_1</i> ‘to speak incoherently’	108
22	<i>calaa_1</i> ‘to walk’	111
23	<i>calaa_4</i> ‘to have a particular time’	112
24	<i>baanaano_1</i> ‘to change into’	118
25	<i>laagaa_7</i> ‘to feel’ (usually temperature, pain etc)	122

26	<i>kathaa_balaa_I</i> ‘to talk’	124
27	<i>tairii karaa_I</i> ‘to make’	126

LIST OF ABBREVIATIONS

Glossing Abbreviations

1	first person
2	second person
3	third person
ACC	accusative
ADJ	adjective
ADV	adverb(ial)
AUX	auxiliary
CAUS	causative
DAT	dative
FUT	future
GEN	genitive
INF	infinitive
LOC	locative
NEG	negation, negative
OBJ	object
PL	plural
PRF	perfect
PRS	present
PROG	progressive
PST	past
PTCP	participle
SG	singular
TR	transitive
NLP	Natural Language Processing
SCB	Standard Colloquial Bengali

‘Itrans’ Transliteration scheme is used for transliteration

Bengali Script	Itrans	Bengali Script	Itrans
অ	a	ফ	ph
আ	aa/A	ব	b
ই	i	ভ	bh
ঈ	ii	ম	m
উ	u	য	j
ঊ	uu	র	r
ঋ	RRi	ল	l
এ	e	ব	v
ঐ	ei	শ	sh
ও	o	স	s
ঔ	oi	ষ	Sh
ক	k	হ	h
খ	kh	ড়	R
গ	g	ঢ়	Rh
ঘ	gh	য়	y
ঙ	~N	ৎ	t
চ	c	ং	M
ছ	ch	ঃ	H
জ	j	ঁ	.N
ঝ	jh		
ঞ	~n		
ত	t		
থ	th		
দ	d		
ধ	dh		
ন	n		
ট	T		
ঠ	Th		
ড	D		
ঢ	Dh		
ণ	N		
প	p		

CHAPTER 1

INTRODUCTION

The development of useful language resources has become a very important task for both linguists and computational linguists due to the advent of the digital revolution. It is more so for the Bengali language because it is, like most other Indian languages, a low-resource language (Karim et al., 2020). A low-resource language is a language that doesn't have sufficient resources built for computational purposes, such as text and speech corpora, digital lexicons, annotated datasets, and computational tools necessary for developing natural language processing systems and other language technologies. The lack of digital resources hinders the development of accurate and efficient language models and systems, which in turn limits the use and availability of these technologies for speakers of low-resource languages.

Bengali, while it is the seventh most spoken language in the world in terms of the number of native speakers, is considered a low-resource language in the context of natural language processing and machine learning. This is due to factors like the lack of digital resources and limited research and development efforts. While there are some digital resources available for Bengali, such as annotated datasets, corpora, and dictionaries, they are often limited in scope and quality compared to the resources built for languages like English or Chinese. The previous efforts to develop language models and machine translation systems for Bengali have not been as extensive or well-funded as those for other languages.

1.1 Bengali Verb-frame Network: An Introduction

The research presented in this dissertation deals with the development and various aspects of the Bengali Verb Frame Network (henceforth BVFN), a resource that links the verb's meaning with its syntactic behaviour and presents it in a precise, easily-understandable

digital format. It is, in essence, an application-neutral resource that can be applied for various purposes and in various domains ranging from natural language processing to language teaching. A dependency-based formalism is employed to develop the verb frames, and it is a comprehensive collection of verb frames that simultaneously represent layered information regarding the syntax and semantics of the verbs in Bengali. By identifying the selectional restrictions and semantic roles played by the verbs in Bengali sentences, this resource can be used to improve the accuracy and efficiency of many different language technologies. This particular resource can be most useful for verb-sense disambiguation. An ambiguous verb is broken down into different senses, and a verb-frame is created for every sense. It resolves lexical ambiguity as it provides separate verb frames for different verb senses. Overall, BVFN can be an important tool for linguists, language learners, and natural language processing researchers working on the Bengali language.

1.2 Verb-frame in the Present Context

The notion of verb-frame usually refers to the argument structure of a verb or the set of syntactic or semantic roles associated with a particular verb. In other words, the verb-frame describes how a verb typically combines with its arguments in order to form complete sentences. The verb frame can go beyond the simple thematic description of argument structure information. It can describe, for example, the ontological properties of the arguments or the event semantics related to a particular verb. According to Sabine Schulte (2009), verb-frame describes the “functional complements that are evoked by a verb”. The term ‘frame’ itself comes from Fillmorean frame semantics (Fillmore, 1976; Lakoff, 1987) in which ‘frame’ refers to something more than just subcategorisation information. Now, the term is used in the generic sense in order to refer to any kind of argument structure representation. In the section 1.3, we will look at the concept of verb-frame from the perspective of event semantics.

1.3 Dependency Grammars vis-a-vis Constituency Grammar

Constituency grammars interpret sentences as a linked chain of ‘constituent structure’.

A constituent structure is a group of words that operates as one unit. The basic idea is that constituents like noun phrases, verb phrases etc., are the middle layer that mediates between words and the sentences they are part of. It considers a sentence as a hierarchical structure composed of nested phrases. Each phrase consists of one or more words and can be further divided into sub-phrases. The relationships between phrases are typically represented using labelled brackets or tree structures. Constituency grammar is also known as Phrase-structure grammar. This school of grammar can be contrasted with dependency grammar which treats sentences as a series of relations among words termed ‘dependency’. Each word in the sentence is considered a node, and the relationships between words are represented as labelled arcs or dependencies. The dependencies typically indicate the syntactic role of each word in the sentence, such as subject, object, modifier, etc. The structure is often represented as a dependency tree, where the root of the tree represents the main predicate of the sentence, and the edges represent the dependencies between words. The concept of dependency grammar primarily developed from the work of Lucien Tesnière (Tesnière, 1959). Most dependency grammars are verb centric. Dependency relations usually connect verbs with other words in a sentence. Unlike constituency grammars, dependency structures are not ordered hierarchically. The tree below illustrates this point

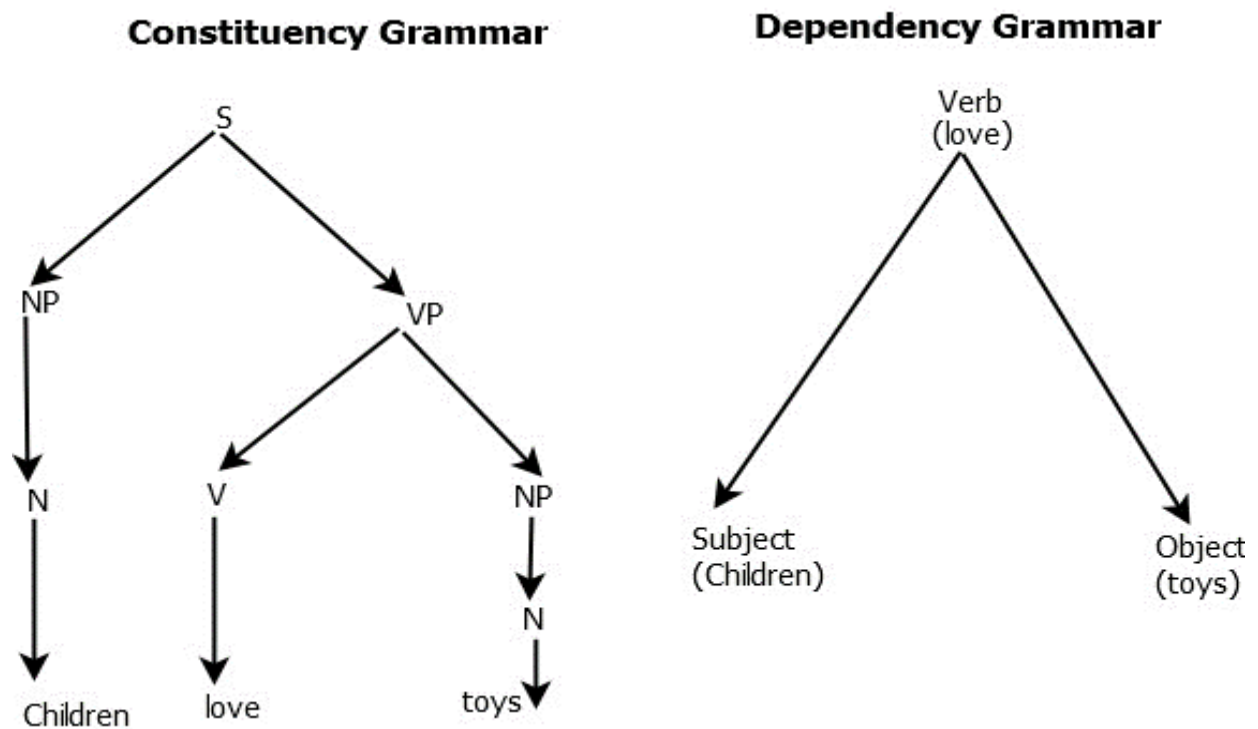


Figure 1.1: Tree Structure of Constituency and Dependency Grammars

Dependency grammars are widely used in natural language processing tasks such as syntactic parsing, machine translation, and information extraction. They offer a straightforward representation of the grammatical relationships between words, and their simplicity makes them computationally efficient.

On the other hand, Constituency grammars are often used in formal syntax and linguistic analysis. They provide a detailed hierarchical representation of the sentence structure, which can be useful for studying the syntax and semantics of a language. However, parsing and processing constituency-based structures can be computationally more complex than dependency-based structures.

1.3.1 Dependency Formalism in the Context of Verb-frame

Dependency grammar focuses on the relationships between individual words, while constituency grammar emphasises the hierarchical structure of phrases within a

sentence. The point of developing a verb-frame is to represent the syntactic-semantic relation the verb has with other elements in a sentence. Though the concept of verb-frame can be applied within both dependency and constituency-based grammar formalism, verb-frame, as a resource, is more useful for dependency grammar formalism. This is due to the fact that verbs play a central role in most of the dependency formalism.

The core part of the Bengali Verb-frame Network is a dependency framework that is based on Paninian Computational Grammar (Bharati et al., 1995). Dependency relations represent the verb's relationship with the other elements in a sentence. The sum of these relations constitutes a verb's dependency structure. In this way, by design, dependency grammars can be used to identify possible verb-frames.

1.4 Framework: Paninian Dependency Model

Paninian grammar is a dependency formalism developed by Panini (4th century B.C.) for classical Sanskrit in ancient India. In Paninian dependency, the verb is at the centre of all dependency relations. It treats a sentence as a series of relations between the verb and the various nominals in a semi-hierarchical order. Panini's rules are preserved in the book 'Aṣṭādhyāyī'. Though it deals with the derivational and inflexional word-formation process, Panini's 'Aṣṭādhyāyī' is not a grammar book in the general sense of the word. Rather, it should better be understood as a rule-governed device to generate linguistic expressions (Joshi, 2009, p. 2; Mishra, 2009, p. 40). There are around 4000 rules in 'Aṣṭādhyāyī'.

The central part of Paninian grammar is its kāraka relations. Karaka is the structural-semantic relation that is assigned to the nominal elements (nouns and pronouns) in *relation to the verb in a sentence. The grammar provides explicit rules about how to recognise and interpret a karaka, usually by the existence of specific vibhakti (i.e., the*

suffixes) attached to the nominals. Paninian Grammar specifies 6 karaka roles i.e., ‘*karta*’, ‘*karma*’, ‘*karaṇa*’, *sampradāna*, *apadāna* and *adhikaraṇa*.

Karta is the logical subject of a sentence; it is the most independent of all karaka. *Karma* is the direct object, *karaṇa* is the instrument of action, *sampradāna* is the recipient/beneficiary of an action. *Apadāna* can be defined as the fixed point of departure, while *adhikaraṇa* refers to location. These six karaka relations described by Panini, to some extent, correspond to the thematic roles. We can map them to theta roles in the following way

- (1) *karta*(k1) ‘agent/theme/experience/actor’
- (2) *karma*(k2) ‘theme/patient’
- (3) *karaṇa* (k3) ‘instrument’
- (4) *sampradāna* (k4) ‘recipient’
- (5) *apadāna* (k5) ‘source’
- (6) *adhikaraṇa* (k7p) ‘location’

These karaka roles are situated on a level that can be described as a coupling point of “meaning and morphosyntactic surface structure” (Kipersky, 2009, p. 16). It should be noted here that karaka roles are not identical to semantic roles. There can be a large number of semantic relations. But the karaka roles are a fixed number of relations that capture only a certain level of semantics. Though we can roughly map the last four karakas to their thematic counterpart, the ‘*karta*’ and ‘*karma*’ are entirely different from agent and theme. These two karakas are roughly equivalent to syntactic criteria like sentential subject and sentential direct object, respectively (Bharati, Chaitanya and Sangal, 1995). This is because thematic roles are purely semantic, whereas karaka relations are syntactic-semantic in nature.

The status of karakas has been a matter of intense academic debate for more than a century which falls outside the scope of this study. It is fairly apparent that the karaka system

is not identical to what is known as ‘case’ in Latin or other classical grammars. Bhavananda (1579 A.D.) defined karaka as something which is syntactically connected with the action verb and is conferred with any one of the following six properties, i.e., agency, objecthood, instrumentality, recipient-hood, ablative-hood and locus-hood (Bhavananda, quoted in Matilal, 1991, P. 278). It proves that the notion of karaka is more functional than theoretical. Through the karaka system, Panini was able to explain the functionality of various grammatical forms which occur within a sentence, usually as an affix, and contrast them with each other. Then, the karaka system establishes those grammatical forms as representations of certain semantic relations. (Matilal, 1991, P. 263-64). We here have similar goals as we attempt to build up a verb-centric dependency framework in this dissertation. Hence, we have adopted the Paninian Karaka Approach within our model. The reasons are as follows:

1. The Paninian dependency framework was originally created for Classical Sanskrit. Bengali, being an NIA language, has inherited many structural features of Sanskrit.
2. Indian languages are inflectionally rich. Information is encoded mostly at the morphological level. The Paninian Karaka Approach works better in these kinds of languages.
3. Paninian Grammar is better suited for flexible word-ordered languages; Bengali fits into this category (Bharati, Chaitanya and Sangal, 1995).

1.5 The Objectives

The goal of this dissertation is to develop and present a bank of Bengali verb-frames representing the most common Bengali verbs. The verb-frames have been developed as a network that is inter-referenced and can be extended to represent all verbs at a later period. In order to accomplish the primary goal and along with it, the following objectives have been identified.

- i) A well-rounded corpus needs to be developed first. The corpus should demonstrate the everyday use of standard colloquial Bengali.
- ii) The Bengali verbs for which verb frames will be created need to be identified. Those verbs then should be analysed for sense disambiguation.
- iii) Two different annotation frameworks, one for dependency roles and another for ontology, have to be established.
- iv) Verb frames will be created and populated. Thereafter, verb frames need to be analysed, and the important observations regarding this resource need to be recorded.
- v) At last, a verb classification is to be attempted based on the similarity of the verb frames.

1.6 The Research Question

We start from the assumption that a verb's meaning correlates with its syntactic behaviour (Levin, 1993). If this is true, then meaning difference should lead to differences in syntactic environment in which a verb occurs. In other words, if a given verb has two distinct senses, we should always get two distinct verb-frames. This is our hypothesis here and the research is motivated by this hypothesis. The opposite position is that there is no correlation between syntactic behaviour of a verb and its meaning. Bengali Verb-frame Network consists of verb-frames representing different verb senses. Our hypothesis will be true if all verb-frames of a given verb are different from each other.

1.7 Problem Statement and Scope

The focus here is to develop a bank of Bengali verb frames. The research initiative is inspired by the gap between the necessity and the availability of digital verb lexicon in Bengali. VerbNet has been developed for English, which represents English verbs' all

possible syntactic alternations in a very consistent way (Kipper Schuler, 2005). Begum (2017b) developed a Hindi verb frame as a part of HyDT (Hyderabad Dependency Treebank). Bengali does not have any such exhaustive resources (Ghosh, 2014, p. 2). Sense disambiguation is one of the cardinal issues in many NLP researches today. Apart from interests in NLP and computational linguistics, the change of the verb’s meaning as it occurs in different contexts is in itself a matter of serious interest from the viewpoint of theoretical and core linguistics. This line of analysis will inevitably relate the lexical ambiguity of verbs with the setting of its argument structure. Here, we segment lexical Bengali verbs into different verb senses. A verb’s meaning is considered a spectrum in which different points manifest different verb senses. In the planned verb-frame repository, each verb-frame will represent a verb sense instead of a lexical verb. This is a distinctive feature of Bengali Verb-frame Network. Thus, the proposed Bengali Verb Frame Network is uniquely positioned to tackle verb-sense disambiguation. There are two separate annotation schemes used in the development of the verb-frame. One is to annotate dependency structure, and another is to tag ontological properties of the verbal arguments. It is intended as a multi-layered, interconnected resource that holds many types of information represented in a table of verb-frame.

1.8 Justification and Likely Benefits

The Bengali Verb Frame Network will be an important resource and will bridge a gap in the existing research. It is a tool that can be used to analyse the grammatical structure of Bengali sentences and identify the different types of arguments associated with Bengali verbs. The justification for developing a Bengali verb-frame is manifold. First, it can help linguists and language researchers gain a deeper understanding of the Bengali language. By analysing the syntactic-semantic structure of Bengali verbs and their arguments, future

researchers can identify patterns and regularities that can shed light on the underlying grammar of the language.

Secondly, a Bengali verb frame can be used to develop natural language processing (NLP) tools for Bengali. The Bengali verb frame will be specifically designed for verb-sense disambiguation. By incorporating this resource, NLP tools can potentially resolve sense-disambiguation as they rely on accurate linguistic analysis of text to perform tasks such as machine translation, sentiment analysis, and text classification. By providing a comprehensive representation of the syntactic structure of Bengali verbs and their arguments, a verb frame can help improve the accuracy and efficiency of NLP tools for Bengali.

Third, a Bengali verb-frame can be used to develop language learning resources for Bengali learners. By providing a clear and structured representation of Bengali verb syntax, learners can easily understand and internalise the rules of the language.

After developing the verb-frames, we will study and analyse the verb-frames in the fifth chapter of this thesis. Classification of the Bengali verbs will be attempted on the basis of the verb-frame information. Such verb classification is uncharted territory in the case of Bengali and will be a separate resource in itself.

1.9. The organisation of the Thesis

This dissertation consists of 6 chapters, including the introductory chapter. Chapter two reviews prior works in the relevant field. This chapter is divided into three main parts. The First part introduces the important theoretical models in verb semantics like Jackendoff's Lexical Conceptual Structure, Fillmore's Semantic Frame and Pustejovsky's Generative Lexicon etc. The second part presents the overview of several well-known verb-based NLP resources like Levin's verb classification, VerbNet, PropBank, FrameNet and Hindi verb

frame. The third section deals with the theoretical background and the development of the concept of ontology in Computational Linguistics.

Chapter three is the methodology chapter. It enumerates the basic assumptions behind our initiative, along with some discussions on the theoretical frameworks. It then details the development process of the verb frame and the basic resources needed for creating such a knowledge base, and introduces four sections of our verb-frame.

Chapter four is titled ‘Architecture of the Verb-frame’. It describes the annotation schemes used in the development of the verb frame. Each verb frame has four sections: dependency relation, lexical type, dependency marker and ontology. There are two annotation schemes used to annotate dependency relation and ontology. Lexical types are tagged using the linguistic nomenclature like ‘N’ for nouns, ‘ADV’ for adverbs etc. The only other category, ‘dependency marker’, represents the post-positions and affixes that signal the dependency relations. There is a section allotted for discussion on the dependency markers in this chapter.

The purpose of chapter five is to present verb-frames and analyse it. After a brief introductory paragraph, basic statistics regarding the Bengali Verb-frame Network are provided here. Thereafter the chapter includes the verb-frame files along with the descriptions. The importance and implicit role of each element in the verb-frame are explored, and certain complicated structures are analysed in the verb-frame approach. Then, the classification of the Bengali verbs is established on the basis of verb-frame. The last section of this chapter documents the observations from the corpus analysis conducted along with the verb frame development.

Chapter Six is the last chapter of the thesis. This chapter is titled ‘Conclusion and Future Research’. It evaluates the contribution of the research. It then includes the concluding remarks and elaborates on our plan to expand the Bengali Verb-frame Network.

CHAPTER 2

BACKGROUND

In this chapter, we will look into the various theoretical models as well as existing NLP resources which analyse verb semantics in one way or another. This chapter is divided into three sections. The first section presents influential theoretical approaches to verb semantics, and the second section discusses the NLP resources, which are based on those theoretical models. In the third section, a brief introduction to ontology is provided.

2.1 Theoretical Approaches

The last paragraph of Chomsky's 'syntactic structures' goes:

More generally, it appears that the notion of "understanding a sentence" must be partially analysed in grammatical terms. To understand a sentence, it is necessary (though not, of course, sufficient) to reconstruct its representation on each level, including the transformational level where the kernel sentences underlying a given sentence can be thought of, in a sense, as the 'elementary content elements' out of which this sentence is constructed. In other words, one result of the formal study of grammatical structure is that a syntactic framework is brought to light, which can support semantic analysis. Description of meaning can profitably refer to this underlying syntactic framework, although systematic semantic considerations are apparently not helpful in determining it in the first place. The notion of "structural meaning" as opposed to "lexical meaning", however, appears to be quite suspect, and it is questionable that the grammatical devices available in language are used consistently enough so that meaning can be assigned to them directly. Nevertheless, we do find many important correlations, quite naturally, between syntactic structure and meaning; or, to put it differently, we find that the grammatical devices are used

quite systematically. These correlations could form part of the subject matter for a more general theory of language concerned with syntax and semantics and their points of connection. (Chomsky, 1957, p 107-108)

The general theory that Chomsky explains here has to solve two major problems. One is the Problem of Meaning, which means that the theory has to “develop a formal treatment of semantic intuitions” (Jackendoff, 1990, p. 1). Another is the Problem of Correspondence. It is a problem that semantic theorists face when they try to grasp the interaction between words’ meaning and syntactic structure. Therefore, the task of a general theory of semantics is not only to account for the native speaker’s knowledge of meaning but also to provide a formal representation of syntax-semantics interaction. The basic problem in developing such a formalism is that it needs to reflect the complexities and difficulties and real-life language usage, not just to deliver a “ready to wear logical formalism” (Pustejovsky, 1995, p.2). Various theoretical frameworks were developed in the last few decades that try to arrive at such a unified theory. Semantic theories and frameworks usually fall along two lines, primitive-based and relation-based. The primitive-based theories assume that the word meaning can be defined in terms of a fixed number of semantic primitives and rely on the semantic decomposition strategies to arrive at those semantic primitives (e.g., Jackendoff, 1972; Katz, 1972; Lakoff, 1971; Schank, 1975; Wilks, 1975). On the other hand, the relation-based theories attempt to explain word meaning in terms of a network of relational ties in the lexicon (e.g., Brachman, 1979; Carnap, 1956; Collins & Quillian, 1969; Fodor, 1975; Quillian, 1968). Some of the theoretical approaches presented here which were very influential in the later developments.

2.1.1 Early Initiatives in Generative Semantics

Chomsky's quote, presented in section 2.1, adequately voices the confusion among generative linguists of the 50s' and 60s' in determining the role of semantics in their theoretical model. Chomsky (1957) acknowledged the role of semantics but tended to disregard structural meaning. Hence, the focus is completely on syntax, and semantics is treated as a lower branch of syntax. In 1964, Katz and Postal's widely influential book 'An Integrated Theory of Linguistic Descriptions' was published and subsequently accepted as the mainstream position of generative semantics. It was an advancement over earlier approaches (e.g., Katz and Fodor, 1963), which merely stated "Synchronic linguistics description minus grammar equals semantics" (Katz & Fodor, 1963, p. 172), where grammar is syntax + phonology. In practice, it just means semantics is anything that one doesn't want to explain through formalism. It should be noted here that this kind of negative connotation with semantics existed prior to the generative tradition, perhaps starting with Bloomfield (Bloomfield, 1933). It was quite understandable because there was a relative scarcity of semantic formalism in comparison to syntactic one. Yet, Katz and Postal did not try to create semantic formalism as an independent domain. Their formalism assumes-

KP Hypothesis (strong form): - Deep structure is the source of all semantic information.¹

This line of argument ensures that the meaning difference between sentences created with identical lexical items has to be explained solely as a difference in deep structure. One thing that Fodor (1970, 1980) did which was consequential for the later developments was to point out that simple semantic feature analysis does not go much beyond a few clearly defined concepts. He noticed that, for all practical purposes, it is impossible to define most

¹ This hypothesis also has a weak form. Which states, "Semantic projection rules operate exclusively on underlying phrase markers; hence transformations do not change meaning." From there, it is a simple rhetorical step to reach the 'strong hypothesis'.

lexical items in a precise way. Hence, he argues that there is no reason to believe that lexical concepts have any internal structure.

To some extent, this is indeed a fact. We can consider the colour terms as examples. How do we make a categorical analysis on terms like red or purple? How do we semantically relate those colour terms to each other? Fodor did not offer any good solution other than denying the compositional structure of meaning. Furthermore, this line of argument erases the motivation to create a separate semantic formalism. It was only in the '80s Chomsky (1981) finally incorporated some level of semantic formalism in the form of thematic roles in his contemporary work, i.e., **Government and Binding** framework. Other semanticists like Jackendoff (1990) argued that even these thematic roles are not semantic representations. These should arguably be understood as some deeper-level structural representations. Jackendoff viewed them as an intermediate level between the structural level and the semantic level. Hence, as a consequence of lesser focus and motivation, most of the detailed semantic formalisms did not arise from generative tradition. Rather, some of the very important later approaches use dependency-based formalisms instead of the constituency-based models of generative linguists.

2.1.2 Jackendoff's Lexical Conceptual Structure

Ray Jackendoff (1990) accepted Fodor's (1980) position that the meaning postulate approach and feature analysis are inadequate to be developed into a complete semantic formalism but rejected his claim that each lexical item is an isolated unit that exists without any compositional structure. He developed his lexical conceptual structure framework that seeks to analyse the meaning through the mapping between linguistic structure and conceptual processes. He based his framework on Gruber's work (Gruber, 1965) but expanded it into a complete system. The concept of conceptual structure forms the

foundation of LCS. According to Jackendoff, conceptual structures are not part of language; rather, they should be understood as part of thought. Conceptual structure is said to be built on some essential units, which are known as **conceptual constituents**. These constituents are **Thing, Event, State, Action, Place, Path, Property, and Amount**, each representing one major ontological category in conceptual structure. These are universal semantic categories.

Consider the sentence below:

(1) Steve went into the room

[Event GO ([Thing STEVE]_i,]Path TO ([Place IN ([Thing ROOM]_m))]_k)]_j

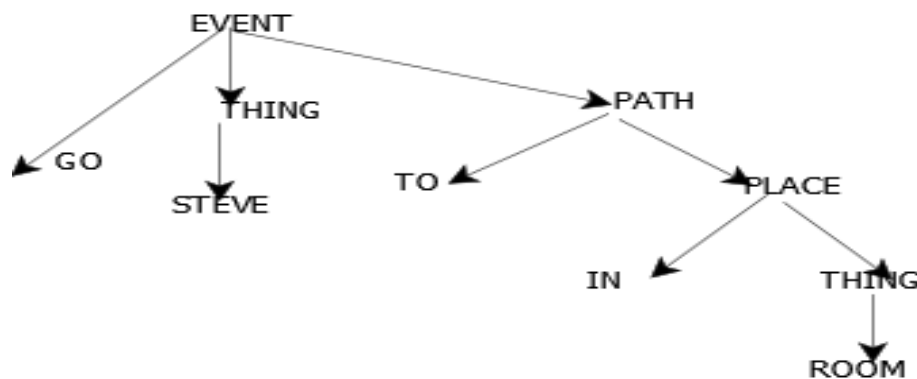


Figure 2.1: Conceptual Structure of sentence (1)

The conceptual structure of sentence (1) is shown above in both label-bracketing as well as in tree diagram format. These conceptual constituents can be mapped onto the grammatical categories. Thing usually refers to the NPs, and Event usually indicates a verb etc. But, the reverse of this situation is not true because NP can be a Thing (The stick), an Event (The war), or a Property (friendliness); a sentence can express an Event (The wheel is spinning) or a State (He is sick) and so on. This mapping is partly language specific. NP can express almost any conceptual category; a PP can express a Path, Place, and in English, it can mean Property; a sentence can represent a State or an Event. The below example which are from Jackendoff (1990) illustrates the correlation between syntactic structure and conceptual structure.

(2) a. *Syntactic structure*

[S[NP John] [VP ran [PP into [NP the room]]]]

b. *Conceptual structure*

[Event GO ([Thing JOHN], [Path TO ([Place IN ([Thing ROOM])])])]

Jackendoff adopted Gruber's localist approach and kept it as an inseparable section of his system. As discussed, the localist approach argues that location and motion are essential for the conceptualisation of all other events. The distribution of these semantic primitives seems to be constrained by some rules. Some of the most important rules are given below.

a. [PLACE] → [_{Place} PLACE-FUNCTION ([THING])]

b. [PATH] → [_{Path} {
TO
FROM
TOWARD
AWAY-FROM
VIA
} ([{THING}
[PLACE]})]

c. [EVENT] → { [_{Event} GO ([THING], [PATH])]
[_{Event} STAY ([THING], [PLACE])]

d. [STATE] → { [_{State} BE ([THING], [PLACE])]
[_{State} ORIENT ([THING], [PATH])]
[_{State} EXT ([THING], [PATH])]

e. [EVENT] → [_{Event} CAUSE ([{THING}
[EVENT]), [EVENT])]

Figure 2.2: Specialised Formation Rules for Conceptual Structures

These rules explain how each conceptual constituent can be elaborated as a function of other conceptual constituents.

Jackendoff (1985) does not consider the thematic roles to be semantic primitives. They are not independent and can be understood as arguments of conceptual constituents. For instance, the thematic role ‘Source’ can be defined as ‘the object from which motion proceeds’; thus becomes an argument of the Path function FROM. LCS presents conceptual roles in a two-tier structure, i.e., thematic tier and action tier. The thematic tier represents motion and location; on the other hand, the action tier deals with the actor-patient relationship.

LCS introduced many new types of machinery and a novel approach to the study of semantics, which has had a profound influence on later semantic research. The stated goal of Jackendoff’s work is to develop a semantic model which is compatible with the formalisms of generative syntax. This formalism is in no way secluded from the generative approach, but it treats semantics as an autonomous organisation. Unlike many other semantic frameworks, LCS formalism is not created just to meet the demand of syntax. The theoretical foundation of this framework follows that semantic theory and formalism must be able to explain the ‘language of thought’ and its transformation into words, phrases and sentences. Jackendoff understood that the task needed a different vocabulary and organisation than what was available. This is the motivation behind LCS.

2.1.3 Fillmore’s Semantic Frame

Charles Fillmore first put forth the concept of the semantic frame as an extension of his earlier case theory in his seminal paper ‘Case for Case Reopened’ (1977a). Later it developed into a complete system of semantic theory rather than an ancillary idea. The frame semantics, as it developed in later years, does not have much in common with his earlier case theory, though, in some of his papers, he used the term ‘case frame’ interchangeably with ‘semantic frame’. Fillmore’s, just like Jackendoff, summarily rejected the contemporary componential analysis or semantic feature analysis. He maintained that meaning cannot be

disintegrated into a list of features. Rather, the human mind looks for a prototype to understand a category. In this respect, he mentioned, in his seminal 1977 paper, the famous colour terms research of Berlin and Key (Berlin & Key, 1969). This research showed that major colours like red are recognised more or less directly. On the other hand, minor colours like pink are understood by their difference from red “along a certain dimension and to a certain degree” (Fillmore, 1977c, pp. 55-56). Hence, semantic frame theory originated from his rejection of componential analysis and his attempt to theorise the perception of meaning.

The basic idea behind the frame semantics is quite accurately presented by the following slogan “Meanings are relativised to scenes” (Fillmore, 1977a, p. 59). The central point is that the meaning of a lexical unit cannot be fully understood unless we are looking at the entire context in which the word occurs in terms of real-world events. Each word, phrase or sentence is a part of larger settings in the real world. Those settings or the real world activities are semantic frames, whose participants are called Frame Elements. He argues that children first learn the entire setting, then the individual words, which are the elements of the setting. Hence, the argument goes, the meaning of an individual lexical item must be understood by its position in the semantic frame it ‘activates’.

Let us take the commercial transaction frame as an example. Its frame elements are the buyer, the seller, money and the commodity. English words like ‘buy’, ‘sell’, ‘cost’, ‘money’, and ‘pay’ can each activate the entire event of a commercial transaction.

- a) John bought a toy from Steve.
- b) Steve sold a toy to John.

These two sentences represent the same real-world event but from different perspectives. Sentence (a) represents the perspective of a buyer, whereas sentence (b) represents the perspective of the seller. We can say sentences (a) and (b) are two ‘scenes’ from the commercial event frame. Fillmore’s central argument is that we cannot understand

the ‘scenes’ as represented in (a) and (b) if we do not have a prototypical commercial event in our minds. For, the concepts like buyer, seller, money, and commodity all are understood only in the context of a commercial transaction event.

The concept of prototype or prototypical frame is the most important tenet of frame semantics. Fillmore adopts the concept of the prototype as proposed by psychologist Eleanor Rosch. Prototypes are certain members, as experimental evidence suggests that are instantly recognised as the better instances of a category than other members. In other words, prototypical elements are representatives of a category. Fillmore employed the same basic idea in his frame theory. The prototypical frame consists of the essential entities of participants of the frame. For example, in the case of commercial transactions, the buyer, the seller, money and the commodity are the essential elements. One cannot think of a commercial transaction taking place without these four elements. Hence, a prototypical commercial transaction event consists of these above-mentioned elements, which are obligatory for its execution. Optional elements such as the time of transaction can be added to the frame.

Therefore, the semantic frame is a prototypical frame, a prototype of real-world events. Frame can be described as an abstraction from the numerous experiences of similar events, similar enough to form a category in the human mind. The real-world experiences are called ‘scenes’ in Fillmorean terms. Scenes refer to the experiential, mental reality, whereas frame is related to linguistic choices. Both scenes and frames can activate each other. In his seminal paper, Fillmore always talked about ‘scenes’ in plural form. Clearly, he was referring to all the relevant experiences combined, which would then be generalised into a frame.

The semantic frame approach, for the most part, was initiated in order to provide a better alternative to the structuralist approach, which focuses on the segmentation of meaning. Fillmore’s criticism of compositional analysis is well grounded. Relevant research

indeed shows that, in the process of understanding a meaning, cognitive processes do not break it into more basic segments. Rather, meaning is understood as a whole unit. Fillmore himself talked about the potential application of his model in the domain of text analysis, metaphor analysis, ambiguity resolution etc. The semantic frame approach was later realised into ‘FrameNet Project’, which is a great NLP resource for complex computational purposes.

2.1.4 Generative Lexicon

Lexicon is usually perceived as a static organisation. Even generative linguists tend to treat the lexicon as a default basket where all the idiosyncrasies are placed. Meaning is one such idiosyncratic property of a word; therefore, it belongs to the lexicon. The traditional organisation of the lexicon ascribes an enumerable number of senses to each word in a fixed, static format. Hence, it fails to account for the creative use of a word in a novel context. For computational purposes, the system has the “need to specify, ahead of time, the contexts in which a word might appear; failure to do so results in incomplete coverage” (Pustejovsky & Boguraev, 1993, p. 194). James Pustejovsky first introduced and outlined the theory and framework of Generative lexicon in his 1991 paper ‘The Generative Lexicon’. Later it was elaborated by Pustejovsky & Boguraev (1993) and Busa (1996). Pustejovsky argued that we could deal with lexical ambiguity in a better way if we regard the lexicon as a generative device. The framework of lexical knowledge he suggests is dynamic, open-ended by definition and organised in a multi-layered structure of generalisation. This model proposes four types of different knowledge representation structures to successfully capture lexical meaning in a layered organisation. These are 1) **Argument Structure**, 2) **Event Structure**, 3) **Qualia Structure** 4) **Lexical Inheritance Structure**. These four structures are interconnected through a set of generative devices.

Argument Structure: - According to Pustejovsky and other theorists of this framework, the starting point of semantic mapping should be argument structure which refers to the

number and specification of logical arguments of a word as well as the description of its syntactic realisation. Argument structure in generative lexicon framework is not different from what is established in generative tradition. Pustejovsky calls argument structure a ‘minimum specification’. It represents the mapping of logical arguments into syntactic realisation. It is a necessary part, but it cannot complete the semantic information of lexical units by itself.

Event Structure: - Events or aspectual structure of verbal semantics has become a standard component of linguistic theories, especially those dealing with verbal semantics (Vendler, 1967; Dowty, 1979). Event-based interpretation of lexical units is another layer of multi-layered organisation that constitutes generative lexicon. At the most basic level, there are three classes of events, namely **State, Process and Transition**. Pustejovsky, although mostly following earlier noted works, also developed a more nuanced and elaborative sub-eventual structure. In his development of event structure, he was clearly influenced by Jackendoff’s LCS, as he himself acknowledged (Pustejovsky, 1991b).

Qualia Structure: - Qualia structure is what defines “essential attributes of objects, events, and relations, associated with a lexical item” (Pustejovsky & Boguraev, 1993, p. 203). It deals with the semantics of nominals just like argument structure does with verbs. In this way, instead of positing nominal construction as a passive argument, GL treats nouns on an equal footing with respect to verbs. Qualia Structure identifies four roles of a nominal construction, i.e., **Constitutive role, Formal role, Telic role, and Agentive role**. These are four different aspects of a word’s meaning that together compose the semantics of a word in relation to the other lexical items in a sentence or phrase.

The constitutive role describes the physical composition and basic physical attributes like weight etc., of an object.

Formal roles are those properties of an object which distinguish it in the larger domain.

Pustejovsky listed the following elements under formal role (Pustejovsky, 1991a).

- a. Orientation
- b. Magnitude
- c. Shape
- d. Colour
- e. Dimensionality
- f. Position

Telic role is the purpose and function of an object in terms of the real world.

The agentive role is the combined factors that are involved in the origin or the ‘bringing about’ of an object.

These roles, combined into one, select or license the correct sense of a word in a particular sentence. Hence, the dynamics of these roles have both semantic and grammatical consequences. Here is an example.

“novel(*x*)

Const: narrative(*x*)

Form: book(*x*), disk(*x*)

Telic: read(T,y,*x*)

Agentive: artifact(*x*), write(T,z,*x*)” (Pustejovsky, 1991a)

The above example shows the Qualia structure of the lexical item ‘novel’. It is a narrative which is available either in the form of a book or a disk and is created for the purpose of reading. ‘y’ symbolises the reader. It is an artefact that came into existence through the act of writing, whose event type is ‘transition’. The writer is symbolised by ‘w’. Thus, the above systematic description perfectly unveils the innate world knowledge of the lexical item ‘novel’.

Qualia structure presupposes that a lexical item has many aspects of meaning or semantic substructure within it. A book is both ‘literature’ as well as ‘physical object’ and anything else as the context demands. Researchers face this problem whenever they try to do ontological annotations of lexical items. Different aspects of semantics get foregrounded depending upon the context requirements. Thus, it attempts to resolve ambiguity that arises out of confusion about the class membership of a lexical item.

Lexical Inheritance Structure: - Lexical inheritance structure primarily means a set of relational mechanisms among lexical items. NLP research and traditional semantics have always been familiar with a static inheritance structure, i.e., a fixed set of relational concepts like hyponyms and hypernyms. Apart from the mentioned fixed lexical inheritance structure, The Generative Lexicon proposes another kind of inheritance which Pustejovsky calls ‘projective inheritance’. This inheritance link essentially comes from the qualia structure. For example, two concepts like ‘prisoner’ and ‘escape’ seem to share something in common because whenever we think of escape, we are bound to imagine a prisoner. Just like the event structure, this section of the theory is heavily loaded with higher-level mathematics. Pustejovsky finds a way to connect different lexical concepts within a network of inheritance structures.

The reason Pustejovsky stresses the need for a generative lexicon is that the traditional lexicons cannot adequately pinpoint accurate sense in a given condition. It tends to give a generalised meaning, a sort of median position among all the available senses. A lexical item gets a specific sense from the context itself in which it is used. Traditional lexicography does not possess any mechanism to capture the network a word establishes with the context of its occurrence. It may not be a big difference for the human user. But the computational lexicon needs to be structured in a way so that machines can understand the context-specific sense. Traditional lexicons, obviously, cannot account for the creative, novel use of a lexical item.

Pustejovsky and Boguraev established that “There are serious problems with positing a fixed number of “bounded” word senses for lexical items” (Pustejovsky & Boguraev, 1993, p.197) because “as far as a text analysis system is concerned, this reduces to a search within a finite space of possibilities” (Pustejovsky & Boguraev, 1993, p.197).

They have shown that their analysis is explanatorily adequate to account for many known types of lexical ambiguity. It provides nearly complete information regarding the contextual sense of a lexical unit. Generative lexicon argues against the view that word meaning is fixed and ambiguity can be explained in terms of multiple word entries in the lexicon. This framework is one of the newest important contributions in applied as well as theoretical semantics as it not only shows the weakness of traditional lexicography but also details an alternative mechanism.

2.1.5 Verb in the Indian Grammatical Tradition

Indian grammatical tradition refers to the rich and diverse body of linguistic scholarship that has developed in India over thousands of years. Panini is considered to be the father of the Indian tradition of linguistics, and his *Aṣṭādhyāyī* is the foundation of it. Most of the later works by other ancient Indian grammarians are commentaries on *Aṣṭādhyāyī*. Following Panini, several other grammarians further developed and expanded upon his ideas. Notable among them are Katyayana, Patanjali, and Bhartrhari. Katyayana’s *Varttika* is a commentary on Panini’s *Ashtadhyayi*, providing further explanations and clarifications. Patanjali’s *Mahabhashya* is another influential commentary on Panini’s work, offering interpretations and discussions of the Paninian sutras. Bhartrhari’s *Vakyapadiya* explores the philosophy of language and the relationship between language, thought, and reality.

Paninian Grammar, as outlined in Panini’s *Ashtadhyayi*, is divided into four main parts. Each part focuses on specific aspects of linguistic analysis and provides a systematic

framework for understanding and generating Sanskrit sentences. The four sections of Paninian Grammar are:

1. *Śivasūtras*: The *Śivasūtras* serve as the foundation of Paninian Grammar. It introduces various linguistic elements, such as vowels, consonants, and the rules of sound changes, which are essential to understand the formation and transformation of words in Sanskrit.
2. *Ashtadhyayi*: The *Ashtadhyayi* is the core text of Paninian Grammar, consisting of eight chapters (*Adhyayas*) with around 4,000 sutras (rules). It is a comprehensive treatise that covers the morphology, syntax, and semantics of the Sanskrit language. The *Ashtadhyayi* provides a detailed analysis of word formation, inflectional processes, and syntactic rules. It defines various grammatical categories, such as nouns, verbs, particles, and compounds, and establishes rules for their combination and modification.
3. *Ganapatha*: The *Ganapatha* is a part of Paninian Grammar that focuses on the formation of verbal roots and their verbal derivatives. It provides a systematic classification of verbal roots into various groups called *Ganas*. The *Ganapatha* classifies roots based on their phonetic characteristics and identifies the rules for forming different verb forms, such as tense, voice, and mood. It establishes patterns and conventions for inflecting verbs and generating a wide range of verbal forms.
4. *Dhatupatha*: The *Dhatupatha* is a critical part of Paninian Grammar that deals with the roots (*Dhatu*) of Sanskrit words. It is a list of 2000 verbal roots along with the subclassification information and diacritic markers to denote the morpho-syntactic properties.

These four parts of Paninian Grammar work together to form a comprehensive system for analysing and generating Sanskrit sentences. The *Shiva Sutras* establish the phonetic

principles, the *Ashtadhyayi* provides the rules of morphology and syntax, the *Ganapatha* focuses on verb forms, and the *Dhatupatha* deals with the roots of words. Grammar formalism analyses a sentence hierarchically at four levels of the linguistic domain, starting from semantics and ending at the phonetic realisation. The derivation of the sentence is explained in the figure below.

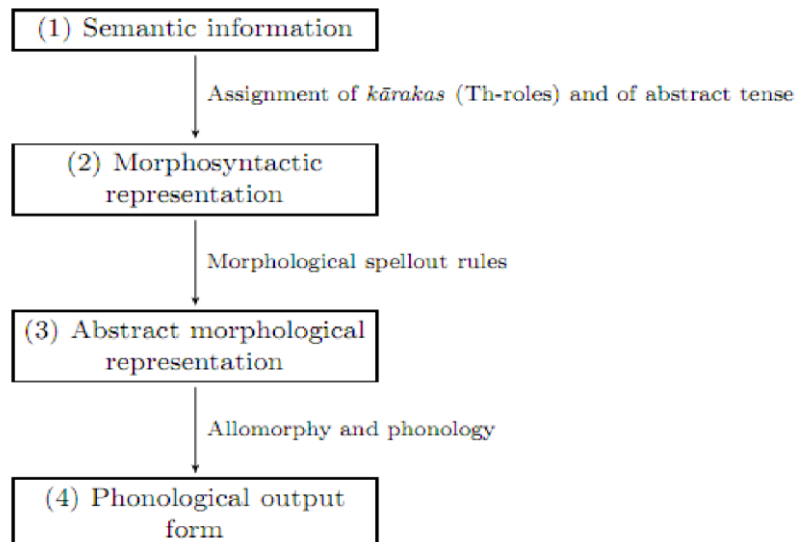


Figure 2.3: Derivation of Sentence in Paninian Grammar

The Paninian grammar conceptualises a sentence as a chain of dependency relations between verbs and various nominals. These dependency relations are called ‘karaka’, and they are assigned at the first level, which is the level of semantic information. There are six karaka roles; they are karta, karma, karana, sampradana, apadana and adhikarana (see Chapter 1, section 1.5). Panini also talks about the genitive relation between two nominals; he calls it *śeṣa* ‘the remainder’. The assignment of ‘karaka’ is crucial for the other following steps because morpho-phonological realisation is dependent upon it. The grammar recognises the verb as the primary constituent that imparts action or state and is responsible for the overall organisation of a sentence.

The emphasis on semantic roles and the organisation of these roles around the verb sets the Paninian Karaka approach apart. Instead of relying solely on syntactic relations like subject and object, the Paninian approach focuses on the conceptual roles played by participants in relation to the action denoted by the verb. This allows for a more nuanced and comprehensive analysis of sentence structure and meaning. Overall, the Paninian verb-centric approach has played a foundational role in shaping formal linguistics. Its influence can be seen in dependency grammar, semantic role theory, computational linguistics, and language typology, providing valuable insights into the analysis and understanding of sentence structure, meaning, and language processing.

2.2 NLP Resources

In this section, some of the NLP resources that have been developed and widely used for various computational purposes will be presented. Many of these have been developed from one or the other theoretical frameworks discussed in the earlier section. There is a large number of NLP resources of various kinds that have developed in many languages, including Indian languages. The purpose of this section is to present some widely used NLP resources dealing with verb semantics which are also considered standard in the respective field. These NLP resources are similar to Bengali Verb-frame Networks. In the course of the development of BVFN, we have studied each of these NLP projects extensively, and it has enriched our understanding of the requirements of the NLP systems as well as problems that need to be addressed in the process of development of the same.

2.2.1 Levin’s Verb Classification

Beth Levin’s work (1993) is the most exhaustive attempt to investigate English verbs and systematically document the correlation between the syntactic and semantic behaviour of English verbs. She started with the hypothesis that a verb’s meaning determines its syntactic

behaviour. Levin classified 3100 English verbs into 47 broad classes and further subdivided those into 193 second and third-level groupings. She also displays a wide range of syntactic alternations that those verbs take. Verbs that belong to the same verb class show similar syntactic behaviour, hence have identical verb frames. A verb frame represents the syntactic setting in which a verb occurs; Levin (1993) used it to establish verb classes in English. The underlying idea is that semantically similar verbs yield similar verb frames. Her book is the first thorough attempt to classify English verbs on the basis of syntactic explanation. She published a book detailing her work entitled 'English Verb Classes and Alternations' (1993), in which she examined and exemplified most noted syntactic alternations typical of English verbs like causative-inchoative alternation, unexpressed object alternation, locative alternation and many others.

Many of the alternations have been noticed by various scholars. In the first part, those alternations are examined and exemplified. In the second part, she attempted to classify English verbs on the basis of their participation or non-participation in those alternations. For example, '*Break Verbs*' occurs in section 45.1; and it is characterised by causative/inchoative alternation, middle alternation as well as instrument subject alternation. Levin first made the upper-level classification on the basis of semantic similarity. Then she sub-classified on the basis of the alternations discussed above. In this way, she proves her hypothesis that semantics, indeed, is a good indicator of understanding the syntactic behaviour of a given verb. The relationship between verbs' meaning and syntactic usage may seem arbitrary at first glance, but once we look at subtle syntactic alternations, we discern the control of semantics over syntactic operations. Having detailed so many very complex and subtle syntactic alternations involving English verbs, Levin has done pioneering work. It signifies a leap in the research of verbal semantics. Later projects, like VerbNet, are completely based on Levin's verb classes.

2.2.2 VerbNet project

VerbNet (Kipper-Schuler, 2005) is currently the largest online verb lexicon available for English. This project is administered by the University of Colorado. VerbNet project has employed the model of Levin's verb classes and then extended this classification through refinement and the addition of Subclasses. The verb net has 274 first-level verb classes, 23 thematic roles, and 94 semantic predicates. 5257 verb senses were attested in the net verb project. VerbNet presents a detailed verb-frame for each verb class. The verb-frame contains thematic roles, argument structure, selectional restriction (ontological information of the arguments) and event-based interpretation following Moens and Steedman (1988). Semantic information, in VerbNet, is "expressed as a conjunction of Boolean semantic predicates such as 'motion,' 'contact,' or 'cause.' Each semantic predicate is associated with an event variable E that allows predicates to specify when in the event the predicate is true ($start(E)$ for preparatory stage, $during(E)$ for the culmination stage, and $end(E)$ for the consequent stage)."

Class <i>Hit-18.1</i>			
Roles and Restrictions: Agent[+int_control] Patient[+concrete] Instrument[+concrete]			
Members: bang, bash, hit, kick, ...			
Frames:			
Name	Example	Syntax	Semantics

Class <i>Hit-18.1</i>			
Basic Transitive	Paula hit the ball	Agent V Patient	cause(Agent, E) manner(during(E), directed motion, Agent) !contact(during(E), Agent, Patient) manner(end(E), forceful, Agent) contact(end(E), Agent, Patient)

Table 2.1: Verb Frame for class ‘Hit’ in VerbNet project

VerbNet is compatible with Word-Net, FrameNet and other NLP resources, and it is being used in many NLP applications. VerbNet has been mapped to Propbank predicates as well as FrameNet frames. All these resources are complementary to each other, thus creating a super network of lexical items, senses and syntactic information.

2.2.3 FrameNet Project

FrameNet is a project housed at the International Computer Science Institute in Berkeley, California. The theoretical framework of the FrameNet project is drawn from Charles Fillmore’s frame semantics, which is elaborated on in section 2.1.3. The basic idea is that the meaning of a word can only be understood on the basis of a semantic frame. A semantic frame can be thought of as a conceptual structure describing an event, relation, or object and the participants in it. For example, the word ‘buy’ evokes the semantic frame of a commercial transaction which involves a buyer, a seller, the object which is bought and the price. The verb ‘buy’ obligatorily requires a buyer, goods and optionally a seller and a price. FrameNet website informs us that its lexical database contains 1,200 semantic frames, 13,000 lexical

units and 202,000 example sentences. Lexical Unit (LU) refers to the pairing of a word with its meaning.

FrameNet is a complete semantic database of the predicates in English. The purpose of the FrameNet is to group predicates into a shared semantic network which Filmore defines as a ‘frame’. Unlike Levin’s verb alternation and VerbNet, FrameNet does not take into account the syntactic context for its classification. This NLP resource starts with a brief description of a concept, then segments it into different ‘frame elements’- all of which themselves are concepts, it differentiates between core frame elements and non-core frame elements, and then it establishes some frame-frame relations. It is only at the end the lexical units which signify the frame are introduced. For example, the FrameNet entry for ‘Revenge Frame’ starts with the description of the concept. Then it identifies all the core and non-core frame elements. The core frame elements are ‘Avenger’, ‘Injured party’, ‘Injury’, ‘Offender’ and ‘Punishment’. After that, it describes frame-frame relations under an inheritance network. In the end, it introduces Lexical Units like “avenge.v, avenger.n, get back (at).v, get even.v, payback.n, retaliate.v, retaliation.n, retribution.n, retributive.a, retributory.a, revenge.n, revenge.v, revengeful.a, revenger.n, sanction.n, vengeance.n, vengeful.a, vindictive.a” etc. FrameNet is quite unique in pursuing the abstract notion of ‘frame’, whereas the other resources described here mostly analyse verb semantics.

2.2.4 PropBank

‘The Proposition Bank’ is another useful tool for semantic role labelling and annotation. It consists of two sections

1. Frame files
2. An annotated corpus of the Wall Street Journal portion of the Penn Treebank II.

Propbank mostly remains close to the syntactic level as sense difference is taken into account only if there is corresponding syntactic alternation occurs. The stated purpose of the PropBank is to create a highly efficient annotated corpus to train data for machine learning purposes. It represents the argument structure of the verb without making much distinction between what is known as complement and adjunct in generative tradition. Here is an example,

Frameset accept.01 “take willingly”

Arg0: Acceptor

Arg1: Thing accepted

Arg2: Accepted-from

Arg3: Attribute

Ex: [Arg0 He] [ArgM-MOD would][ArgM-NEG n’t] accept [Arg1 anything of value] [Arg2 from those he was writing about]. (wsj 0186)

(Palmer, Kingsbury and Gildea, 2005, p 5)

The arguments are tagged using a novel semantic labelling method quite different from the traditional way of marking verbal arguments. Though primarily semantic role structures were defined only for verbal propositions, later, attempts were made to include nouns, adjectives and complex predicates (Bonial et al., 2014). ‘Arg0’ represents all the proto-agent features (Dowty, 1991). PropBank is nearer to FrameNet, especially in one aspect. Instead of marking arguments with traditional thematic roles like agent, patient, theme etc., it delves deeper into the event and marks arguments for their actual role in the context. Unlike VerbNet, PropBank does not classify verbs in terms of syntactic alternations and semantic commonality. But VerbNet class membership is employed to project framing for a given PropBank entry.

2.2.5 Hindi Dependency Treebank (HDTB)

This is an ongoing project initiated by IIIT Hyderabad to develop a multi-layered as well as multi-representational Hindi/Urdu treebank consisting of syntactic, phrase level and word-level structures. Various universities and institutions are involved in this project. This project is a successor of ‘Hyderabad Dependency Treebank for Hindi’ in which a pilot Hindi Treebank was developed and presented at ICON-2009 (International Conference on Natural Language Processing-2009). It is a dependency-based Treebank, as the name suggests. A novel dependency annotation scheme AnnCorra was also developed for the aforementioned project. This dependency annotation scheme is based on Paninian Computational Grammar (Bharati, Chaitanya, and Sangal, 1995). It is now widely used for annotation purposes. AnnCorra is detailed in the next chapter.

Along with the development of the Hindi Treebank, Begum et al. (2008, 2017a) also created a lexicon of Hindi verb frames. Different verb senses are identified and treated as different verbs in this database. Verb Frame represents verbal argument structure which is tagged using AnnCorra tag sets. As per the creators, a total of 486 verb frames were created for 486 different verb senses belonging to 300 different verbs. Another 201 verb senses were found to share the identical verb frames created in the verb-frame lexicon. Hence, the verb-frame lexicon represents verb-frame for a total of 687 verb senses (Begum, 2017b, pp -117). This is a remarkable development of NLP research in Indian languages, and subsequently, this resource is used for many NLP tasks.

2.3. Ontology

The term ontology has different connotations in different domains of knowledge. The term has its origin in philosophy, where it means the study of being in general. From the knowledge representation context, Sowa defines ontology as “a catalogue of the types of things that are assumed to exist in a domain of interest D from the perspective of a person

who uses a language L for the purpose of talking about D” (Sowa, 2010). Similarly, in linguistics, ontology is usually thought of as an interconnected network of concepts”. Hence, it is quite clear that ontology is regarded as a way of representing human conceptualisation in all these domains. In artificial intelligence, Gruber defined ontology as an “explicit specification of a conceptualisation” (Gruber, 1993, p. 908), where conceptualisation “is an abstract, simplified view of the world that we wish to represent for some purpose”. In other words, from a machine learning point of view, ontology is usually a formally structured schema that organises real-world objects and abstractions in a hierarchical order. Here, in this dissertation, the term ontology is used in the sense of a technology resource, in contrast to that of a philosophical concept. Ontologies are usually of two types. The first one is called **domain-specific ontology**. It represents the hierarchy of concepts in a particular area of knowledge, like physics or biology. The second is known as **upper ontology** or universal ontology, or top ontology. These are general-purpose ontologies seeking to represent a vast area of knowledge.

The history of ontology in information science can be traced back to the 1970s. AI researchers have become increasingly aware of the need to employ a formal representational structure for machine learning. Some scholars held the idea that the application of ontology in information science is applied philosophy. This, in turn, created interest in studying philosophy. Thomas Gruber’s seminal paper titled “Toward Principles for the Design of Ontologies Used for Knowledge Sharing” came in 1993, in which he used ontology as a technical terminology closely related to semantic networks.

For Gruber, ontology is both a specification mechanism as well as a theory of such specification. He defined ontology for AI systems as a formal specification of elements. Gruber proposed five distinct criteria to design such a formal ontology: clarity, coherence, extendibility, minimal encoding bias, and minimal ontological commitment.

1. Clarity: - An ontology should clarify each section of its modelling. Each term in the ontological schemata should be objectively defined.
2. Coherence: - An ontology should be internally consistent.
3. Extendibility: - The foundational concepts of an ontology should allow extension and modification needed for new tasks. The developers of an ontology should anticipate this and design accordingly.
4. Minimal Encoding Bias: - Encoding bias is the consequence when ontological modelling is developed “purely for the convenience of notation or implementation” (Gruber, 1993, p.910). A good ontology is supposed to reduce this bias to the least degree.
5. Minimal Ontological Commitment: - It means an ontology should make as few claims as possible regarding the world.

These criteria are not complementary in nature. In practice, many of these criteria oppose each other, and an ontology developer needs to make a compromise among them. For the sake of clarity, we need a strong theory so that all concepts can be adequately and objectively defined. On the other hand, Minimal Ontological Commitment demands the weakest possible theory to be adhered to. Hence, Gruber suggested a good ontology is always a trade-off among those criteria. Gruber’s observation has become instrumental in the later development of the ontology. Many formal ontologies have been developed in the last three decades, both upper and domain-specific. All NLP tools now use some kind of ontological model.

CHAPTER 3

METHODOLOGY

When a verb has multiple possible senses, the system has to disambiguate among them and identify its correct meaning as per the context. The basic assumption is that a verb's syntax is determined by its meaning (Levin, 1993). Hence, we expect to get a different verb-frame for each senses of a particular verb. The objective here is to create a repository of verb-frames that will capture fine-grained, context-sensitive information associated with each sense of a verb. Since the focus is on verb sense-disambiguation, primarily, verbs with high lexical ambiguity have been taken and analysed. Still, considerable effort is directed to represent all unique argument structures involving Bengali verbs in BVFN for completeness and coherence.

3.1 Lexical Ambiguity and Verb-frame

There are two major ways ambiguity occurs in natural languages: lexical and structural ambiguity (Ovu, 2011). Lexical ambiguity is a situation where a lexical item has more than one possible meaning. On the other hand, structural ambiguity refers to the multiple possible interpretations of a syntactic structure. Bengali Verb-frame Network is designed to be used to resolve lexical ambiguity in verbs. This section details the approaches and assumptions to the verb sense disambiguation using verb-frame.

3.1.1 Homonymy and Polysemy

There are two major types of lexical ambiguity: homonymy and polysemy. Homonymy occurs when two or more words have the same pronunciation or spelling but different meanings. Among Bengali verbs, there are some well-known instances of homonymy, such as the verb *paRaa* 'to read' and the verb *paRaa* 'to fall' are indeed two different verbs but spelt and pronounced alike. Polysemy, on the other hand, refers to the

multiple related meanings of a single word. Bengali Verb-frame Network deals with 89 most common, frequent verbs, which are analysed for lexical ambiguity and a total of 300 verb senses was identified. Most verb senses in BVFN are instances of polysemy. We do not differentiate between homonymy and polysemy in the verb-frame network. Each verb sense is treated as a new verb, and a verb-frame file is created for every verb sense.

3.2 Event Semantics

The event structure of a verb and the verb-frame interact at the syntax-lexicon interface because it is where verbs get mapped onto the verb-frame. Van Hout (2000) has shown that different verb frames are expressions of different event structures. This proposal solves many implicit problems. One of them is the problem of language acquisition. If the complexity of the verb-frame is nothing but the different combination of events and sub-events, then it follows from there that children acquire this knowledge as they try to express such combinations. However, another related question arises- what are the possible combinations that a single predicate can refer to? The only constraint on the event structure seems to be that the event, expressed by the verb, along with its sub-events, must form a coherent ‘semantic frame’ (Goldberg, 2010). Thus, we arrive at a coupling point of verbal arguments and event structure. Verb frame, in the context of this study, is understood as the multi-layered representation of argument structure, which, in turn, is the realisation of event structure at the deep level.

3.2.1 Event Structure and Verb Sense

Event structure refers to the level of semantics concerned with real-world events as expressed by the language. The event structure of a sentence, in its entirety, tells us which real-world activity or state the sentence has expressed. Since, in natural languages, verbs represent

the activities and states, understanding event structure is very useful and handy for verb sense disambiguation.

Event structure formalism can be used for sense disambiguation by representing the meaning of a word or phrase in terms of its event structure. It can distinguish between different senses of a word with different event structures, even if they share some surface-level similarities. For example, consider the sentences below in which the verb ‘fall’ is used in two different contexts.

1. The leaves fell from the tree.
2. Sales fell by 20%.

In sentence 1, the verb ‘fall’ means ‘to drop from a higher to a lower place’, while in the second sentence, it refers to an event of ‘declining or decreasing in value or quantity’. These two different senses have two different event structures, with the former involving a physical action and the latter involving a change in quantity or value. By representing the event structure of each sense of the verb ‘fall’, it is possible to distinguish the two meanings and disambiguate them in context. For example, the event structure of the physical sense of ‘fall’ might be represented as an activity with an endpoint, while the event structure of the quantitative sense might be represented as a state that changes over time.

Overall, event structure formalism provides a powerful tool for sense disambiguation, enabling linguists and computational systems to represent the meaning of words and phrases in a detailed and nuanced way. Thus, using event structure to analyse the underlying structure of language, we can gain insight into the complex interplay between language, thought, and perception and develop effective methods to understand and process natural languages. For the development of the Bengali Verb Frame Network, the understanding of event structure was instrumental in the identification of distinct verb senses.

Although no specific event structure framework was used, the event structure analysis has been a crucial part of the analytical process.

3.2.2. The Approach to Resolve Lexical Ambiguity by Using Verb-frame

Lexical ambiguity is a phenomenon that occurs when a single word or phrase has multiple meanings or interpretations in different contexts. One way to potentially resolve lexical ambiguity is through the use of verb-frames. As it has been described above, a verb-frame representation is nothing but an event structure invoked by the verb. Therefore, verb-frames for different senses of a verb will be different. Thus, the verb-frame provides a structure that helps to disambiguate the meaning of the verb by providing clues about its intended interpretation. Developing verb-frames for each meaning of a particular verb is a feasible approach to resolving lexical ambiguity. This approach involves creating a separate verb-frame for each distinct meaning of a verb based on its usage in different contexts. It requires a significant amount of linguistic knowledge and analysis, as it involves identifying the different ways in which a verb can be used in natural language. This can be done using techniques such as corpus analysis, in which large collections of text are analysed to identify patterns of language use, and semantic analysis, in which the meaning of words and phrases is analysed based on their relationships to other words and concepts. The tools of dependency grammar can help us in this regard. For the point of dependency, grammar is exactly this kind of analysis.

3.3 Computational Paninian Grammar

Computational Paninian Grammar (CPG) is a framework for computational linguistics inspired by Paninian grammar. It provides a deep level of syntactic-semantic interpretation of a sentence and has been successfully applied to several tasks. The framework has been applied to modern Indian languages, using which the '*Anusaaraka*'

machine translation system has been built. It uses a constraint-based parser which requires karaka frames and transformation rules (Bharati & Sangal, 2010). The focus here is not parser or any other computation tool. But the same principle is followed in the development of BVFN. The analysis of a sentence starts from the main verb. The verb is related to the other elements in the sentence through karaka relations (or other dependency relations). The karaka is realised by the existence of bhakti/post-position. The karaka-vibhakti/post-position correspondence is only a part of the analysis. The lexical or phrasal type of the argument and the selectional restriction is also explored and presented in BVFN.

Dependency roles in BVFN are marked by using a dependency annotation scheme named AnnCorra (Bharati et al., 2006). It was a tagging scheme developed as a part of CPG. Based on the Paninian dependency framework, the AnnCorra is built to facilitate NLP researchers to annotate sentences at many levels. AnnCorra consists of three different tag sets 1. PoS tags 2. Chunk tags 3. Dependency tags. In Bengali Verb-frame Network, the AnnCorra dependency tags are used to annotate the verbal arguments. AnnCorra has 37 dependency tags which are completely modelled on the Paninian karaka approach. AnnCorra tagging scheme has been followed for dependency relations in the development of verb-frames. AnnCorra has a vast repository of dependency roles, but only some are designed to tag verbal arguments. Other dependency relations are used in another part of a sentence; thus, they are not useful for the purpose of developing verb-frames. Furthermore, the Bengali language has some unique constructions which require special treatment. Therefore, the dependency format employed here can be called a modified version of AnnCorra. Chapter 4 includes a complete description of the dependency roles along with definitions and examples.

AnnCorra Dependency Scheme

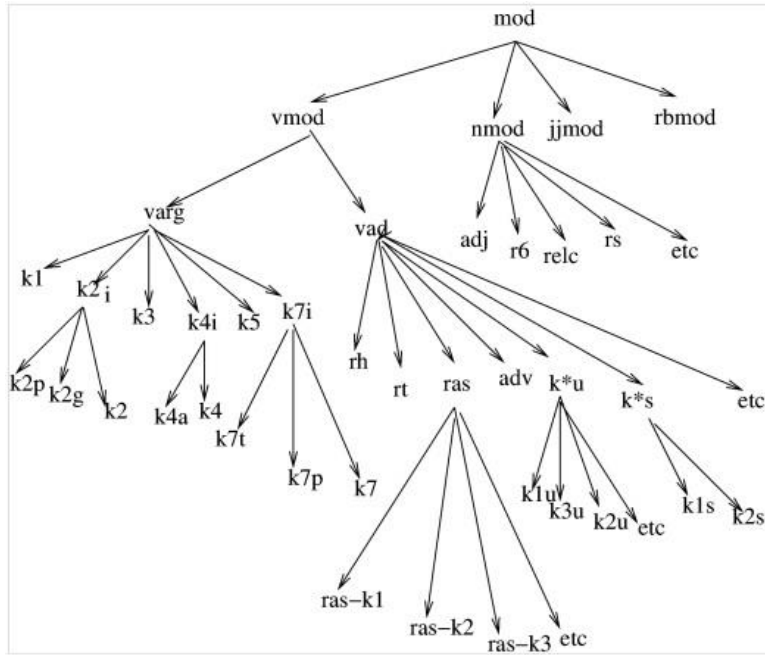


Figure 3.1: AnnCOorra Dependency Scheme

(Adopted from Begum, 2017b)

The AnnCorra dependency roles are of two type's 1. Karaka relations 2. Non-karaka relations.

1. Karaka Relations: -AnnCorra Karaka relations consist of all six paninian karakas along with numerous sub-classes of them. Karaka relations in AnnCorra start with 'K' and are then followed by a numerical. For instance,

karta=k1

karma=k2

Karma has sub-classes: k2p and k2s.

karaṇa=k3

Sampradāna=k4

apadāna=k6

Adhikarana has three subclasses. The classification of adhikarana is captured in the following way:

- (i) kālādhikaraṇa ('location in time') (k7t)
- (ii) deshādhikaraṇa ('location in space') (k7p)
- (iii) viṣayādhikaraṇa ('location other than place and time) (k7)

2. Non-karaka Relations: -

AnnCorra has a number of non-karaka dependency relations; for example, relations such as direction (rd), reason (rt) etc.

3.4 Motivation

The inter-referenced verb-frames developed under the CPG framework play an important role in automatic dependency annotation as well as in parsing (Bharati et al., 2009; Begum, 2017b). The motivation for developing verb-frames are as follows

- i) The verb-frames serve as a detailed lexicon for the argument structure of Bengali verbs.
- ii) It can directly aid verb sense disambiguation
- iii) It can be used to prepare a demand chart for a Bengali parser.
- iv) Verbs have been classified on the basis of verb-frames.

3.5 Modus Operandi

There are six steps in the development of the Bengali Verb-frame Network.

Step 1: Selection of Verbs: -

For the selection of ambiguous verbs, the monolingual dictionary ‘Adhunik Bangla Obhidhan: Bangla Academy’ has been used. To some extent, this dictionary is also consulted for sense differentiation. Eighty-nine verbs have been selected for this study. All the verbs are very common and have frequent usage in the Bengali language.

Step 2: Sense Identification and Differentiation: -

Sense differentiation is the next step. Different senses represent different event structures invoked by the same verb in different contexts (see Chapter 1, section 1.3.1). Primarily, we have used Indo-WordNet to identify different senses of a verb. Of the 89 Bengali verbs, we have identified 300 different senses. Each verb sense is given a SENSE_ID.

Step 3: Corpus Development: -

Corpus analysis is needed to get the real-time usage of the verb. We have used the TDIL corpus, which consists of 1270 text files. These are exclusively academic texts. Apart from that, we have also included a good number of sentences taken from film dialogue, news articles as well as social media comments so that the corpus represents both formal as well as informal language use. In total, the corpus we have used for the verb-frame development has more than one hundred thousand sentences. Sentences are directly taken from the corpus. If there is any complex sentence or if the sentence is too large, then only the relevant clause is taken. There is no other pre-processing done on the sentences. We have analysed at least five sentences for each verb tense.

Step 4: Creating the verb-frame Files

We have generated 300 text files for 300 verb senses altogether. These 300 verb senses come from 89 Bengali verbs. Files are named after the SENSE_IDs of verb senses. Then,

these text files are populated with verb-frames. Henceforth, we will mention each of them simply as a ‘file’.

Step 5: Populating the Verb-Frame

Bengali verb-frame Network consists of 300 files, each containing information about one verb sense. Each file has two sections 1. General Information 2. Verb-frame. The General Information section includes some basic details about the verb sense, such as:

1. Sense ID
2. SENSE
3. Verb Class
4. Verbs in the Same Class
5. Morphosyntactic Type
6. Example Sentence

Sense ID represents a particular verb sense in BVFN. It is made up of the name of the verb, followed by an underscore and then a numerical such as *bhabaa_1*, *balaa_2* etc. The next header is the ‘SENSE’. Here, the English gloss is given for the verb sense. It is observed that many verb senses do not have proper English equivalents. In those cases where an exact equivalent is not available, we tried to put the most important aspect of meaning within parentheses. Thus, the English gloss given within the general information section should be regarded as an approximate meaning. ‘VERB_CLASS’ refers to the dependency-based verb classification of verbs that we have detailed in chapter 5, section 5.8.

‘VERBS_IN_THE_SAME_CLASS’ is the next header. It enlists other members of the same verb class. The fifth header in the verb-frame file is the ‘MORPHOSYNTACTIC_TYPE’. There are four morphosyntactic types: ‘transitive’, ‘intransitive’, ‘di-transitive’ and ‘causative’.

Finally, the ‘EXAMPLE SENTENCE’ header provides one example sentence to illustrate the context in which the verb sense is used in the corpus.

A verb-frame is divided into four parts, and it is developed in a tabular format comprising four columns. Each one provides a different type of information. Four columns of the verb-frame are:

1. Dependency Relations
2. Lexical Type
3. Dependency Marker
4. Ontology

The verb-frame section starts after the general information section. Some verb senses have marked for more than one verb-frame. The concept, assumptions, and annotation schemes used to populate the verb-frames are described in Chapter 4.

Step 6: Example Sentences: -

One example sentence is placed immediately before each verb-frame. Example sentences are transliterated and glossed. The ‘Itrans’ scheme is used for transliteration. Leipzig Glossing rules are followed in the glossing of example sentences.

CHAPTER 4

ARCHITECTURE OF THE VERB FRAME

The objective of this chapter is to provide details on the structure of the verb-frame. Verb-frames are designed as a table, and it has four rows. Each row represents one particular type of relevant information regarding the argument of a verb. The four rows of the verb-frame are 1. Dependency Relation 2. Lexical Type 3. Dependency Marker 4. Ontology. In this chapter, the notations that are used to annotate dependency relation, lexical type and ontology are explained with examples. The only remaining section, the dependency marker section, enlists the Vibhakti or the post-positions, which signals the dependency relations and is a surface realisation of it. Table 4.1 shows the correspondence between the affixes/post-position and the dependency role it signals.

4.1. Dependency Relation

The tag-set of dependency relations has been developed after a thorough preliminary corpus analysis. The dependency tags that are used in BVFN are mostly AnnCorra dependency tags with a few additions. The additional dependency relations represent some specific construction in Bengali which were not available in AnnCorra. Along with Paninian karakas and their subtypes, five non-karaka relations features are on the list.

1. **k1(karta):** - It is the Paninian karta, i.e., the logical subject of the sentence. Thematically it can represent an agent, theme or experience. Usually, ‘k1’ in Bengali is not signalled by any vibhakti/post-position. In rare cases, it can be inflected with locative case markers *e* or *te*.

Ex: - 4.1. *baccha(k1) talmal karte karte cal-che*
child staggering walk-PROG
The child is staggering.

2. **k1s ((vidheya karta - karta samanadhikarana):** - It is the ‘subject complement’ dependency relation and occurs with copula/copula-like verbs in Bengali. It is marked by k1s in verb-frames. Bengali copula verb *haoya* is usually omitted from the sentence when it has no other semantic function except linking the subject with subject complement, such as in the following:

Ex: -4.2. *raam (k1) khub buddhimaan (k1s)*
 Ram very intelligent
 Ram is very intelligent.

3. **k1im (impersonal subject):** - Bengali impersonal subject is a subject construction which does not agree with the verb and takes the genitive suffix *er/r*. It is very different from the dative subject found in Hindi and other Indian languages. The tag for impersonal subject is ‘k1im’. Bengali impersonal subject is discussed in detail in the section 5.9.2 of the 5th chapter.

Ex: - 4.3. *aamaa-r (k1im) khide(k2) pey-eche*
 I-GEN hunger feel
 I feel hunger.

Three dependency relations **pk1, jk1 and mk1** which are used in causative sentences, explained below.

4. **pk1 (causer):** - Causativisation is executed through morphological processes in Indian languages, including Bengali. When a transitive or intransitive verb is causativised, it adds one more argument to its existing argument structure. Semantically the extra argument can be termed as causer who/which makes the causee (another argument) to do the activity or to have undergone through the process indicated by the verb. In the annotation scheme, the causer is tagged as pk1.

Ex: - 4.4. *jaadukar (pk1) rumaal-ke (jk1) phul(k2) baanaa-len*
 magician handkerchief-ACC flower make-PRF
 The magician has transformed the handkerchief into flower.

5. **jk1(Causee):** - jk1 is the entity who/which is made to perform an activity or is made to undergo a process by the causer pk1. An Ex is given in sentence 4 above.
6. **mk1 (madhyastha karta ‘mediator causer’):** - Hindi has two levels of causativisation. Each level adds one more argument. AnnCorra marks first level causer with pk1. In the case of the second level causativisation, the two causers are marked with pk1 and mk1. Bengali lacks this morphology of second-level causativisation altogether. However, underlying semantics is not much different. Hence, we have retained this label. mk1 refers to a situation when somebody “causes an activity with the help of someone else” (*Bharati et al, 2012*). The second causer is marked with mk1. In the case of Bengali, instrumental post-position *diye* signals mediator causer in a sentence.

Ex: - 4.5. *aamraa(pk1) atithi-ke (mk1) diye baacchaa-der (jk1)*
 We guest-ACC post-position children-GEN
puraskaar(k2) deoy-aa-laam
 gift give-CAUS-1

We have distributed the gifts to the children through the guest.

7. **ras-k1(associative karta):** - ras-k1 represents a comitative case relation. It suggests accompaniment. The accompanying agent is marked with ras-k1 in BNFN. In Bengali, this dependency relation is signalled by the post-position *saathe*.

Ex: - 4.6. *raam(k1) baabaa-r (ras-k1) saathe baajaar(k2) jay*
 Ram father-GEN post-position market go
 Ram goes to market with his father

8. **cok1 (coordinating karta):** - It indicates an activity which by nature involves more than one participant, and every participant is equal in terms of the event structure. ras-k1 and cok1 are structurally very similar and are marked by the same post-position *saathe* in Bengali. The difference is semantic as the former represent a sense of association and the latter represent a sense of co-ordination in which both the agents are equal participants in the activity

designated by the verb. This dependency role distinguishes social interaction verbs like *kathaa balaa* ‘to talk’, *prem kara* ‘to be involved in romantic relation’ etc.

Ex: -4.7. *raam(k1) siitaa-r(cok1) saathe prem kare*
 Ram Sita-GEN post-position to have a love relation
 Ram is in a romantic relationship with Sita.

9. **k2 (karma):** - k2 is the direct object of a sentence. In Bengali generally *-ke* Vibhakti marks k2 if the object is animate. If the direct object is inanimate, *-ke* marker is optional.

Ex: - 4.8. *tini(k1) aamaa-ke(k2) Daak-len*
 he/she me call-PRF
 He/she has called me

10. **k2p (Goal, Destination):** - It is regarded as a sub-type of k2. In Bengali, the locative marker *-e/te* denotes this dependency relation. It refers to the culmination point of a motion and is used in the verb-frames of the motion verbs like *aasaa* ‘to come’, *jaaoyaa* ‘to go’.

Ex: - 4.9. *uni(k1) gechi-len okhaane(k2p)*
 He/she go-PRF there
 He/she has gone there.

11. **k2s (object complement):** - Object complement is tagged as k2s in verb-frames. Object complement can occur without any overt dependency marker. However, often it is marked by post-positions like *bole*, *hisaabe*, *moto* etc.

Ex: - 4.10. *aami (k1) tomaa-ke (k2) sabsamay bandhu bole bheb-echi*
 I you-ACC always friend as perceive-PRF
 I have always perceived you as my friend.

12. **k3 (instrument):** - It denotes the instrument of action. Post-position *diye* or *dvaaraa* marks k3 in Bengali.

Ex: - 4.11. *aami(k1) bled(k3) diye nakh(k2) kaaT-i*
 I bled post-position nail cut-1.PRS
 I cut my nail with bled.

13. k3s (secondary instrument): - Few verbs like *baa.Ndhaa* ‘to tie’ can take two different instrumental arguments. The first one is a primary or active instrument. The second one is a secondary or passive instrument. In Bengali, this dependency role is generally marked by *saathe* post-position.

Ex: - 4.12. *ghoraaTaa-ke (k2) gach-er(k3s) saathe baa.Ndh-la*
 The horse-ACC tree-GEN post- position tie-3SG.PRF
 He/she has tied the horse to a tree.

14. k4 (recipient/beneficiary): - k4 is the recipient of action. This dependency relation features in double object sentences like the one below. The direct object is tagged as k2, and the indirect object is tagged as k4.

Ex: -4.13. *o (k1) amay(k4) tiris haajaar Taakaa (k2) dich-Che*
 he me thirty thousand Rs. Give-PRS.PROG
 He is paying me thirty thousand rupees.

15. k5 (source): - k5 manifests the source of an activity. The post-position *theke, haye* marks this dependency relation in Bengali.

Ex: - 4.14. *baba(k1) delhi(k5) theke phir-echen*
 Father Delhi f from return-PRS.PRF
 Father has returned from Delhi.

16. k5a (on behalf of): - This is a sub-type of k5 in which the verb takes the identical ablative vibhakti/post-positions. But, unlike k5, the dependency relation k5a does not indicate the source or origin of the activity; rather, it indicates a sense which can be roughly translated as ‘on behalf of’. This dependency role mainly occurs with communication verbs like to speak, to opine etc.

Ex: - 4.15. *aamaader sabar(k5a) hoye harenbabu(k1) balben*
 all of us on behalf of Mr.Haren speak+FUT
 Mr. Haren will speak on behalf of all of us.

17. K5prk (Source Material): - It is another sub-type of k5. It occurs with the creation verbs like *tairi karaa* ‘to make’. It indicates the source material of creation. Instrumental post-position *diye* marks this k5a.

Ex: - 4.16. *maaTi(k5prk) diye pratimaa(k1) tairi kare*
 clay post-position the idol make
 Clay is used to make the idol.

18. k7p/k7t/k7 (location): - This is Paninian *adhikarana karaka*. These are the three subtypes of the same dependency relation. k7p suggests location in space, k7t location in time and k7 location elsewhere. Bengali locative suffix *-e/te* marks this dependency relation.

Ex: -4.17. *egulo(k2) oi jayga-y(k7p) daa-o*
 these there-LOC put-IMP
 Put these things below.

Ex: -4.18. *oi samay-e(k7t) cale jaa-bo*
 that time-LOC reach-1.FUT
 I/we will reach by that time.

Ex: -4.19. *biShay-e(k7) kathaa bal-o*
 topic-LOC speak-IMP
 Speak on the topic.

19. rt (reason, purpose): - Dependency role reflecting reason or purpose is tagged ‘rt’ in verb-frames. The occurrence of the post-position *janya* is the overt syntactic clue to this dependency relation.

Ex: - 4.20. *o(k1) ei kaj-er(rt) janno amay(k4) tiris haajaar*
 he this work-GEN for me thirty thousand
 Rs. Give-PROG
 Taakaa(k2) di-chChe
 He is paying me thirty thousand rupees for this work.

20. rd (direction): - The argument marking direction is tagged as ‘rd’. This dependency relation features in the verb-frames directional verbs like *pheraa_2* ‘to turn’, *ghora_1* ‘to revolve’ etc.

Ex: - 4.21. *ami(k1)pichan(rd) phir-laam*
 I behind turn around-1.PRF
 I turned back.

21. **rkc (Metaphorical Nearness):** - This dependency relation does not appear in AnnCorra as the construction is unique to Bengali. The post-position *kaache* ‘near’ is the indicator of this dependency relation. When the post-position *kaache* is used in its literal sense of spatial proximity, it is annotated as k7p dependency relation. But the word *kaache* sometimes indicates metaphorical proximity in various situational/social/emotional contexts instead of the spatial sense. All the metaphorical uses of the post-position *kaache* are subsumed under the dependency relation rkc. Some of them are very interesting and display very innovative use of this locative post-position by the language. One sentence which features rkc relation is given below.

Ex: - 4.22. *cheleTaa(k1) ramenbaabu-r(rkc) kaache pare*
 The boy Ramenbabu-GEN post-position tutor
 The boy gets tutored by Ramenbabu.

Sentence 4.22 represents an event in which the learner/student is being tutored by a teacher. Bengali employs the post-position *kaache* which literally means ‘near’, to convey the sense of ‘getting tutored’ in this scenario. The second argument or the participant who does the tutoring is tagged as rkc. This dependency relation is attached to the learning verbs like *shekhaa_1* ‘to learn’, *paRaa_1* ‘to study’ etc. This structure does not have an equivalent in Hindi. Apart from it, there are also other contexts as well in which the sense of metaphorical proximity is used and is annotated with rkc, for example.

Ex: - 4.23. *onaa-r(k1) kaache bhaalobaasaa(rkc) ache*
 S/he-GEN post-position love has
 S/he has love with him/her.

22. **rsd (durative expressions):** - Durative expressions are tagged as rsd. Unlike k7t, instead of suggesting a point in time, rsd signifies the duration of an activity. Bengali mainly has two

durative expressions 1. *dhare*- it suggests any temporal duration 2. *theke...parjanta* – it suggests the starting and culminating point of an activity.

Ex: - 4.24. *dudin dhare (rsd) bRRiShTi(k1) ha-cche*
 Two days DUR rain happen-PROG
 It has been raining for two days.

Ex: - 4.25. *shanibaar theke sombaar parjanta (rsd) chuti (k1) thaak-be*
 Saturday from Monday to vacation stay-FUT
 There will be a vacation from Saturday to Monday.

23. ras-neg (negation in associative): - In a sentence which states that *karta* accomplishes the activity without the associative participant, the associative participant is marked as res-neg.

Ex: - 4.26. *raam(k1) bon-ke(ras-neg) chaaRaai ge-lo*
 Ram sister-ACC without go-PST
 Ram went without his sister.

24. vmod (Verb Modifier): - In Bengali Verb Frame Network, vmod refers to the perfective participle clause (Thompson, 2010, pp. 411-425) like the following.

Ex: - 4.27. *raam-er Daak(vmod) shun-e shyaam(k1) phir-lo*
 Ram-GEN call hear-PTCP Shyaam return-PRF
 Shyaam turned around as he heard Ram calling.

In AnnCorra, it is an underspecified tag whose purpose is to represent various non-finite verbal structures (Bharati et al., 2006, p. 56). In Bengali, only perfective participle structures needed to be marked in verb-frames. vmod should not be confused with lexical modifiers of verbs like a single adverb and others. Such lexical elements are not to be marked at the dependency level.

There are a total of 26 dependency relations, described above, used in the development of the Bengali Verb-frame Lexicon. Not all dependency relations have a similar frequency of occurrence; some dependency roles like k3s or rkc have very limited occurrence.

4.2. Lexical Type

This is the second row of the verb-frame architecture. Here lexical types of the verbal argument are annotated. An argument can be a phrase or a complementiser clause. It can also be a single lexical item. In this case, it will be regarded as a phrase that has only one item element under it. There are only five notations to mark the lexical types: N, ADJ, ADV, INF and CL.

‘N’ stands for noun phrase. Most Bengali verbs take NPs in subject, object, as well as in other positions.

Similarly, ‘ADJ’ represents adjectival phrases. The dependency relation *k1s* is usually marked with ‘ADJ’ in the verb-frames.

‘ADV’ represents adverbs and adverbial phrases. In the context of BVFN, the ‘ADV’ is mostly employed to denote adverbs of place. It is almost always attached to the dependency relation ‘rd’ which refers to direction. Directional verbs like *pheraa_1* ‘turn’ take an adverbial argument.

Cognition and perception verbs like *jaanaa* ‘know’, *dekhaa* ‘see’, *shonaa* ‘hear’, *bhaabaa* ‘think’ can take a clausal argument as the object. Clause means the complementizer clause like the English ‘that’ clause and Hindi ‘ki’ clause². It is annotated as CL.

The last notation under Lexical Type is ‘INF’. It stands for infinitival structures. In Bengali, verbs like *paaraa* ‘to be able to’ can take infinitival construction as object.

4.3. Dependency Marker

² Bengali has two complementizers ‘*je*’ and ‘*bole*’ (Bayer, 2001). ‘*je*’ acts like the clause initial complementiser like English ‘that’ and Hindi ‘ki’, while ‘*bole*’ is a quotative complementizer.

Dependency markers are the affixes or particles that inform the existence of dependency relations. In Paninian grammar, the inflections which mark karaka relations are called *vibhakti*. Each karaka has its own set of *vibhaktis* in Sanskrit. This system, to a lesser extent, is inherited by most of the modern Indo-Aryan languages. Bengali has only three *vibhaktis* *-ke*, *-e/te* and *-r/er* which mark accusative, locative and genitive cases, respectively. *-ke vibhakti* marks the object. However, there is a distinction between animate object and inanimate object. *-ke Vibhakti* is mandatory for animate object but optional if the object is an inanimate entity. *-e/te* is the locative (*adhikarana*) *vibhakti*. The suffix *-e* gets attached to the consonant stem, whereas the suffix *-te* is attached to the vowel stem. The last one, *-r/er* is the genitive suffix in Bengali. The usage distinction between *-r* and *-er* is similar to that of *-e* and *-te*. The suffix *-r* is added to a vowel stem; on the other hand, *-er* is added to the consonant stem. Genitive relation is not important in analysing verb-frame because genitive is the relation between nominals and does not represent an argument of any verb. Thus, it does not appear in verb-frames. But, in Bengali, there is an impersonal subject which is marked with *-r/er vibhakti*.

These *vibhaktis* mark direct object, indirect object, locative relation and impersonal subject in Bengali. All other dependency relations are marked by post-positions in Bengali. Post-positions are function words derived from either nominals or verbal forms. In a sentence, their position is immediately after the argument. Most post-positions are highly flexible in their meaning as well as their usage. They usually have many functionalities apart from being able to mark dependency relations. Table 4.1 below shows Bengali *Vibhakti* and Post-positions and the dependency roles they are attached to. We have identified these patterns of correspondence through corpus analysis. It gives a good overview at a glimpse, though by no means is it a complete description of the usages of *vibhakti/post-positions*. It is clear from the table 4.1 that it is not always a one-to-one correspondence. Some dependency roles, like k5 or k7 can be expressed in many different ways.

While depending on the context, the same dependency marker can designate different dependency relations.

<u>Dependency Relations</u>	<u>Vibhakti/Post-position</u>
k1	0
cok1	0
	<i>saathe</i>
	<i>moddhye</i>
ras-k1	<i>saathe</i>
k1im	<i>r/er</i>
k1s	0
k2	0
	<i>ke</i>
k2p	0/e/te
k2s	0
	<i>hisaabe/bole</i>
	<i>moto</i>
k3	<i>diye/dvaaraa</i>

k3s	<i>saathe</i>
k4	<i>ke</i>
k5	<i>theke</i>
k5prk	<i>diye/dvaaraa</i>
k5a	<i>theke/pakShe/haye</i>
k7p	<i>e/te</i>
	<i>kaache</i>
	<i>maddhye</i>
k7t	<i>e/te</i>
k7	<i>e/te</i>
	<i>upar</i>
	<i>maddhye</i>
	<i>niye</i>
rt	<i>janya</i>
	<i>e/te</i>
rd	<i>dike</i>
	<i>upare/nice</i>
rkc	<i>kaache</i>

rsd	<i>dhare</i>
	<i>theke...parjanta</i>
ras-neg	<i>chaaRaa</i>
	<i>bina</i>
	<i>badale</i>
	<i>baad diye/baade</i>
	<i>maddhye</i>
pk1	<i>0</i>
mk1	<i>ke diye</i>
jk1	<i>ke</i>
vmod	<i>0</i>

Table 4.1: Dependency Relations and Dependency Markers

4.4 Ontology

The term ‘ontology’ originated in Philosophy, where it means the study of existence. After the advent of artificial intelligence and machine learning, the term has begun to imply any kind of hierarchical knowledge representation. Here, we will consider ontology as a technology resource in contrast to a philosophical concept. Ontology is now becoming an important aspect of knowledge engineering and information science research. Most WordNets either directly use some kind of ontology framework or are mapped to any established upper ontology. Most noted ontological models are developed in some sort of machine-readable format. Ontological tagging

shows what a linguistic item refers to in the real world. An ontological annotation framework has been developed for the Bengali Verb-frame Network. The ontological schemata used here are not adopted from any existing resource; rather, we have built the ontological schemata from scratch.

Ontologies can broadly be classified into two categories. First one is called **domain-specific ontology**. It represents the hierarchy of concepts in a particular area of knowledge, like physics or biology. The Second is known as **upper ontology** or universal ontology, or top ontology. These are general-purpose ontologies seeking to represent a vast area of knowledge. Since the purpose of developing an ontological framework is to annotate verbal arguments, the ontology has to be an upper ontology. For, arguments of verbs are not restricted to one specific domain. However, we did not attempt to produce a complete upper ontology just like Cyc (Lenat, Prakash and Shepherd, 1986) or SUMO (Pease, Niles & Li, 2002). Rather, we have employed a collection of ontological features and used the presence or absence of the said features to annotate the data. There are many reasons for such a simplistic approach.

- i) We are only annotating the nouns. Hence, we don't require features that may not be useful in tagging nouns.
- ii) The focus is to facilitate verb-sense disambiguation and also to classify verb senses on the basis of their shared semantic similarities. A small number of concepts can neatly execute the above tasks.
- iii) Large ontological frameworks are not built for this kind of annotation task. We have specific needs and require an ontology that is tailored-made for this specificity.
- iv) Apart from that, ontological frameworks like SUMO have thousands of terms, making them unusable for verb-frame annotation. So many features would have been confusing for the readers/users of the verb-frame as well as problematic for machine learning.

The ontological model we have developed features natural classes and uses the + and – symbols to denote the presence or absence of a feature. A word will be analysed in terms of binary opposition where + means the existence of a feature, - means the absence of a feature. The word ‘man’ can be tagged as [+concrete, +natural, +living, +animate, +human]. The last feature entails all the previous features. Hence, the word ‘man’ can be annotated just as [+human]. By stating it as a human entity, we are implying all the other features like concrete, living and natural. The feature ‘animate’ includes both humans and animals. We can analyse the word ‘animal’ as [+animate -human], as only the other entities that belong to the class of animate (except human) are animals. Since this ontological model is based upon binary opposition, we can understand that [-human] means animal. The feature +animate need not be mentioned while tagging human or animal entities. Thus, human beings are marked as [+human], and animals are marked as [-human]. In this way, we can analyse a lot of concepts using a small number of features.

The feature matrix that is currently being used under ontology is as follows-

1. ± abstract
2. ± concrete
3. ±object
4. ± liquid
5. ±region
6. ±artefact
7. ±living
8. ±animate
9. ±human
10. +organisation
11. +time
12. + content

- 13. + body part
- 14. +energy
- 15. + event
- 16. + language representation
- 17. + sound
- 18. + currency

4.4.1. Defining Ontological Terms

We have made this ontology format to be self-explanatory. Only by looking at a feature's name anyone can anticipate its function, thereby greatly enhancing the readability of verb-frames. There are a total of 18 unique ontological features used in verb-frames. It is a semi-hierarchical feature matrix which starts from two contrastive features named **physical** and **abstract**. The entity **abstract** is not further sub-classified. On the other hand, **physical** is the top feature which is bifurcated into two features **concrete** and **region**. The feature **concrete** is again bifurcated into two features: object and **substance**. In this way, ontology is developed on the basis of binary contradiction in which **the physical** is the main node from which most other features descend. Apart from that, there are eight features, namely **event**, **currency**, **language representation**, **content**, **sound**, **body part**, **energy** and **time**, which are not part of the hierarchy.

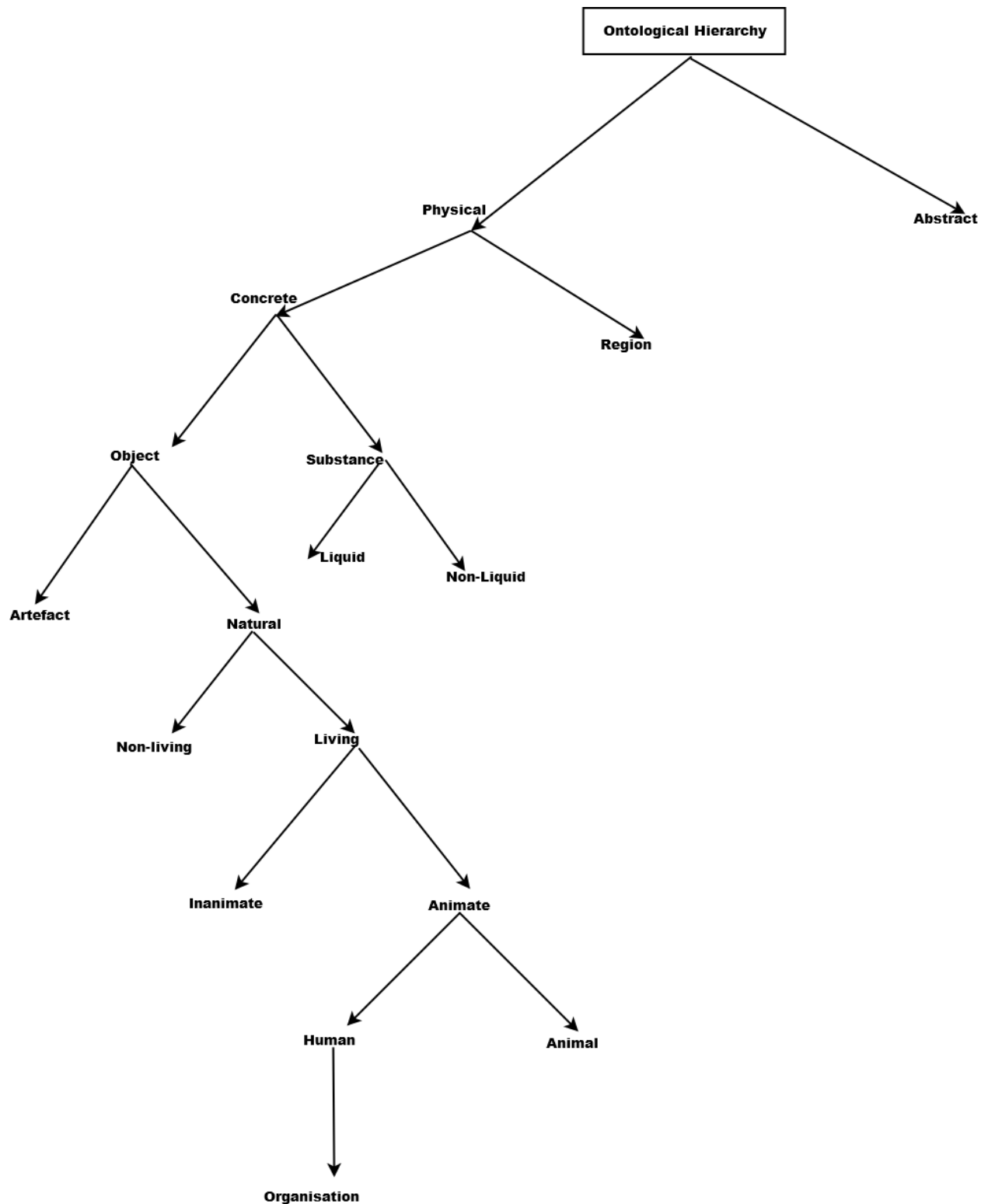


Figure 4.1: Ontological Hierarchy

Physical: - Physical can be defined as an entity that can be touched and seen as well as occupies a portion of three-dimensional space. It is one of the two top nodes in the ontology framework. It is the opposite of abstract, as the entities which are not abstract are defined as

physical. Physical includes both concrete as well as region. It is denoted as **-abs** in the ontology framework.

Abstract: - Abstract can be defined as the non-physical entities which cannot be touched, may or may not be seen but are usually felt, recognised or understood. Abstract entities do not occupy three-dimensional space. It is denoted as **+abs** in the ontology framework.

Concrete: - Any physical entity apart from a region is a concrete entity. Object and substance are the two elements that are classified under concrete. It is marked as **+con**

Region: - Region is defined as any specific reference to a section of three-dimensional space on which concrete entities can stay, and events can take place. It is marked as **+reg**

Object: - Concrete entity which is constructed by some parts and the parts do not resemble the whole in their properties and shape, is defined as an object. For example, a table is an object. It is made of some parts. None of the parts can be called a table. If we break the table into pieces, none of the broken pieces will resemble a table. It is marked as **+obj**. The category object is bifurcated into two ontological categories Natural and Artefact.

Substance: - Substances are those concrete entities which are not objects, like sand, powder, soil, pebble, stone, water, oxygen etc. If we isolate some part from an instance of sand, we will get two instances of sand. It is marked as **-obj**. Substance is divided into two categories Liquid and Non-liquid.

Liquid: - Liquid substances like water and milk are tagged as **+liq** in verb-frames.

Non-liquid: - The category non-liquid comprises solid and gaseous substances. It is marked with **-liq** in verb-frames.

Natural: - Naturally originated objects are defined as Natural in ontology. Natural consists of both living and non-living natural entities. It is contrasted with artefacts, i.e., man-made objects. Natural is tagged by **-art** in the ontology.

Artefact: - Any man-made object is designated as an artefact. It is **+art** in the ontology.

Living: - All living entities are marked as **+liv**. It is further branched into two categories: animate and inanimate.

Non-Living: - Naturally originated non-living objects are tagged as **-liv** in the ontology.

Animate: - Among the living entities, animate are those who can move of their own volition. Human beings and the animals are grouped together into the category of animate. It is marked as **+anim**. Human and animal are sub-categories of animate.

Inanimate: - The living entities which cannot move of their own volition are inanimate. It is marked as **-anim**.

Human: - It is marked as **+human**. +human implies a single human being, a group of humans or an organisation. The category organisation comes under the node human.

Animal: - Textual reference to animals is tagged as **-human** in the ontology model.

Organisation: - A socially recognised body of humans which has some specific structure and functions and is usually engaged in various activities of human society is an instance of an organisation. It is a sub-node under the category human. It is marked as **+org**.

All of these above 17 features are structured in a hierarchical design where physical and abstract are two most primal nodes and all other 15 nodes are descended from the concept physical. Each concept manifests itself in terms of binary contrast. Each parent node is bifurcated into two child nodes; the children nodes share all the characteristics of the parent nodes, mutually exclusive to each other. We have another eight features which are not a part of the hierarchical order. These features are hard to fit in any hierarchical modelling; most of them can convey the characteristics of more than one above-mentioned ontological features, depending upon the context. Features like body-part can point to a human or an animal. Yet, these features are very important not only for sense-disambiguation but also for understanding the event structure of certain verbs. These features are event, content, time, currency, sound, language representation, body part, and energy.

Event: - Any activity, sequence of activities or process is defined as an event. Instances of an event are lecture, wave, war, hibernation, working, flowing etc. It is marked by **+event** in the ontology.

Currency: - Direct references to money and payment in the text are annotated as currency in ontology. It is annotated as **+cur**. It is especially important in the case of verbs of transfer like to give, to pay, to send etc., whose semantics invariably contain the event of a monetary transaction.

Time: - Time is marked by **+time** in the ontology.

Language Representation: - This ontological category should be understood as a reference to the use of specific linguistic items. The mode of representation can be verbal speech, written documents, digital documents or even a thought. Its use is mostly limited to the verb-frame of speech and write verbs. The feature is annotated as **+LR**.

Sound: - Both speech and non-speech sound is marked by **+sound** in verb-frames.

Energy: - Elements like light, fire, thunder, electricity, and plasma which do not belong to any of the substance categories solid, liquid and gas are **+energy**.

Content: - Content can be defined as any object or event that contains or has the potential to contain some intellectual content. For example, books, songs, a novel, a CD etc.

Body Part: - Human or animal body part is tagged as **+BP**.

Ontologically Unrestricted Argument: - Sometimes, no ontological generalisation is possible because there is no selectional restriction on the verb. In other words, the argument can fall under any ontological category. It is considered to be a case of ontologically unrestricted argument, and its position is left blank in the verb-frame.

4.4.2 Importance and Issues:

This section will show the significance as well as issues regarding ontology in BVFN.

4.4.2.1 Role of Ontology in Bengali Verb Frame:

Ontology is an added component in verb-frames which is otherwise built on Computational Paninian Grammar. The task of ontology is to provide some other kind of information which cannot be obtained by structural-semantic analysis of the Paninian dependency framework. Dependency roles show the semantic connection between the verb and the argument. On the other hand, ontology tells us what kind of entity the argument is in the real world. Ontological information is not a relationship between the verb and the argument but rather a property of the argument itself. Ontological features do not denote a linguistic entity but rather the real-world object or concept that the linguistic item signals. In this way, ontology is language-independent and universally applicable with regard to the language. We need such language-independent resources when we try to systematically decode the real-world information encoded within linguistic structures.

Consider the following examples taken from the corpus,

Ex – 4.28 *tini aamaa-ke Daak-len*
 he/she I-ACC call-PRF
 He/she has called me.

Ex – 4.29 *Svaamii-ke daakla strii*
 Husband-ACC call-PRF wife
 Wife called her husband.

Ex – 4.30 *poShaa kukurTi-ke daaklen bRRiddha*
 pet dog-ACC call-PRF old man
 The old man has called his pet dog.

The verb sense used in the above sentences is the verb *Daakaa_1*³ ‘to call’(as per the BVFN) which implies the activity of verbal calling meant to attract the attention of the

³ The complete verb-frame of *Daakaa_1* and *Daakaa_2* is given in the section 4.5.

hearer. This activity needs two elements; the agent who enacts the activity of calling and a patient who is being called. As illustrated in the example sentences above, the agent or the first argument of the verb *Daakaa_1* is always a human. For, only humans have the ability to call someone (at least in the real-world scenario). The second argument or patient can be a human or an animal, i.e., an animate entity. Hence, in the verb-frame of *Daakaa_1* (see section 4.5 for complete verb-frame), the ontological tagging of the the first argument is ‘+human’ and the second argument as ‘+anim’. In this way, the framework is able to capture the generalisation regarding the verb sense in question.

VERB_FRAME

DEPENDENCY_RELATION	k1	k2
LEXICAL_TYPE	N	N
DEPENDENCY_MARKER	0	ke
ONTOLOGY	+human	+anim

Now, let us look at the corpus sentences of another sense of the same verb *Daakaa_2* ‘to make a noise’, which implies the activity of animals making a sound in Bengali.

Ex – 4.31 *gaadhaa theke-theke Daak-chila*
 donkey at times sound-PST.PROG
 The donkey was braying

Ex – 4.32 *kukur gheu gheu kare Daak-che*
 dog bark post-position sound-PROG
 The dog is barking.

Similarly, these examples clearly show that the agent of the verb *Daakaa_2* can only be an animal. Therefore, the ontology is tagged as ‘-human’ in the verb-frame.

VERB_FRAME

DEPENDENCY_RELATION	k1
LEXICAL_TYPE	N
DEPENDENCY_MARKER	0
ONTOLOGY	-human

In the above examples, the contextual difference between two verb senses is clearly revealed in the form of a tabular verb-frame. Machines can make use of such verb-frames to identify the correct sense from the context. Here the verb-frame is different in terms of both the dependency relations as well as ontology. But, sometimes two senses of an ambiguous verb appear to have identical dependency structures. Then, ontology becomes the most important marker of contextual difference. The Reverse is also true. In other words, ontological information complements dependency information. Often, ontology presents novel aspects of the context which cannot otherwise be revealed.

4.4.2.2 Issues and Limitations:

Adam Pease, in a 2011 lecture, said, “creating a large ontology does consist of a lot of trivial assertions about the world. Things that people already know. But the fact is computers do not know them” (Pease, 2011). He was then talking about one of the primary criticisms directed at SUMO (Suggested Upper Merged Ontology). Ontologies are usually built for machine learning purposes. The information that ontologies represent may not appear unique to people, but the computer needs this kind of world knowledge to perform NLP tasks. The ontology, which is used for the development of BVFN, is specifically developed to categorise nominal arguments of Bengali verbs. The objective is to arrive at some categorical generalisations regarding the argument themselves which is otherwise consistent with the dependency structure information and complements the same. Though ontology performs well in its task, it has some limitations. Some limitations are coming from the purpose of the

ontology itself. For example, the ontological model cannot account for a supposed fantasy fictional or mythological situation where everything, even a stone, can talk. We cannot include this kind of fantasy or mythology scenarios in the ontology because then it will be open-ended and can no longer be used for meaning disambiguation. For, in fantasy situations, everything is possible. We won't be able to find necessary constraints on which we can base generalisation. There is another kind of limitation which is more subtle. The ontology will not be able to take care of the scenario where a robot or speech generation system is talking. The ontology used here is an upper ontology; it is not designed to take care of specific situations such as robotic speech. This information, if necessary, can be conveyed to the machine by other means. For example, we can also develop a domain-specific ontology such as ontology of speech technology, to handle such scenarios. Hence, the ontological information in the Bengali verb-frame provides categorical generalisation regarding verbal arguments.

The design of the verb-frame resulted in a multi-layered web of information represented in a tabular format. Each column in the table manifests one mode of information. Ontology provides information regarding the constraints on the arguments of a verb. For this reason, it is sometimes designated as the 'selectional restriction' (Kipper Schuler, 2005, pp. 35-36). Ontology informs what the entity is and what it is not in terms of the real world. Another thing that needs to be noted here is that ontology is used to annotate only the nominal arguments. Adjectival, adverbial, infinitival and clausal arguments are not meant to be marked using ontology. In the case of those arguments mentioned above, the ontology position is marked by a cross 'x'.

4.5 Full Verb-frame of Daakaa_1 and Daakaa_2

Complete verb-frames of the verbs Daakaa_1 and Daakaa_2 are given below.

Daakaa_1 ‘to call’

SENSE_ID Daakaa_1

SENSE ‘to call’

VERB_CLASS 2(k1+k2)

VERBS_IN_THE_SAME_CLASS bhaabaa_1

EXAMPLE_SENTENCE tini aamaa-ke Daak-len
he/she me call-PRF
He/she has called me

VERB_FRAME

DEPENDENCY_RELATION	k1	k2
LEXICAL_TYPE	N	N
DEPENDENCY_MARKER	0	ke
ONTOLOGY	+human	+anim

***Daakaa_2* ‘to make a noise’ (animals)**

SENSE_ID Daakaa_2

SENSE ‘to make a noise(animals)

VERB_CLASS 1 (k1)

VERBS_IN_THE_SAME_CLASS calaa_1

EXAMPLE_SENTENCE gaadhaa theke-theke Daakchila
donkey at times sound-PST-PROG
The donkey was braying at times

VERB_FRAME

DEPENDENCY_RELATION	k1
LEXICAL_TYPE	N
DEPENDENCY_MARKER	0
ONTOLOGY	-human

CHAPTER 5

DEVELOPMENT AND ANALYSIS

The present thesis attempts to build a repository of verbal argument structures annotated at multiple levels. The Bengali Verb Frame Network, developed as part of the research, is a linguistic resource similar to Hindi verb frames (Begum, 2017b). Lexical ambiguity, a significant challenge for any linguistic analysis, is handled efficiently by treating the different senses of a verb, each as a separate entry in the BVFN. Each ambiguous verb is broken down into different senses that convey different but related meanings. Sometimes, these meanings are homonymous, while most of the time, they are polysemous. Thus, there are 300 verb senses for which 362 verb frames have been created. The mention of ‘verb’, henceforth in this chapter, should be understood as a verb sense unless specified otherwise.

5.1 Organisation of the Chapter

The chapter is organised as follows: Section 5.2 enumerates some important characteristics of Bengali verbal system, section 5.3 discusses the statistics of BVFN with histogram and pie-chart, section 5.4 introduces some basic concepts and assumptions, section 5.5 presents verb-frames of some selected Bengali verbs, the topic of section 5.6 is the verbs which have multiple verb-frames, the observations regarding the lexical types in BVFN are described in section 5.7, section 5.8 discusses dependency relation and its statistics of usage, section 5.9 discusses ontology and its importance in verb-frame network, section 5.10 provides a detail discussion on some specific structures in Bengali in the context of verb-frame, section 5.11 classifies the verbs on the basis of verb-frames, section 5.12 is corpus observations, and finally the chapter ends with a concluding paragraph in section 5.13.

5.2 Some Important Features of the Bengali Verbal System

Before going into the verb-frame analysis, it is necessary to identify and list a few essential features of verbs in Bangla. These features are well known and much discussed in the literature of linguistics.

1. Finite verbs in Bangla are inflected for person, honorification, tense and aspect (Dash, 2015, p. 119). It also means they are not inflected for gender and number, unlike Hindi and many other Indian languages.
2. The verb agrees with the subject in person and honorification unless it is an impersonal verb (Dasgupta, 2006, p. 1).
3. In the case of impersonal verbs, the verb appears in its default form, which is third person, neutral.
4. Causativisation is a morphological process in Bengali. Causative verbs are derived from simple verbs through affixation.
5. Each Bengali verb has at least 42 finite and four non-finite verbal forms. Thompson enumerated these four types of non-finite verbal forms as verbal nouns, imperfective participles, i.e., infinitives, perfective participles and conditional participles (Thompson, 2010, p. 146).

Form	Affix	Example (with <i>kar</i> 'do' verb root)
Verbal Noun	-aa	karaa
Verbal Noun (causative)	-a+no	karaano
Imperfective Participle	-te	karte

Perfective Participle	-e	kare
Conditional Participle	-le	karle

Table 5.1: Non-finite Verb Forms in Bengali

6. A verb is listed in the dictionary in its ‘verbal noun’ form, formed by adding *-aa* to the stem.

7. An imperfective participle is marked by *-te* marker attached to the verb root and can form an infinitival clause. On the other hand, the perfective participle is marked by *-e* suffix.

5.3. Verb Senses: Number and Statistics

BVFN consists of the verb-frame files of 300 unique verb senses which are derived from 89 verbs. All of the 89 verbs are very common, frequent verbs in the language. Since the primary objective of this initiative is to resolve verb sense disambiguation, mostly ambiguous verbs are chosen. These verbs show remarkable semantic flexibility in relation to their context. It is their semantic elasticity that drives wider use of those verbs resulting in lexical ambiguity. Sometimes, even more than ten senses of certain verbs were detected and analysed. Certain unambiguous verb, i.e., verbs that have only one verb sense, has also been included for various reason, such as their unique characteristics or for the completeness of the work. Verb-frame files were developed for each verb sense. Table 1 below enlists all the verbs included in the Bengali Verb Frame Network and the number of senses relative to each verb.

Index	Verb	Senses	Index	Verb	Senses
1	baa.Ndhaa	2	45	galaa	2

2	baaRaa	1	46	gaRaa	4
3	balaa	4	47	ghoraa	3
4	bhaabaa	2	48	ghumono	2
5	calaa	10	49	haoyaa	7
6	Daakaa	3	50	jaagaa	2
7	deoyaa	6	51	jaanaa	5
8	Dobaa	3	52	jharaa	1
9	jaaoyaa	5	53	jho.Nkaa	2
10	jamaa	4	54	kaaje laagaa	1
11	jvalaa	3	55	kaamaano	2
12	kaaTaa	6	56	kenaa	1
13	khaaTaa	3	57	khaoyaa	4
14	maraa	2	58	khaoyaano	1
15	melaa	1	59	khelaa	3
16	milaa	2	60	kholaa	10
17	naamaa	3	61	kothaa balaa	1
18	phaa.Nsaa	2	62	laagaa	8
19	pheraa	2	63	laagaano	8
20	phoTaa	3	64	laRaa	1
21	saaraa	2	65	maar khaoyaa	2
22	aanaa	1	66	maaraa	2
23	aanaano	1	67	mane karaa	4
24	aasaa	7	68	meTaa	2
25	baanaano	4	69	neoyaa	5

26	bakaa	2	70	paaoyaa	6
27	balaano	1	71	paaraa	1
28	banaa	4	72	paataa	1
29	basaa	3	73	paaThaano	1
30	basaaano	2	74	paRaa	3
31	bhaalobaasaa	3	75	raakhaa	6
32	bhaangaa	5	76	shekhaa	1
33	bharaa	3	77	shonaa	5
34	bikri_karaa	1	78	shoyaa	2
35	bojhaa	4	79	Taanaa	6
36	caaoyaa	2	80	tairii karaa	2
37	caapaa	5	81	thaakaa	7
38	cenaa	3	82	thaamaa	2
39	chaaR_deoyaa	2	83	Thelaa	3
40	daa.Nraano	2	84	tolaa	8
41	dekhaa	6	85	uthaa	11
42	dekhaano	6	86	uthaano	2
43	dhaakkaa khaoyaa	2	87	haa.Ntaa	2
44	dharaa	6	88	paataano	1
89	ghoraano	3			
Total					300

Table 5.2: List of Verb Senses

Hence, there are 300 verb-frame files in Bengali Verb Frame Network. But the number of verb frames created is much larger because many verbs (i.e., verb sense) have more than one verb-frame. First, we have developed verb-frames, at least one verb-frame for

each verb; but it can be more than one. Then, verbs are grouped into various verb classes and sub-classes based on the similarity of verb frames.

As we can see from the table one verb ‘uthaa’ has as many as 11 senses. But it is an outlier, most verbs in our network have less than 4 senses. The median value of senses per verb is 3.

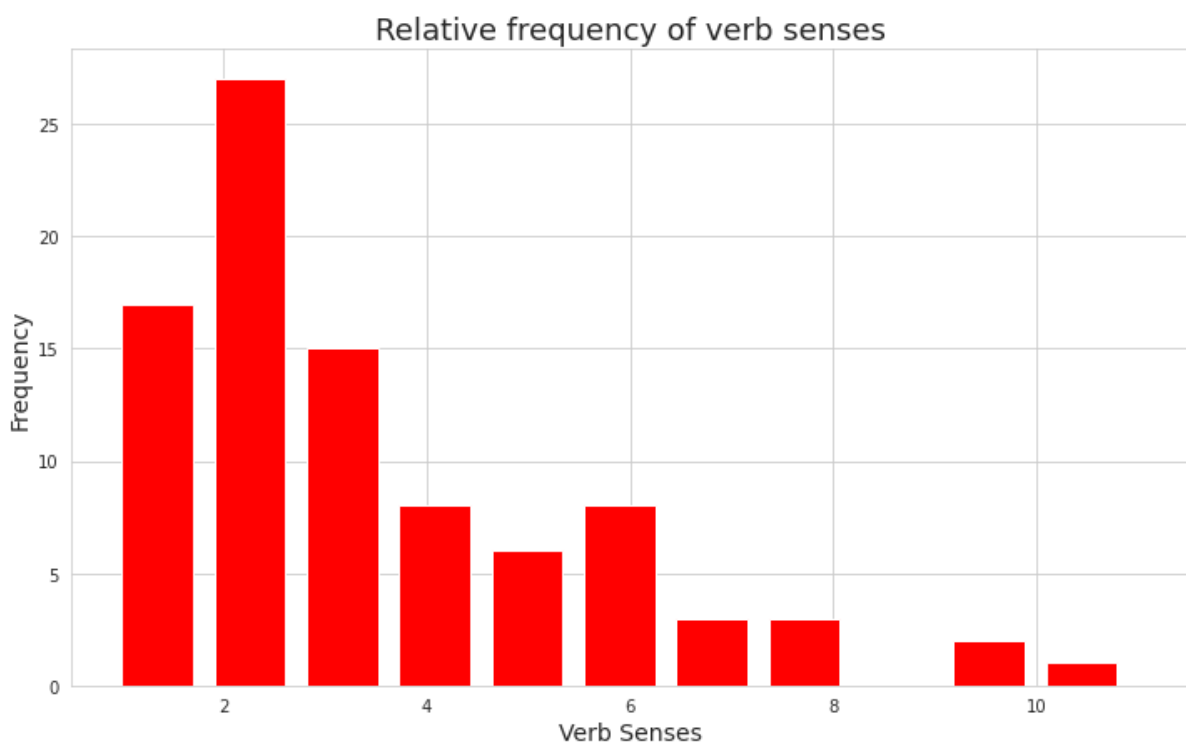


Figure 5.1: Relative Frequency of Verb Senses per Verb (Histogram)

The above histogram reflects the numerical frequencies of verb senses as they occur in the verb-frame network. Among the 89 verbs, 27 verbs have two senses which are statistically the most frequent. As it is evident from the histogram, verbs having more than 7 senses are quite rare. Only six such verbs are noted. Frequency distribution is elaborated in the pie chart below.

Frequency of verb senses in Bengali Verb Frame Network

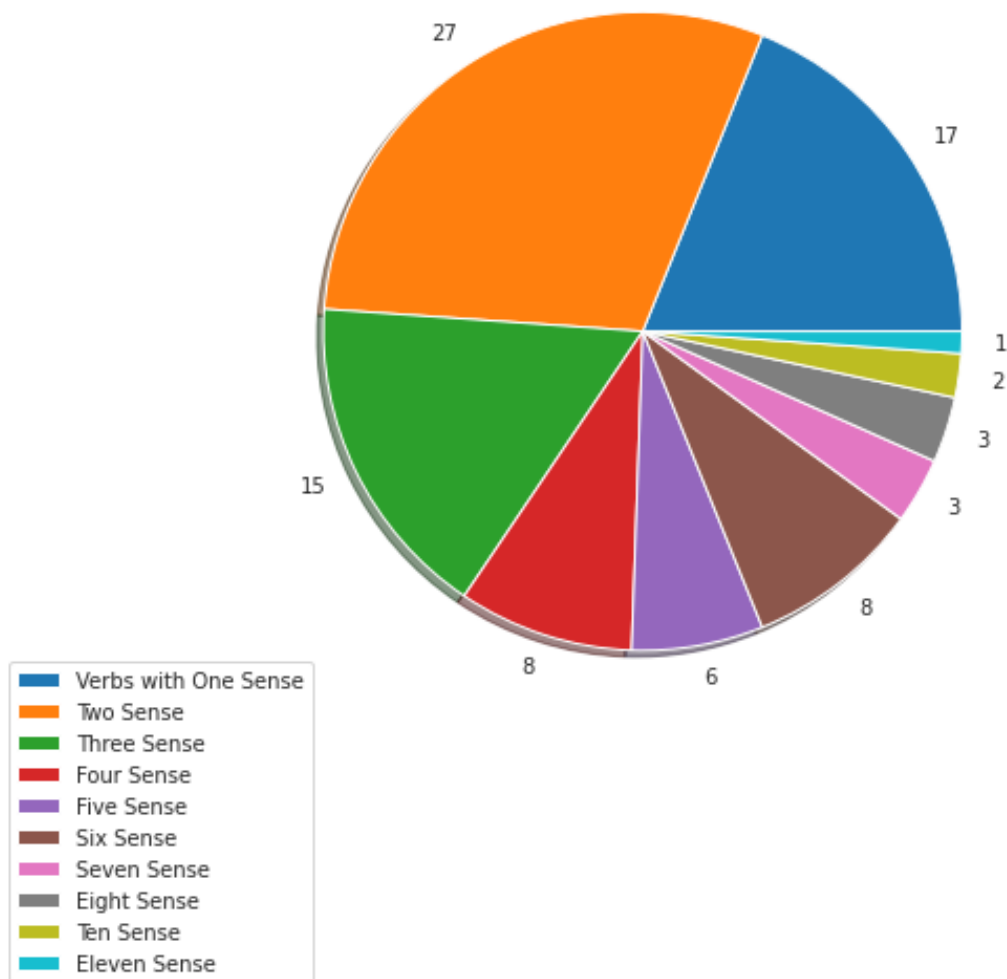


Figure 5.2: Relative Frequency of Verb Senses per Verb (Pie Chart)

5.4. Some Basic Concepts

5.4.1 Essential and Non-essential Roles

Categorisation of arguments is an unavoidable task for any verb centric formalism. The problem here is one of the mapping problems, as not all arguments are on the same semantic layer. Martha Palmer, in her work on ‘inference-driven mapping’, introduced the notion of semantic roles (Palmer, 1985, p. 39). Moerdler (1987) took Palmer’s nomenclature as she

defined the concept of obligatory, essential and non-essential roles. Obligatory roles are syntactically defined. It is of not much relevance in the case of Indian languages. Indian languages are free word order languages and have relatively lenient syntax. In Indian languages, even the subject can be omitted from the sentence. Moreover, obligatory roles are always subsumed under essential roles. Hence, we have only made the distinction between essential and non-essential roles. Essential roles are semantically required to be present in the argument structure. These roles can be anticipated from the verb's meaning even if they are not present on the surface structure. These arguments are always implied whenever a verb is used in a sentence. On the other hand, non-essential roles are neither syntactically nor semantically mandatory. They can be derived or dropped according to the requirement of the context. Essential roles form the core verb-frame, which is described below. Any non-essential role added to the core verb-frame makes it an extended verb-frame. Non-essential roles are not considered in the creation of verb-frames.

5.4.2 Core Verb-frame

Core verb-frame is the argument structure that represents a verb's semantics. It includes all of the essential roles for a verb and disregards the non-essential roles. Verb-frame can be extended; new arguments can be added and subtracted as per the demand of the context. However, core verb-frame of a verb (which means verb-sense in this context) is identical in every context as it counts only the essential arguments. For example, verb *jaaoyaa_1* 'to go' has the dependency structure [**k1,k2p**]. But this verb-frame can be extended with an extra non-essential argument 'rt' as in the sentence below.

Ex: - 5.1. *Sumit(k1) paRaashonaa-r(rt) janya kolkaataa(k2p) jaa-cche*
 sumit study-GEN for Kolkata go-3.PROG
 Sumit is going to Kolkata for the purpose of study.

This external argument **rt** neither has any relation to the semantics of the verb nor is it licensed by its semantics. The meaning of the verb *jaaoyaa_1* ‘to go’ does not predict the existence of any dependency relation such as **rt**(reason). In the above example, the argument whose dependency relation is marked **rt** comes solely from the context.

5.4.3 Case Syncretism

Case syncretism refers to a linguistic phenomenon where a single linguistic form can signal more than one grammatical case. In case of syncretism, two or more distinct cases may collapse into a single form, resulting in the loss of grammatical distinctions. Case syncretism can be partial or complete. In partial syncretism, certain grammatical relations merge in certain contexts while remaining distinct in others. Complete syncretism, on the other hand, involves the complete loss of distinctions between multiple cases, resulting in a single-case form used for multiple functions. In the present context, cases are labelled as different dependency relations. In Bengali, case syncretism between dative and accusative relations is an example of complete case syncretism. The dependency marker *-ke* is used to indicate both *k2* and *k4* dependency relations. On the other hand, partial case syncretism is also displayed by various dependency markers. For example, in the verb-frame of *saaraa_1* ‘to recover’ which is given in section 5.5, the *e/te* suffix, which is primarily a locative marker, represents ‘*rt*’ (reason) dependency relation. The conflation between multiple dependency roles is important information that needs to be understood and analysed as part of verb-frame analysis.

5.5 Verb-frames

In this section, the verb frames of a few selected verbs, chosen from different semantic domains, have been illustrated. There are four verbs *balaa*, *kaaTaa*, *naamaa*, *saaraa* represented in this section.

5.5.1 Verb-frames of the Verb *balaa*

Verb-frames of all the senses of the verb *balaa* will be demonstrated below. There are four different senses of the verb *balaa*: *balaa_1* ‘to tell’, *balaa_2* ‘to chant’, *balaa_3* ‘to be called/pronounced as’ and *balaa_4* ‘to speak on behalf of’ is identified and included in the BVFN.

An essential feature of the ‘verbs of communication’ is that these verbs can take a clause as direct object. ‘*balaa_1*’ is a communication verb whose event structure is characterised by a sender, recipient and the message itself.

***balaa_1* ‘to tell’**

SENSE_ID	balaa_1		
SENSE	‘to tell’		
VERB_CLASS	3[k1,k2,k4]		
VERBS_IN_THE_SAME_CLASS	deoyaa_1, deoyaa_2, deoyaa_3, pathaano_1		
MORPHOSYNTACTIC_TYPE	di-transitive		
EXAMPLE_SENTENCE	siimaa aamaa-ke ghaTanaaTaa bal-be		
	Seema me the incident tell-3.FUT		
	Seema will tell me the incident		
VERB_FRAME_1			
DEPENDENCY_RELATIONS	k1	k2	k4
LEXICAL_TYPE	N	N	N
DEPENDENCY_MARKER	0	0	ke
ONTOLOGY	+human		+human

EXAMPLE_SENTENCE aami tomaa-ke bal-echilaam je se bhalo lok nay
 I me tell-1.PST that s/he good person NEG
 I told you that he/she is not a good person

VERB_FRAME_2

DEPENDENCY_RELATIONS	k1	k2	k4
LEXICAL_TYPE	N	CL	N
DEPENDENCY_MARKER	0	0	ke
ONTOLOGY	+human	x	+human

EXAMPLE_SENTENCE siimaa aamaa-ke ghar-er byapar niye bol-be
 seema me home-GEN regarding post position tell-3.FUT
 Seema will tell the fact regarding home

VERB_FRAME_3

DEPENDENCY_RELATIONS	k1	k7	k4
LEXICAL_TYPE	N	N	N
DEPENDENCY_MARKER	0	niye	ke
ONTOLOGY	+human		+human

The verb-frame file starts with some general information about the verb. First is SENSE_ID. Next is SENSE. The meaning of the verb *balaa_1* is ‘to tell,’ i.e., the verb expresses an event of telling something to someone. Then comes VERB_CLASS. The verb ‘*balaa_1*’ belongs to verb class 3. Besides the verb class, the dependency arrangement from this verb class is provided within the parenthesis. As we have discussed in the previous section, this verb class is reserved for di-transitive verbs. VERBS_IN_THE_SAME_CLASS

row represents the other verbs which have the same dependency structure. Last is MORPHOSYNTACTIC_TYPE which is di-transitive in this case.

The verb *balaa_1* has three verb-frames. The table of verb-frame has four rows. The first row depicts the dependency relation, which is [k1,k2,k4] in the case of the first verb-frame. Purpose of second row is to mark the lexical category of the arguments. In the case of this verb-frame of the verb ‘*balaa_1*’, all three arguments are nominal entities. The next row is the dependency marker row. Only the third argument is marked by *-ke vibhakti*. Lastly, there is ontology. As the verb ‘*balaa_1*’ is a communication verb whose event structure is characterised by one person communicating something with another, both the agent and recipient have to be human. Hence, we have marked both first and third arguments as +human entity, while the direct object can not be marked with ontological feature because there is no ontological constraint.

The VERB_FRAME_2 is identical to that of the first one in nearly all aspects. The only difference is that lexical type of the second argument is clause instead of a nominal. This is an important feature of the ‘verbs of communication and cognition’ as these verbs can take a clause as their direct object. On the other hand, VERB_FRAME_3 has different dependency structure. Here, k2 is replaced by k7 and is no longer a double object construction. Bengali post-position *niye* here marks the new dependency relation. *niye* is specifically used with the dependency relation k7 and can be replaced with locative marker *e/te* in most cases.

***balaa_2* ‘to chant’**

SENSE_ID	balaa_2	
SENSE	‘to chant’	
VERB_CLASS	2[k1,k2]	
VERBS_IN_THE_SAME_CLASS	baa.Ndhaa_2, calaa_2	
MORPHOSYNTACTIC_TYPE	transitive	
EXAMPLE_SENTENCE	bacchaTi ram-ram bal-che	
	The child ram-ram chant-3.PROG	
	The child is chanting Ram-Ram.	
VERB_FRAME		
DEPENDENCY_RELATION	k1	k2
LEXICAL_TYPE	N	N
DEPENDENCY_MARKAR	0	0
ONTOLOGY	+human	+LR

While *balaa_1* requires three arguments: the speaker, the topic of communication and the hearer, *balaa_2* manifests an event where the speaker speaks something, but it is not intended for any hearer. It is a transitive verb that requires two dependency relations k1 and k2. Ontological characteristic of its argument structure is very distinct. Ontological property of its first argument is ‘human’, and the second is ‘language representation’.

***balaa_3* ‘to be called/pronounced as’**

SENSE_ID	balaa_3		
SENSE	‘to be called/pronounced as’		
VERB_CLASS	9[k1,k2,k2s]		
VERBS_IN_THE_SAME_CLASS	daakaa_3, bhaabaa_2		
MORPHOSYNTACTIC_TYPE	di-transitive		
EXAMPLE_SENTENCE	benaaras-ke	kashiio	bal-e
	Benaras-ACC	Kashi	call-3.PRS
	Benaras is also called Kashi		
VERB_FRAME			
DEPENDENCY_RELATIONS	k1	k2	k2s
LEXICAL_TYPE	N	N	N
DEPENDENCY_MARKER	0	ke	0
ONTOLOGY	+human		

Dependency structure of the verb *balaa_3* is unique because it takes k2s or object compliment. Bengali accusative marker *-ke* is attached to the object. The event structure of *balaa_3* needs either a human or an organisational agent. Hence, k1 is marked with +human. No ontological feature can be ascribed to the two other arguments because these can just be anything.

***balaa_4* “to speak as representative”**

SENSE_ID balaa_4

SENSE ‘to speak as representative’

VERB_CLASS 25 [k1,k5a,k7p]

VERBS_IN_THE_SAME_CLASS thaakaa_8

MORPHOSYNTACTIC_TYPE transitive

EXAMPLE_SENTENCE aamaa-der sabar hoye harenbabu okhan-e bal-ben
all of us on behalf of Mr.Haren there-LOC speak-FUT
Mr.Haren will speak there on behalf of all of us

VERB_FRAME_1

DEPENDENCY_RELATION	k1	k5a	k7p
LEXICAL_TYPE	N	N	N
DEPENDENCY_MARKER	0	hoye/pakShe	e/te
ONTOLOGY	+human	+human	+reg

EXAMPLE_SENTENCE aami puro graam-er haye tomaa-ke bal-chi
I whole village-GEN on behalf of you-ACC speak-1.PROG
I am speaking to you on behalf of the whole village

VERB_FRAME_2

DEPENDENCY_RELATION	k1	k5a	k2
LEXICAL_TYPE	N	N	N
DEPENDENCY_MARKER	0	hoye/pakShe	ke
ONTOLOGY	+human	+human	+human

EXAMPLE_SENTENCE pradhaan graam-er haye samasyagulo bal-ben
head village-GEN on behalf of problems speak-3.FUT
The village head will convey the problems on behalf of the village

VERB_FRAME_3

DEPENDENCY_RELATION	k1	k5a	k7
LEXICAL_TYPE	N	N	N
DEPENDENCY_MARKER	0	hoye/pakShe	ke

The sense of *balaa_4* is different from the rest. It is not a generic act of speaking, uttering or pronouncing. Its event structure involves a social occasion where someone speaks on behalf of certain groups, organisations or others. The dependency relation k5a represents that ‘on behalf of’ sense. Post-positions like ‘*hoye*’, ‘*pakShe*’ denote this dependency relation.

5.5.2 Some other Verb-frames

All verb-frames 3 more lexical verbs *kaaTaa*, *naamaa*, *saaraa* are presented below. These three verbs have total 11 senses, Hence, there are 11 verb-frames in this section.

kaaTaa (6 senses)

The verb *kaaTaa* has six senses. All six verb frames are given below. The verb *kaaTaa_1* is the primary sense of the verb *kaaTaa*. It refers to the act of literal cutting through instrument. Beth Levin (1993) has classified these verbs under ‘verbs of contact’, which bring about a ‘change in the state’ of the object. ‘Cut’ type of verbs demand three arguments essentially, k1, k2 and k3.

***kaaTaa_1* ‘to cut’**

SENSE_ID kaaTaa_1

SENSE 'to cut'

VERB_CLASS 6 [k1,k2,k3]

VERBS_IN_THE_SAME_CLASS baa.Ndhaa_1, bhaangaa_1

MORPHOSYNTACTIC_TYPE transitive

EXAMPLE_SENTENCE aami bleD diye nakh kaaT-i

bled post-position nail cut-1.PRS

I cut my nail with bled

VERB_FRAME

DEPENDENCY_RELATION	k1	k2	k3
---------------------	----	----	----

LEXICAL_TYPE	N	N	N
--------------	---	---	---

DEPENDENCY_MARKER	0	0	e/te/diye/dvaaraa
-------------------	---	---	-------------------

ONTOLOGY +human +obj +art

kaaTaa_2 is a currency verb. It suggests the activity of monetary deduction

***kaaTaa_2* ‘to deduct’ (amount of money)**

SENSE_ID kaaTaa_2

SENSE ‘to deduct’(amount of money)

VERB_CLASS 21 [k1,k2,k5)

VERBS_IN_THE_SAME_CLASS kenaa_1

MORPHOSYNTACTIC_TYPE transitive

EXAMPLE_SENTENCE aamar maaine theke tirish shataamsha aaykar kaaTe
my salary from thrity percent income-tax deduct
Thirty percent income tax is deducted from my salary

VERB_FRAME

DEPENDENCY_RELATION	k1	k2	k5
LEXICAL_TYPE	N	N	N
DEPENDENCY_MARKER	0	0	theke
ONTOLOGY	+cur	+cur	+cur

kaaTaa_3 is another sense of the verb *kaaTaa*. It is a physical motion verb which suggest the activity of digging a place.

***kaaTaa_3* ‘to dig’**

SENSE_ID	kaaTaa_3	
SENSE	‘to dig’	
VERB_CLASS	2[k1,k2]	
VERBS_IN_THE_SAME_CLASS	Taanaa_1, Taanaa_3, Taanaa_6 , Thelaa_2, dharaa_1, khaoyaa_1, khaoyaa_2 , tolaa_7	
MORPHOSYNTACTIC_TYPE	transitive	
EXAMPLE_SENTENCE	raasta keT-eche rakShiiraa the road dig-PRF guards The guards have dug up the road	
VERB_FRAME		
DEPENDENCY_RELATION	k1	k2
LEXICAL_TYPE	N	N
DEPENDENCY_MARKER	0	0
ONTOLOGY	+human	+reg

The verbs *kaaTaa_4* and *kaaTaa_6* are very different from the rest. These two verbs can be described as the abstract use of the verb *kaaTaa*. These two senses are also intransitive, while others are transitive.

***kaaTaa_4* ‘to recover or be okay’**

SENSE_ID	kaaTaa_4		
SENSE	‘to recover or be okay’		
VERB_CLASS	1 [k1]		
VERBS_IN_THE_SAME_CLASS	calaa_1, ghumono_1, baaRaa_1, Daakaa_2, phoTaa_2, naamaa_3		
MORPHOSYNTACTIC_TYPE	intransitive		
EXAMPLE_SENTENCE	manoj-er neshaa kaaT-la abasheShe		
	Manoj-GEN intoxication recover-PRF at last		
	Manoj recovered from intoxication at last		
VERB_FRAME			
DEPENDENCY_RELATION	k1		
LEXICAL_TYPE	N		
DEPENDENCY_MARKER	0		
ONTOLOGY	+abs		

kaaTaa_5 is a metaphorical use of the verb *kaaTaa*.

***kaaTaa_5* ‘to mark with a cross or a line’**

SENSE_ID kaaTaa_5
SENSE ‘to mark with a cross or a line’
VERB_CLASS 2 [k1,k2]
VERBS_IN_THE_SAME_CLASS balaa_2, saaraa_2, bhaabaa_1, kholaa_2, kholaa_4, bhangaa_2, bakaa_2
MORPHOSYNTACTIC_TYPE transitive

EXAMPLE_SENTENCE shikShak bhul uttar kaaT-bei
 teacher wrong answer mark with a cross-FUT
 The teacher will certainly cross the wrong answer

VERB_FRAME

DEPENDENCY_RELATIONS	k1	k2
LEXICAL_TYPE	N	N
DEPENDENCY_MARKER	0	0
ONTOLOGY	+human	+content

***kaaTaa_6* ‘to pass the time’**

SENSE_ID	kaaTaa_6
SENSE	'to pass the time'
VERB_CLASS	3 [k1]
VERBS_IN_THE_SAME_CLASS	calaa_1, ghumono_1, baaRaa_1, Daakaa_2, calaa_6
MORPHOSYNTACTIC_TYPE	intransitive

EXAMPLE_SENTENCE	samay	aar	kaT-che	naa jena
	time	particle	pass-PROG	NEG indeed
	Time is not being passed, indeed			

VERB_FRAME	
DEPENDENCY_RELATIONS	k1
LEXICAL_TYPE	N
DEPENDENCY_MARKER	0
ONTOLOGY	+time

5.5.3 naamaa

The verb *naamaa* has three senses. The verb *naamaa_1* ‘to get down’ is the primary sense of the verb *naamaa*.

***naamaa_1* ‘to get down’**

SENSE_ID naamaa_1

SENSE ‘to get down’

VERB_CLASS 14 [k1,k5,k7p)

VERBS_IN_THE_SAME_CLASS pheraa_1

MORPHOSYNTACTIC_TYPE intransitive

EXAMPLE_SENTENCE gari theke naam-ba baajaar-e
 car from get down-1.FUT market-LOC
 I will get down from the car at the market

VERB_FRAME

DEPENDENCY_RELATION	K1	k5	k7p
LEXICAL_TYPE	N	N	N
DEPENDENCY_MARKER	0	theke	e/te
ontology	+anim	-abs	+reg

naamaa_2 is the metaphorical extension of the first sense. It suggests the event of going into some activity.

naamaa_2 'to start or be involved in some activity'

SENSE_ID naamaa_2

SENSE 'to start or be involved in some activity'

VERB_CLASS 13 [k1,k7)

VERBS_IN_THE_SAME_CLASS jamaa_1, khaaTaa_3, aasaa_5, milaa_2

MORPHOSYNTACTIC_TYPE intransitive

EXAMPLE_SENTENCE aami byabsaa-y naam-chi

I business-LOC be involved-PROG

I am going to be involved in a business

VERB_FRAME_1

DEPENDENCY_RELATION	k1	k7
LEXICAL_TYPE	N	N
DEPENDENCY_MARKER	0	e/te
ONTOLOGY	+human	+event

EXAMPLE_SENTENCE police tadanta kar-te neme-che

police investigation do-INF be involved-PRF

Police is doing the investigation

VERB_FRAME_2

DEPENDENCY_RELATION	k1	k7
LEXICAL_TYPE	N	INF
DEPENDENCY_MARKER	0	e/te
ONTOLOGY	+human	x

naamaa_3 is the last sense. It means the reduction of value or quantity. This verb always takes abstract agent.

***naamaa_3* ‘to reduce or decrease’**

SENSE_ID	naamaa_3
SENSE	‘to reduce or decrease’
VERB_CLASS	1 [k1]
VERBS_IN_THE_SAME_CLASS	calaa_1, ghumono_1, baaRaa_1, Daakaa_2, kaaTaa_4, phoTaa_2
MORPHOSYNTACTIC_TYPE	intransitive
EXAMPLE_SENTENCE	aaj sakaalei er jvar nem-eche today morning fever reduce-PRF (His/her) Fever has come down today morning
VERB_FRAME	
DEPENDENCY_RELATION	k1
LEXICAL_TYPE	N
DEPENDENCY_MARKER	0
ONTOLOGY	+abs

5.5.4 saaraa

In Bengali Verb Frame Network, the verb *saaraa* has two senses: *saaraa_1* ‘to recover’ and *saaraa_2* ‘to do something’. Both semantic and syntactic distinctions between two senses is very clear. The first one is an intransitive verb. It has a unique dependency structure [k1,rt]. The second is a transitive verb.

***saaraa_1* ‘to recover’**

SENSE_ID saaraa_1

SENSE ‘to recover’

VERB_CLASS 12 [k1,rt]

VERBS_IN_THE_SAME_CLASS khaaTaa_2, jvalaa_2, ph.Naasaa_2

MORPHOSYNTACTIC_TYPE intransitive

EXAMPLE_SENTENCE saarjaari-te ser-eche pragya-r stan kyansaar
 surgery-LOC recover-PRF Pragya-GEN breast cancer
 Pragma’s breast cancer has been cured by surgery

VERB_FRAME

DEPENDENCY_RELATION k1 rt

LEXICAL_TYPE N N

DEPENDENCY_MARKER 0 e/te

ONTOLOGY +abs +event

***saaraa_2* ‘to do/execute something’**

SENSE_ID saaraa_2
SENSE ‘to do or execute something’
VERB_CLASS 2 [k1,k2]
VERBS_IN_THE_SAME_CLASS bhalobaasaa_1, cenaa_1, kholaa_2, balaa_2, kaaTaa_3
MORPHOSYNTACTIC_TYPE transitive

EXAMPLE_SENTENCE samprati miusik bhidio-r shuTiM ser-echen kanikaa
 recently music video-GEN shooting do-PRF Kanika
 Kanika has done shooting of a music video

VERB_FRAME_
DEPENDENCY_RELATION k1 k2
LEXICAL_TYPE N N
DEPENDENCY_MARKER 0 0
ONTOLOGY +human +event

5.6 Verbs with Multiple Verb-Frames: Theory and Statistics

Frame associated with a verb is a knowledge-level abstraction. Language users may not be aware of the frame as it is perceivable only by analysing the language data. We have noted in the introductory chapter that verb-frame is the concrete realisation of the underlying event semantics. Therefore, the same verb occurring in different contextual frames reflects the difference in the underlying event structure. Most of the time, the event structure formation correlates with a particular sense of a verb. But sometimes, a single verb sense can express two or more unique event structures, thus having more than one verb-frame. In fact, there is nothing in the theory that prevents a verb sense from demonstrating more than one verb-frame.

Verbs can occur with many different structures. These different structures can have different configurations of argument structures. BVFN is developed in order to seek the generalisation among those different argument structures and present it in terms of dependency structure, lexical type, dependency marker and ontology. Usually, in Bengali Verb Frame Network, a verb has one frame that perfectly represents the range of the argument structures that the verb can occur with. But verbs can have more than one verb-frame too. Our experience in this regard is more similar to that of the Hindi verb-frame in which vast majority of verbs have only one frame, but small minority has more than one (Begum, 2017b, chapter 6, p 18). Here, in case of BVFN too, some verbs show more versatility than others in that they display two or more configurations of ‘Dependency Roles’ or ‘Lexical Type’. These are not extended verb-frames and should not be confused with them. Multiple verb frames of a verb cannot be narrowed down into a single one as they represent completely different dependency structures or, in some cases, different argument structure compositions in terms of lexical type. However, the number of such verbs in Bengali Verb-frame Network is not very large. Little more than one-sixth of the total 300 verbs carry multiple verb frames. 52 verbs have more than one verb-frame. The highest number of verb frames that a verb has is 4. Only one verb *caaoyaa_1* ‘to want’ has four verb-frames. 43 verbs have two verb-frames, and 8 verbs have three verb frames. In contrast, 248 verbs out of 300 have verb-frame files populated with a single verb-frame.

Comparison: Verbs with Single and Multiple Verb Frames

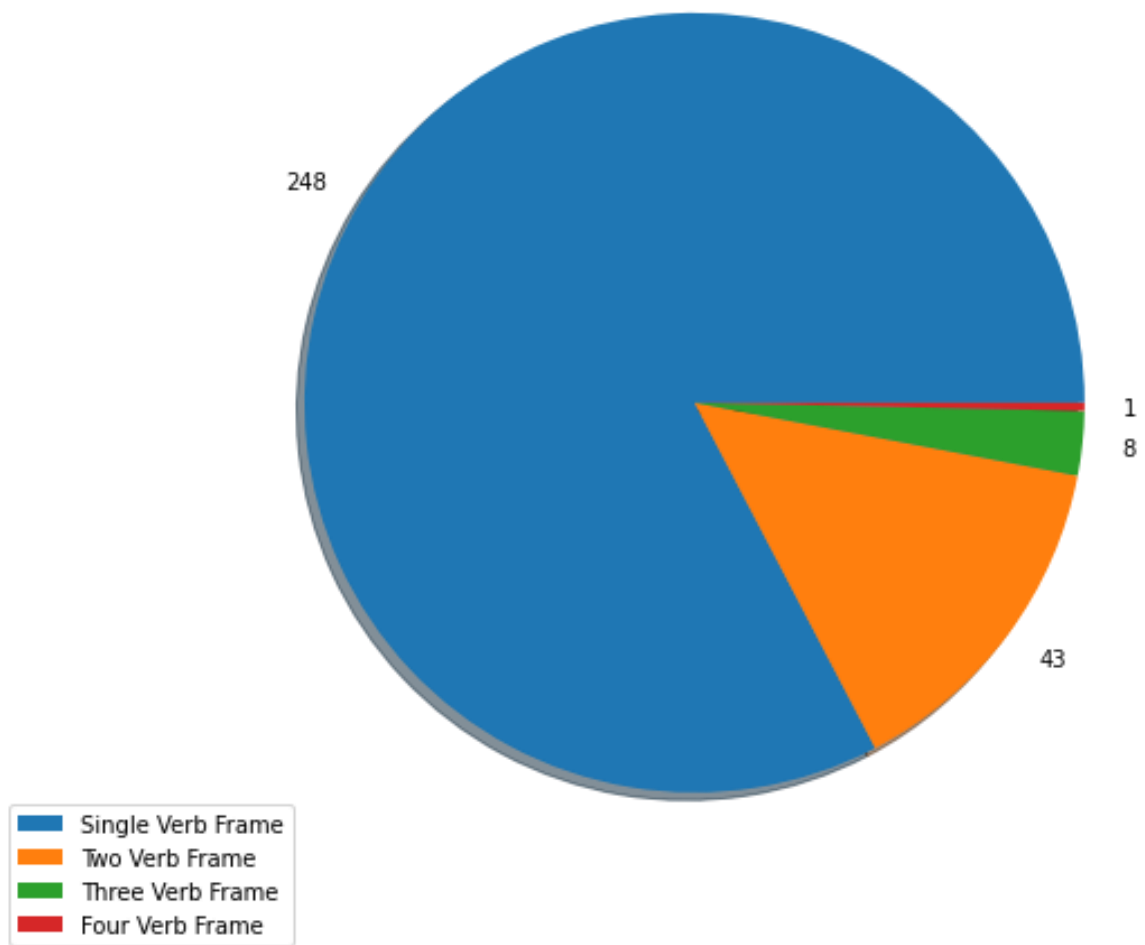


Figure 5.3: Verbs with Multiple Verb-frames (Pie Chart)

5.7 Verbs and Lexical Types

In this section, the observations regarding the lexical category will be enumerated.

5.7.1 Verbs and the Clausal Argument

Verb of cognition and communication like *bhaabaa_1* 'to think', *balaa_1* 'to tell', *dekhaa_1* 'to see', *dekhaa_3* 'to check', *dekhaa_4* 'to monitor', *shonaa_2* 'to get to know', *bojhaa_1* 'to understand', *jaanaa_1* 'to know' etc. are the only classes of verbs which are

licensed for clausal objects. The clause here refers to the complementiser clause structure like the English 'that' clause or the Hindi 'ki' clause. These verbs fall into two distinct categories.

One is communication verbs like *balaa_1* 'to tell', *lekhaa_1* 'to write', *paRaa_1* 'to read'.

Another category is the verbs of cognition like *dekhaa_1* 'to see', *dekhaa_3* 'to check', *dekhaa_4* 'to monitor', *shonaa_2* 'to get to know', *bojhaa_1* 'to understand', *jaanaa_1* 'to know' etc. take clausal arguments. These verbs typically have [k1,k2] structure in which the k2 can either be a nominal or a clause.

Di-transitive verbs like *balaa_1* have three arguments in their dependency structure [k1,k4, k2].

Here also, the object or k2 can either be a nominal or a clause.

***bojhaa_1* ‘to understand’**

SENSE_ID	bojhaa_1
SENSE	‘to understand’
VERB_CLASS	2[k1,k2]
VERBS_IN_THE_SAME_CLASS	saaraa_2, bhalobaasaa_1, cenaa_1, kholaa_2, balaa_2, kaaTaa_3
MORPHOSYNTACTIC_TYPE	transitive

EXAMPLE_SENTENCE	<i>siimaa prashnaTaa bojh-e ni</i>
	Sima query understand-3 NEG.PRF
	Sima has not understood the query

VERB_FRAME_1

DEPENDENCY_RELATION	k1	k2
LEXICAL_TYPE	N	N
DEPENDENCY_MARKER	0	0
ONTOLOGY	+human	+abs

EXAMPLE_SENTENCE	<i>aami bujh-echi je eibhaabe habe naa</i>
	I understand-3.PRF that in_this_way be.FUT NEG
	I have understood that it would not work this way

VERB_FRAME_2

DEPENDENCY_RELATION	k1	k2
LEXICAL_TYPE	N	CL
DEPENDENCY_MARKER	0	0
ONTOLOGY	+human	x

5.7.2 Verbs and the Infinitival Argument

The direct object can be an infinitive structure too. Few verbs that suggest ability, knowledge, skill, need, longing, crave, restraint etc., can occur with such structure. These verbs

are all abstract verbs, and they represent a mental state in relation to something external. It can be an acquisition of ability or liking of something. The Bengali ability verb *paaraa_1* ‘to be able to’ is a typical representative of the group.

***paaraa_1* ‘to be able to’**

SENSE_ID	paaraa_1
SENSE	‘to be able to’
VERB_CLASS	2[k1,k2]
VERBS_IN_THE_SAME_CLASS	jaanaa_1, bojhaa_1, saaraa_2, bhalobaasaa_1, cenaa_1, kholaa_2, balaa_2, kaaTaa_3
MORPHOSYNTACTIC_TYPE	transitive

EXAMPLE_SENTENCE	o naach paar-e
	S/he dance can-3.PRS
	S/he can dance

VERB_FRAME		
DEPENDENCY_RELATION	k1	k2
LEXICAL_TYPE	N	N
DEPENDENCY_MARKER	0	0
ONTOLOGY	+anim	+event

EXAMPLE_SENTENCE	sab kichui raannaa kar-te paa-ren shaamim
	every thing cooking do-PTCP can-3.PRS Shamim
	Shamim can cook everything

VERB_FRAME_2		
DEPENDENCY_RELATION	k1	k2
LEXICAL_TYPE	N	INF
DEPENDENCY_MARKER	0	0
ONTOLOGY	+anim	x

5.7.3 Verbs and the Adverbial Argument

Dependency relation ‘rd’ (direction) is usually represented by adverbs at the lexical level. Directional verbs like *pheraa_2*, *jaaoyaa_4* can take an adverb as their second argument. Adverbs do not occur in object position. It rarely represents other dependency relations as well.

***pheraa_2* ‘to turn around’**

SENSE_ID	pheraa_2		
SENSE	‘to turn around’		
VERB_CLASS	15[k1,rd]		
VERBS_IN_THE_SAME_CLASS	jaaoyaa_4		
MORPHOSYNTACTIC_TYPE	intransitive		
EXAMPLE_SENTENCE	ami	pichan	phir-laam
	I	behind	turn around-1.PRF
	I turned back		
VERB_FRAME			
DEPENDENCY_RELATIONS	k1	rd	
LEXICAL_TYPE	N	ADV	
DEPENDENCY_MARKER	0	0	
ONTOLOGY	+anim	x	

5.7.4 Verbs and the Adjectival Argument

Adjectives have the potential to occur as k2 in case of some verbs but mostly represent dependency relation k1s. The below verb-frame exemplifies adjective as an object.

***bakaa_1* ‘to speak incoherently’**

SENSE_ID bakaa_1
SENSE ‘to speak incoherently’
VERB_CLASS 2 [k1,k2]
VERBS_IN_THE_SAME_CLASS saaraa_2, kholaa_2, kholaa_4, bhangaa_2
MORPHOSYNTACTIC_TYPE transitive

EXAMPLE_SENTENCE lokTaa bhulbhal bak-chila
 the man nonsense speak-PST.PROG
 The man was speaking nonsense

VERB_FRAME
DEPENDENCY_RELATION k1 k2
LEXICAL_TYPE N ADV
DEPENDENCY_MARKER 0 0
ONTOLOGY +human x

Most occurrence of ‘ADV’ in verb-frames has been due to the dependency relation k1s, i.e., the subject complement. Bengali copula verb *hoyaa_1* ‘to be’ and copula-like *banaa_2* ‘to become’ is endowed with such [k1,k1s] dependency structure. The second argument of these verbs is usually an adjective. In case of the verb *haoyaa_1*, both noun and adjective can serve the role of subject complement. In case of the second verb, it has to be an adjective.

5.8 Frequency count of Dependency Relations

Dependency relations are the most prominent part of the verb-frame. k1, or the sentential subject of a canonical sentence, is the most primary relation among all. In Bengali, subject can be omitted from a sentence’s surface structure but can still be inferred from the subject-verb agreement. k1 occurs in 327 verb-frames out of 362. The next frequent

dependency role is k2 which is an obligatory role for all transitive verbs. It occurs 209 times in BVFN. Apart from k1 and k2, most frequent dependency relations are in the order of: k7p, k7, k5, pk1, jk1, k4, k1im, rt in descending order. k7p emerges as the single most important dependency relation among them by having occurred 55 times among 362 verb frames. k7 stands very near as it occurs 41 times. It suggests, as a concept, location is very important in verb semantics. Not only it occurs in the verb-frames of the locational verbs, but k7p also marks the second argument of a large number of intransitive verbs like Dobaa_1 ‘to drawn’, calaa_3 ‘to run a transport’ etc.

Dependency Roles	Frequency
k1	327
cok1	10
k1s	7
pk1	19
jk1	17
mk1	2
k1im	14
k2	209
k2p	7
k2s	9
k3	11
k4	15
k5	20
k5a	6
k7p	55

k7t	6
k7	41
rt	12
rd	9
rkc	4

Table 5.3: Frequency of Dependency Relation in BVFN

5.9 Verb-frames and Ontology

Ontological specifications are essential conditions that determine the nature of the arguments and impose selectional restrictions on them. By doing so, it helps in resolving lexical ambiguity. Such an example can be found in the verb-frames of *kaaTaa_4* and *kaaTaa_6* which is presented in section 5.5. Another example is *calaa_1* and *calaa_4* where ontological feature is the only distinction between two verb-frames.

***calaa_1* ‘to walk’**

SENSE_ID calaa_1

SENSE 'to walk'

VERB_CLASS 1[k1]

VERBS_IN_THE_SAME_CLASS baaRaa_1, calaa_6, Daakaa_2, Dobaa_2, Dobaa_3, jaaoyaa_5,
jamaa_2, jvalaa_3, kaaTaa_4

MORPHOSYNTACTIC_TYPE intransitive

EXAMPLE_SENTENCE	baccha	talma	karte-karte	cal-che
	child	staggering		walk-3.PROG
	The child is staggering			

VERB FRAME

DEPENDENCY RELATION k1

LEXICAL TYPE	N
--------------	---

DEPENDENCY MARKER 0

ONTOLOGY +anim

***calaa_4* ‘to have a particular time’**

SENSE_ID	calaa_4
SENSE	‘to have a particular time’
VERB_CLASS	1 [k1]
VERBS_IN_THE_SAME_CLASS	baaRaa_1, calaa_1, calaa_6, Daakaa_2, Dobaa_2, Dobaa_3, jaaoyaa_5, jamaa_2, jvalaa_3, kaaTaa_4
MORPHOSYNTACTIC_TYPE	intransitive
EXAMPLE_SENTENCE	kriSh~napakSha cal-che dark fortnight be-PROG Now, it is dark fortnight
VERB_FRAME	
DEPENDENCY_RELATION	k1
LEXICAL_TYPE	N
DEPENDENCY_MARKER	0
ONTOLOGY	+time

Ontology plays a pivotal role in verb classification also. ‘+human’ is the most frequent ontology feature in BVFN which is most frequently associated with k1. It occurs 220 times among 362 verb-frames. Sometimes more than one argument of a verb-frame is generalised with the [+human] feature. In the vast majority of verb-frames, the karta or the logical subject of a verb can either be an animate ‘+anim’ (animate includes +human feature) or an abstract ‘+abs’ entity. Except in one case where k7p represents an object ‘+obj’, the dependency relation k7p always represents region ‘+reg’. The same is true for the dependency relation k2p. Only difference is there is not even one exception. These are the general features throughout the verb-frame. There are also other correlations that apply to certain verb groups.

One of the observations pertaining to ontology is that the distinctive ontological pattern of arguments of different verb classes differs from each other, owing to the underlying event

structure that comes from the verb semantics. Volitional physical activity verbs like *caapaa_1* ‘to press’ has the ontology structure of ‘+anim +obj’. These are transitive verbs whose agent is always animate, and the second argument or the theme is an object. The ‘press’ type verbs which follow this pattern are given below.

Verb	Meaning	1st Argument	2nd Argument
<i>Thelaa_2</i>	‘to push forward’	+anim	+obj
<i>Taanaa_3</i>	‘to pull’	+anim	+obj
<i>caapaa_1</i>	‘to press’	+anim	+obj
<i>dharaa_1</i>	‘to catch’	+anim	+obj
<i>tola_1</i>	‘to lift’	+anim	+obj
<i>tola_7</i>	‘to pluck’	+anim	+obj

Ontological patterns can be observed in other verb classes as well. Verbs of instrumental action like *kaaTaa_1* ‘to cut’ or *baa.Ndhaa_1* ‘to bind’ have a remarkably consistent pattern at all sections of verb-frame. These verbs typically have three arguments, i.e., agent, theme and instrument. The ontological pattern observed among them is ‘+human +obj +art’. Thus, ontological properties of verbal arguments are not only generalisable for a single verb as it has been done in Bengali Verb Frame Network, but also these properties can be extended up to the level of verb class.

5.10. Some Special Class of Verbs

In this section, we will look into certain group of verbs through the prism of verb-frame approach. These verb groups⁴ The verb groups are special in the sense that they occur in structures that are theoretically complicated to define and describe, as well as interesting in terms of their event structure. The objective here is to show that the verb-frame mechanism is explanatorily adequate even for such complex structures.

5.10.1 Causative Verbs

Causative verbs imply a meaning which can be roughly translated as cause to do something. Causative predicates take at least two arguments – a CAUSER and a CAUSEE. Causer makes causee to do the action denoted by the verb, and causee is the argument who actually accomplishes the action. “....an actor makes somebody else do something or causes him to be in a certain state. Expressions that indicate this phenomenon is called causatives.” (Agnihotri, 2007, p. 120). For example, the English verb ‘feed’ implies the causation of eating something. Hence, ‘feed’ is the causative counterpart of ‘eat’. Since the causativisation process adds more argument, Comrie (1985) sees causativisation as a ‘valency increasing mechanism’. Causativisation is a very regular process in any language. However, languages employ different ways to signal causativisation. There are mainly three ways in which causativisation is used in world languages. Lexical causatives are very common in world languages. **Lexical causatives** are just like any other transitive verb; only difference is they convey a causative sense in relation to some other verb in the language. English ‘feed’ is an example of lexical causative. Many languages use **periphrastic causatives**, which include two verbs and two clauses. English typically uses ‘make’ (and some other verbs) in the main clause along with a lexical verb in the subordinate clause as an in-

‘I make him swim’.

⁴ Causative, impersonal and human interaction verbs are part of verb classification that we have done on the basis of similarity of verb-frames. The verb classification is described in the next section.

In the above sentence, ‘make’ expresses causation, while the verb ‘swim’ expresses the main action. Causativisation can also be expressed through verbal morphology. Indian languages attach affixes with the main verb to denote causativisation.

The causativisation process in Bengali, just like other Indian languages, is morphological, i.e., causative verbs are formed from the base transitive or intransitive verbs by some morphological process. Unlike Hindi, Bengali does not have two levels of causatives. There is only one causative form for most of the Bengali verbs. A causative verb is marked with the *-no* suffix at the end of the base verb. For example,

Base Verb	Causative Verb
<i>dekhaa</i> ‘to see’ (<i>dekhaa_1</i> in BVFN)	<i>dekhaa-no</i> ‘to show’
<i>Dobaa</i> ‘to sink’ (<i>Dobaa_1</i> in BVFN)	<i>Dobaa-no</i> ‘to make (something) sink’
<i>khaaoyaa</i> ‘to eat’ (<i>khaaoyaa_1</i> in BVFN)	<i>khaoyaa-no</i> ‘to feed’

Therefore, causative verbs in Bengali are marked by *-no* suffix in its dictionary form (which is the verbal noun form of any verb in Bengali; see section 5.1 for details). In Bengali, the dictionary form is not the root form. Causativisation changes root form too. It attaches the *-aa* suffix to the root form of the base verb. Another fact that needs to be asserted here is that all verbs with *-no* suffix are not causative but causative verbs inevitably have *-no* suffix. The verbs which have *-no* suffix in their base form cannot be causativised by regular morphological processes. Most such verbs are intransitive and represent either a bodily action or a natural phenomenon.

The process of causativisation increases verb valency. It invariably adds an extra argument. An intransitive verb, when causativised, is able to take two arguments which is the minimum valency for causative verbs. Similarly, when a transitive verb is causativised, it

takes three arguments. The members of verb class 5 (see section 5.11), which is given below, are examples of such verbs. Di-transitive verbs, on the other hand, have valency 3 in their base form. After causativisation, these verbs can take four arguments.

An intransitive verb, when causativised, at least has a causer (pk1) and one causee(jk1) in its argument structure. It can have other arguments like k7p, depending on the semantics of the verb.

Ex – 5.2. *noukaaTaa (k1) bhaash-lo*
 The boat float-PRF
 The boat has floated.

Ex – 5.3. *ami (pk1) kagoj-er nouka (jk1) bhaash-aa-lam*
 I paper-GEN boat float -CAUS-1.PRF
 I made paper boat float.

Similarly, when a transitive verb [k1,k2] is causativised, it has the default [pk1,jk1,k2] dependency structure.

Ex – 5.4. *shishuTi (k1) dudh(k2) khaa-cche*
 The child milk drink-PROG
 The child is drinking the milk.

Ex – 5.5. *maa (pk1) shishuTi-ke (jk1) dudh (k2) khaa-oyaa-cchen*
 mother the child-ACC milk drink-CAUS-PROG
 Mother feeds milk to the child.

In case of a di-transitive verb, causativisation adds one more argument. But here, the *karta* of the base di-transitive verb became mediator causer mk1 and is followed by *diye* post-position.

Ex – 5.6. *aami (k1) somaa-ke (k4) ciThi (k2) likh-laam*
 I soma-ACC letter write-1.PRF
 I have written a letter to Soma.

Ex – 5.7
uni (pk1) aamaa-ke (mk1) diye somaa-ke (jk1) cithi (k2) lekh-aa-len
 s/he I-ACC post-position soma-ACC letter write-CAUS-PRF
 S/he made me write a letter to Soma.

One or more mediator causers or mk1 can be added in sentences 5.3 and 5.5 as well. But it is not common in Bengali to have a mediator causer in intransitive-causative or transitive-causative sentences.

A total of 6 verb classes in BVFN have been identified, which involves causative verbs.

Verb Class	Base Verb	Dependency Structure
4	Intransitive	[pk1,jk1]
5	Transitive	[pk1,jk1,k2]
30	Transitive	[pk1,mk1,k2]
31	Di-transitive	[pk1,mk1,jk1,k2]
32	Transitive	[pk1,mk1,rd,k2]
33	Intransitive	[pk1,jk1,k7p]

The complete verb-frame of a causative verb *baanaano_1* is given below.

***baanaano_1* ‘to change into’**

SENSE_ID baanaano_1
 SENSE ‘to change into’
 VERB_CLASS 5 [pk1,jk1,k2]
 VERBS_IN_THE_SAME_CLASS dekhaano_1, dekhaano_3, dekhaano_5, khaaoyaano_1,
 MORPHOSYNTACTIC_TYPE Causative
 REMARK causative of banaa_4

EXAMPLE_SENTENCE jaadukar rumaal-ke phul baan-aa-len
 the magician handkerchief-ACC flower be-CAUS-PRF
 The magician has made the handkerchief into a flower.

VERB_FRAME

DEPENDENCY_RELATION	pk1	jk1	k2
LEXICAL_TYPE	N	N	N
DEPENDENCY_MARKER	0	ke	0
ONTOLOGY	+human	+obj	+obj

5.10.2 Impersonal Verbs

An impersonal verb is a verb that does not have a true subject. In the case of Bengali, Dasgupta defines “An impersonal verb is one that does not agree and therefore appears in the unmarked form – third Person, neutral Formality” (Dasgupta, 1982, p. 1). In Bengali, the nominal subject does not take any vibhakti or affixes. But there is another kind of subject construction in which genitive Vibhakti is attached to the subject. This kind of subject construction is a very common structure in Bengali. They are markedly different not only in terms of structure but also there is a clear semantic distinction.

Bengali impersonal subject has two main properties i) It takes the genitive suffix -*r/er* ii) There is no subject-verb agreement. Verb appears in its unmarked form, i.e., third person, neutral form. The below examples illustrate the point

Ex – 5.8

(a) PERSONAL SUBJECT

tumi anek bhul dhaaraNaa kar-echo
 you much wrong idea do-2.PRF
 You have got much wrong ideas

(b) IMPERSONAL SUBJECT

tomaa-r anek bhul dhaaraNaa hay-eche
 you-GEN much wrong idea be-3.PRS.PRF
 You have got much wrong ideas

Ex – 5.9

(a) PERSONAL SUBJECT

aami biye kar-echi
 I marriage do-1.PRF
 I got married

(b) IMPERSONAL SUBJECT

aamaa-r biye hay-eche
 I-GEN marriage be-3.PRF
 I got married.

Ex – 5.10.

a) PERSONAL SUBJECT

raam baaRi jaa-be
 Ram home go-FUT
 Ram will go home

b) Impersonal SUBJECT

Raam-er baaRi jaa-o-yaa habe

Ram-GEN home go-PTCP be.AUX.FUT

Barring the structure, the semantic distinction of the impersonal sentence is equally striking. Sentence 5.10(b) does not have a good English translation. The closest we get – ‘Ram’s going home is possible’. This sentence is also an example of much-discussed Bengali passive construction. Here the auxiliary verb *habe* bears tense and aspect inflection. On the other hand, sentence 5.8(b) and 5.9(b) cannot be termed passive. The situation is quite self-explanatory. In Bengali, the impersonal structure and passive cannot be categorised separately because their usage and function are similar. Rather, it will be argued here that Bengali passive should be treated as a sub-category of impersonals. Observations from the corpus suggests that a Bengali impersonal subject appears whenever the speaker wants to centre the event or the object in his/her utterance, instead of the agent/logical subject.

5.10.2.1 Impersonal Construction in Literature: -

Impersonal construction is attested in many languages. The construction is defined by the invariable third-person agreement on the verb and the noun phrase that can be regarded as the subject stands in a non-nominative case (Trask, 1993). Along with the general impersonal construction, there is an impersonal passive found in many languages. This is a group of sentence constructions characterised by the involvement of passive inflection on the verb and no lexical subject. There is a lot of controversy over the extent of the term. It includes problematic cases like French *On vendit fa maison* literally ‘One sold the house’ which is different from the canonical active transitive sentence only in the employment of the impersonal subject ‘on’.

Impersonal subject construction in Bengali has long been studied by many scholars. Scholars refer to such subjects by many terms like dative subject, impersonal subject, experiencer subject, impersonal construction, genitive subject, passive subject, impersonal passive and so on (Ray, Hai and Ray, 1966; Chatterjee, 1968; Klaiman, 1980; Klaiman, 1981; Dasgupta, 1982; Trask, 1993; Thompson, 2003). M.H. Klaiman's work is the first thorough attempt to study this phenomenon. Attention to this kind of subject construction has come primarily due to the enigma of Bengali passive structures. Bengali passive seems to be very different from the passive construction observed even in the related other Indian languages. Punya Sloka Ray et al. (1966) even rejected the existence of passives in Bengali "A beginner's difficulty is that Bengali has no passive voice as idiomatic usage, though it has a somewhat comparable impersonal clause type." (Ray, Hai and Ray, 1966). This is a very interesting stand. It seems to have been influenced by the fact that Bengali passive sentences do not usually have their active counterpart. Active and passive sentences convey different meanings. Hence, Bengali passives are not passive if one thinks in terms of the English passivisation process. Ray (1966) is of the opinion that Bengali has an impersonal clause structure in place of passive structure. Dasgupta (1982), on the other hand, views the passive structure as a special case in Bengali, the impersonal subject construction as another special case; he treats both as two representations of a more general impersonal construction.

Klaiman's emphasis that impersonal verbs represent non-volitional action seems to be rightly placed. This is indeed the case. The impersonal subject is marked as k1im in BVFN. Impersonal verbs belong to the verb class 16 which has the unique dependency structure [k1im,k2].

***laagaa_7* ‘to feel’**

SENSE_ID laagaa_7

SENSE ‘to feel’ (usually temperature, pain etc)

VERB_CLASS 16[k1im,k2])

VERBS_IN_THE_SAME_CLASS paaoyaa_2, baa.Ndhaa_2

MORPHOSYNTACTIC_TYPE transitive

EXAMPLE_SENTENCE aamaa-r alpa garam lag-che

I-GEN little heat feel-PROG

I am feeling little heat

VERB_FRAME

DEPENDENCY_RELATION	k1im	k2
LEXICAL_TYPE	N	N
DEPENDENCY_MARKER	r/er	0
ONTOLOGY	+human	+abs

5.10.3 Verbs of Human Interaction

These verbs represent the various activities of social coordination among human agents. Coordinating verbs are another special class which is realised as verb class 11 (intransitive) and verb class 17 (transitive) in verb-frame. The semantic novelty of these verbs can only be understood if we look at the event structure of these verbs. The event manifested by these verbs involves at least two agents who are equal participants in the activity. Verbs like talk, marry, and fight are the most prominent examples of these verbs.

These verbs occur within general comitative structure which implies accompaniment. In Bengali, *saathe* ‘with’ post-position is used to indicate any sentence with a comitative meaning. The comitative construction is a very common structure in which almost all verbs,

except the feeling verbs and some bodily activity verbs, can participate. Following is an example of a general comitative sentence.

Ex – 5.11. *se* [k1] *baabaa-r(ras-k1)* *saathe* *baajaar-e (k7p)* *ge-la*
 S/he father-GEN with market-LOC go-PRF
 S/he has gone to market with father.

In the sentence above, there are three arguments: agent [k1], associative agent (ras-k1) and location (k7p). The comitative sentence always involves at least one associative agent who accompanies the main agent in the activity. The second agent is positioned in the event structure of the verb by its association with the main agent.

Verbs of coordination also occur within general comitative structure. The main difference is that these verbs' semantics necessitates that there is more than one agent and all agents are equal in the event structure of the verb. Here the semantic relation is not that of association as both agents contribute equally towards the accomplishment of the activity. The below example illustrates the point.

Ex – 5.12. *amar*[k1] *sushmitaa-r(cok1)* *saathe* *kathaa bal-che*
 Amar Susmita-GEN with talk-PROG
 Amar is talking to Susmita.

Here, the verb *kathaa balaa* represents the activity of 'talking' that requires at least two people and all agents are equal from the perspective of the activity. In Bengali, the verbs like *kathaa_balaa_1* 'to talk', *jamaa_4* 'to gossip', *jaaoyaa_3* 'to match', *shoyaa_2* 'to have sex' etc. are the verbs of coordination. These verbs represent an event where more than one agent needs to coordinate among themselves to accomplish the activity.

***kathaa_balaa_1* ‘to talk’**

SENSE_ID	kathaa balaa_1		
SENSE	'to talk'		
VERB_CLASS	11 [k1,cok1]		
VERBS_IN_THE_SAME_CLASS	jamaa_4, milaa_1, shoyaa_2		
MORPHOSYNTACTIC_TYPE	intransitive		
EXAMPLE_SENTENCE	premikaa-r saathe kathaa bal-che ektaa chele		
	girlfriend-GEN with talk-PROG a boy		
	The boy is talking with his girlfriend.		
VERB_FRAME			
DEPENDENCY_RELATION	k1	cok1	
LEXICAL_TYPE	N	N	
DEPENDENCY_MARKER	0	r/er+saathe	
ONTOLOGY	+human	+human	

In the verb-frame, the intransitive verbs of coordination have [k1,cok1] argument structure, while transitive verbs of coordination have [k1,cok1,k2] structure. The first agent in the sentence is marked with k1, and the other agent/agents are marked with cok1. The second agent/agents are termed as coordinating *karta* as s/he/they have to coordinate with the first agent, and the sense of coordination is semantically pre-determined.

5.10.4 Conjunct Verbs

Conjunct verbs are understood as noun + verb or adjective + verb sequences that act as a single predicate in a sentence. They are similar to that of single verbs as they semantically represent one concept and have argument structure just like the single verbs. Only a few high-

frequency verbs can occur in these structures. Foremost among them is the verb *karaa* ‘to do’. It is considered the default verbaliser in Bengali. This verb can be used to verbalise almost any noun/adjective unless there is another verb earmarked for this particular concept. Thus, most Bengali conjunct verb has ‘*karaa*’ preceded by a noun/adjective.

Conjunct verbs are frequently used in Bengali. Still, it was found during corpus analysis that conjunct verbs have remarkably less ambiguity than single verbs. These verbs are mostly transitive and fall into different semantic domains. A small number of conjunct verbs such as *kathaa balaa_1* ‘to talk’, *chaaR deoyaa_1* ‘to give discount’, *chaaR deoyaa_2* ‘to exempt’, *kaajee laagaa_1* ‘to be useful’, *bikri karaa_1* ‘to sell’ etc. have been included in the verb-frame network.

tairii karaa_1 ‘to make’

SENSE_ID	tairii karaa_1		
SENSE	'to make'		
VERB_CLASS	2[k1,k2]		
VERBS_IN_THE_SAME_CLASS	baaRaa_1, calaa_1, calaa_4, calaa_6, Daakaa_2, Dobaa_2,Dobaa_3,jaaoyaa_5, jamaa_2,jvalaa_3, kaaTaa_4, kaaTaa_6, maraa_1, naamaa_3, phoTaa_2, galaa_1, laagaa_1, ghumono_1, ghoraa_1		
MORPHOSYNTACTIC_TYPE	transitive		
EXAMPLE_SENTENCE	bharat baayoTek ei bhaaksin tairii kar-eche		
	Bharat Biotech this vaccine make-PRF		
	Bharat-Biotech has developed this vaccine		
VERB_FRAME			
DEPENDENCY_RELATION	k1	k2	
LEXICAL_TYPE	N	N	
DEPENDENCY_MARKER	0	0	
ONTOLOGY	+anim	-abs	

5.11 Verb Classification

This section aims to show that we can attain semantically coherent verb classification through the verb-frame mechanism. Verbs are grouped into verb classes on the basis of identical dependency structures. When necessary, verb classes are further analysed into sub-classes based on the lexical type and ontological feature information. This kind of verb classification will be a unique resource in itself and cement the assumption that semantics

indeed determine the syntactic behaviour of a verb, at least to a large extent. We have detected 32 verb classes in Bengali through verb-frame analysis.

The dependency structures of all the 32 verb classes are given in the ‘Table 3’ below.

Verb Class	Dependency Structure		Dependency Structure
1	V_{int}[k1]	18	V_{tr}[k1,k2,k7]
2	V_{tr}[k1,k2]	19	V_{int}[k1,vmod]
3	V_{di}[k1,k2,k4]	20	V_{tr}[k1,k2,vmod]
4	V_{caus}[pk1,jk1]	21	V_{tr}[k1,k2,k5]
5	V_{caus}[pk1,jk1,k2]	22	V_{int}[k1,k1s]
6	V_{tr}[k1,k2,k3]	23	V_{int}[k1,k5]
7	V_{tr}[k1,k2p]	24	V_{int}[k1,k5a,k7p]
8	V_{int}[k1,k7p]	25	V_{int}[k1,k7t]
9	V_{tr}[k1,k2,k2s]	26	V_{tr}[k1,k2,rd]
10	V_{tr}[k1,k2+k7p]	27	V_{int}[k1,rkc]
11	V_{int}[k1,cok1]	28	V_{int}[k1,k5prk]
12	V_{int}[k1,rt]	29	V_{caus}[pk1, mk1,k2]
13	V_{int}[k1,k7]	30	V_{caus}[pk1,mk1,jk1,k2]
14	V_{int}[k1,k5,k7p]	31	V_{caus}[pk1,mk1,rd,k2]

15	V_{int}[k1,rd]	32	V_{caus}[pk1,jk1,k7p]
16	V_{tr}[k1im,k2]		
17	V_{tr}[k1,k2,cok1]		

Table 5.4: Verb Classification

Some selected verb classes will be discussed below. Complete list of all verb classes along with their members is attached in Appendix 2.

5.11.1 Verb Class 1 [k1]

This group is characterised by the presence of only one argument which is the karta [k1]. Only intransitives belong to this group. The verbs that fit into this group are:

baaRaa_1 ‘to increase’, *calaa_1* ‘to walk’, *calaa_4* ‘to get a particular time’, *calaa_6* ‘to continue’, *Daakaa_2* ‘to make a noise (animals)’, *Dobaa_2* ‘to disappear beyond the horizon’, *Dobaa_3* ‘to be ruined’, *jaaoyaa_5* ‘to have lost’, *jamaa_2* ‘to have an enjoyable occasion’, *jvalaa_3* ‘to feel hot’, *kaaTaa_4* ‘to recover’, *kaaTaa_6* ‘to pass the time’, *maraa_1* ‘to die’, *naamaa_3* ‘to decrease’, *phoTaa_2* to appear’, *galaa_1* ‘to melt’, *laagaa_1* ‘to be in use’, *ghumono_1* to sleep’, *ghoraa_1* ‘to revolve’, *uthaa_5* ‘to advance’, *uthaa_6* ‘to grow (abstract phenomenon)’ etc.

Semantically, most verbs in this verb class fall into two subgroups: Sub-class 1A has an abstract subject, while 1B has an animate subject. If we look at the event structure of the verbs, verbs of 1B are physical activity and physical state verbs, whereas verbs of 1A primarily represent an abstract event.

5.11.1.1 Sub-Class 1A

Members of the Sub-class: - *baaRaa_1* ‘to increase’, *calaa_4* ‘to have a particular time’, *calaa_6* ‘to continue’, *Dobaa_3* ‘to be ruined’, *jaaoyaa_5* ‘to have lost’, *jamaa_2* ‘to have an enjoyable occasion’, *kaaTaa_4* ‘to recover’, *kaaTaa_6* ‘to pass the time’, *naamaa_3* ‘to decrease’, *uThaa_6* ‘to grow (abstract phenomenon)’, *metaa_2* ‘to meet expectations, desire, demand etc.’.

These are all traditionally known unaccusative verbs. Levin and Rappaport Hovav have shown that unaccusativity is a syntactic criterion and is not a ‘unified phenomenon’ on the semantic ground (Levin & Rappaport Hovav, 1995, p. 16). Since we are looking for semantic properties, we will not define this class of verbs basis on unaccusativity. Furthermore, unaccusative diagnostics yield very different results in the case of Indian languages (Srishti, 2011, pp. 53-86).

Features:

- i.** The verbs are characterised by the fact that only abstract entities can be the subject.
- ii.** The majority of them are change of state verbs. A few are event verbs and temporal verbs.
- iii.** These verbs do not allow agent. They usually reflect non-volitional action.
- iv.** Sentient beings cannot be the subject.

A typical verb-frame of this sub-class looks as follows:

VERB_FRAME

DEPENDENCY_RELATION	k1
LEXICAL_TYPE	N
DEPENDENCY_MARKER	0

5.11.1.2 Sub-Class 1B

Members of the Sub-class: - *calaa_1* ‘to walk’, *jvalaa_3* ‘to feel hot’, *maraa_1* ‘to die’, *uThaa_5* ‘to advance’, *ghumono_1* ‘to sleep’.

Both typical unaccusative and typical unergative verbs feature in this group. But unergatives numbers more than unaccusatives.

Features: -

- i. Animate karta characterises this group of verbs. Karta can be human, animal or both human and animal.
- ii. They usually represent volitional action. Though *maraa_1* ‘to die’ is an exception, it can be interpreted as volitional or non-volitional depending on context.

VERB_FRAME

DEPENDENCY_RELATION k1

LEXICAL_TYPE N

DEPENDENCY_MARKER 0

ONTOLOGY +anim

5.11.2 Verb Class 2 [k1,k2]

A large number of verbs belong to the verb class 2. It is a class of transitive verbs which have the simplest verb-frame among all transitive verbs, i.e., only two arguments: the subject and the direct object. We can further sub-classify the verbs based on the other elements in verb-frames. Such differentiation inevitably leads to semantically definable verb

groups, as we will see now. Thus, just like we have seen in verb class 1, the verb-frame mechanism enables us to form semantically meaningful verb classes.

5.11.2.1 Sub-class 2A

These are verbs of cognition and perception. They allow both a noun phrase and a clause to function as k2.

Members of the Sub-class: - *bhaabaa_1* ‘to think’, *dekhaa_1* ‘to see’, *dekhaa_3* ‘to check’, *dekhaa_4* ‘to monitor’, *shonaa_2* ‘to be aware of’, *bojhaa_1* ‘to understand’, *jaanaa_1* ‘to know’.

Features: -

- i. The subject of these verbs is thematically an agent. The agent has to be +human.
- ii. For most of these verbs, k2 is ontologically unrestricted.

VERB_FRAME_1

DEPENDENCY_RELATION	k1	k2
LEXICAL_TYPE	N	N
DEPENDENCY_MARKER	0	0
ONTOLOGY	+human	

VERB_FRAME_2

DEPENDENCY_RELATION	k1	k2
LEXICAL_TYPE	N	CL

DEPENDENCY_MARKER	0	0
-------------------	---	---

ONTOLOGY	+human	x
----------	--------	---

5.11.2.2 Sub-class 2B

Members of the Sub-class: - *Taanaa_1* ‘to attract (gravitational)’, *Taanaa_3* ‘to pull’, *Taanaa_6* ‘to draw’, *Thelaa_2* ‘to go forward through pushing’, *dharaa_1* ‘to catch’, *khaoyaa_1* ‘to eat’, *khaoyaa_2* ‘to drink’, *tolaa_7* ‘to pluck’ etc.

These verbs can be best described as the verbs of agentive motion. The agent of action or the *karta* uses some force on an object, and the object gets motion as a result. The agent can even be an object or an artefact. The source of the motion is external to the object.

Features: -

- i. k1 is usually a +animate noun.
- ii. The second argument k2 is marked + obj.
- iii. These verbs represent a volitional act done by an agent.

VERB_FRAME

DEPENDENCY_RELATION	k1	k2
LEXICAL_TYPE	N	N
DEPENDENCY_MARKER	0	0
ONTOLOGY	+anim	+obj

5.11.2.3 Sub-class 2C

This sub-class represents human activity in various social contexts. The activities like speech activity, mental activity, activity in a social occasion, activity involving some event etc, can be distinguished and the verbs can further be sub-classified on the basis of that. For example, the verbs like *Daakaa_1*, *bakaa_1*, *bakaa_2* will belong to the speech activity group. Here, we are not attempting to do more fine-grained analysis. We just want to note that various activities can be further sub-classified and analysed accordingly. Feelings verbs are not included within this class. Instead, they are the topic of the following sub-section.

Members of the Sub-class: - *basaa_3* ‘to engage in some activity’, *Daakaa_1* ‘to call’, *bakaa_1* ‘to speak incoherently’, *bakaa_2* ‘to rebuke’, *calaa_7* ‘to carry on’, *tolaa_2* ‘to bring a topic for discussion’, *uThaa_4* ‘to get promoted’, *cenaa_1* ‘to recognise’.

Features: -

- i. The verb represents the activities of conscious beings. Hence, the karta is always +human.
- ii. The verb *basaa_3* only allows infinitival structure as k2. In the case of other verbs, the object is nominal and is marked by the ontological feature +event.

The verb-frame prototype applicable for this verb sub-class is shown below.

VERB_FRAME

DEPENDENCY_RELATION	k1	k2
LEXICAL_TYPE	N	N
DEPENDENCY_MARKER	0	0
ONTOLOGY	+human	+event

Members of the Sub-class: - *bhalobaasaa_1* ‘to love someone’, *bhalobaasaa_2* ‘to have a fondness to’, *ghRRiNaa karaa_1* ‘to hate’ etc.

Some⁵ Feeling verbs belong to this sub-class. The subject is thematically the experiencer, and the activity is that of feeling something. k2, or the direct object is the entity to whom the feeling is directed.

Features: -

- i. Lexical type of the arguments is either N (nominal) or INF (infinitival structure).
- ii. The direct object is semantically a recipient of the emotion.
- iii. Though animals can have emotions, in the course of corpus analysis, we have not found any animal agent of these feelings’ verbs. Further analysis indicates that the language allows only a few feelings verbs like *pachanda karaa_1* ‘to like’ to have animal agents. Higher-level emotions are almost never ascribed to animals. One possible reason may be that there is no way of knowing what kind of emotion an animal feels. We can only speculate from its activity.
- iv. Thus, only +human entities can act as the k1 for these verbs. The feelings can be directed to any concrete entity. Hence, the k2 is -abs.

Proto-typical verb frame for these verbs is like the following:

⁵ Not all feelings verb have [k1,k2] dependency structure. There are even many intransitive feelings verbs.

VERB_FRAME_1

DEPENDENCY_RELATION	k1	k2
LEXICAL_TYPE	N	N
DEPENDENCY_MARKER	0	ke
ONTOLOGY	+human	-abs

VERB_FRAME_2

DEPENDENCY_RELATION	k1	k2
LEXICAL_TYPE	N	INF
DEPENDENCY_MARKER	0	ke
ONTOLOGY	+human	-abs

5.11.3 Verb Class 3 [k1,k2,k4]

This is the class of di-transitive verbs. Bengali has a very small number of these verbs. These verbs can be subdivided into two classes.

5.11.1.3.1 Verb Class 3A

This class consists of di-transitive communication verbs like *balaa_1* ‘to tell’ and *likha_1* ‘to write’.

Feature:

- i. All communication verbs are not a member of this class
- ii. These verbs typically allow three arguments: agent, theme and recipient.

- iii. Both the agent and recipient are always +human, and the theme is ontologically unrestricted.
- iv. Just like the 2A class discussed above, these verbs can also take clausal objects in place of nominal objects.

VERB_FRAME_1

DEPENDENCY_RELATIONS	k1	k2	k4
LEXICAL_TYPE	N	N	N
DEPENDENCY_MARKER	0	0	ke
ONTOLOGY	+human		+human

VERB_FRAME_2

DEPENDENCY_RELATIONS	k1	k2	k4
LEXICAL_TYPE	N	CL	N
DEPENDENCY_MARKER	0	0	ke
ONTOLOGY	+human	x	+human

5.11.3.2 Verb Class 3B

Verbs of transfer belong to this sub-class. The thematic arrangement of the arguments is identical to that of the previous sub-class. Features of this sub-class are given below.

Members of the Sub-class: - *deoyaa_1* ‘to pay’, *deoyaa_3* ‘to give’, *deoyaa_4* ‘to assign or impose’, *deoyaa_5* ‘to cause an emotion, sensation etc.’, *deoyaa_6* ‘to convey’, *paThaano_1* ‘to send’, *bikri karaa_1* ‘to sell’.

Features: -

- i. This sub-class of verbs differs from the previous sub-class in its inability to accommodate clausal objects.
- ii. Ontological properties are shared between the two sub-classes since both deal with similar arguments. One noticeable difference is that k2, or the direct object is not ontologically unrestricted. The theme of transfer is usually a +obj entity.

The generalised verb-frame of this sub-class is the following:

VERB_FRAME

DEPENDENCY_RELATION	k1	k2	k4
LEXICAL_TYPE	N	N	N
DEPENDENCY_MARKER	0	ke	0
ONTOLOGY	+human	+obj	+human

5.11.4 Verb Class 6 [k1,k2,k3]

Members of the verb class 6 can be called the ‘action with instrument verb’. The semantics of these verbs require an agent, an object and an instrument to act upon the object. Verb class 6 is semantically equivalent to English ‘cut type’ verbs.

Members of the verb class: - *bha~Ngaa_1* ‘to break’, *bharaa_1* ‘to fill’, *caapaa_1* ‘to press’, *kaaTaa_1* ‘to cut’, *baa.Ndha_1* ‘to tie up’, *kholaa_6* ‘to unlock and open’, *maaraa_2* ‘to beat’, *tolaa_1* ‘to lift’ etc.

Features: -

- i. This verb class is the most consistent compared to others. Verb frames show remarkable similarity across the verb class.
- ii. The agent is usually a human because humans are the ones who are supposed to use instruments, especially complex instruments.
- iii. The post-position *diye* or *dvaaraa*, in a sentence, marks the instrument or k3.

VERB_FRAME

DEPENDENCY_RELATION	k1	k2	k3
LEXICAL_TYPE	N	N	N
DEPENDENCY_MARKER	0	0	diye/dvaaraa
ONTOLOGY	+human	+obj	+art

5.11.5 Verb Class 7 [k1,k2p]

k2p is a type of k2 occurs with destination verbs like *aasaa_1* ‘to come’ and *jaaoyaa_1* ‘to go’. Semantically k2p is very similar to k7p and is marked with the same locative suffix *-e* or *-te* (see Chapter 4 for details). These verbs typically represent motion that ends at a destination point.

Members of the verb class: - *aasaa_1* ‘to come’, *calaa_2* ‘to go’, *jaaoyaa_1* ‘to go’, *pouchono_1* ‘to reach’ etc.

Features: -

- i. This is a small class of verbs with well-defined semantic characteristics. Therefore, in terms of verb-frame, these verbs are consistent and similar.
- ii. Syntactically the verbs of verb class 7 and verb class 8 are very similar, almost identical, because the argument, in both cases, is marked with the same Vibhakti. However, k7p refers to the location of an activity, while k2p refers to the culmination point of action.
- iii. It may look like that only animate entities can act as k1 here, just like many other verb classes. However, this presumption will be wrong. Corpus analysis suggests that k1 can be vehicles or other inanimate moving objects.

VERB_FRAME

DEPENDENCY_RELATION	k1	k2p
LEXICAL_TYPE	N	N
DEPENDENCY_MARKER	0	0/e/te
ONTOLOGY	-abs	+reg

5.11.6 Verb Classes 11 and 17 [k1,cok1]/[k1,k2,cok1])

Members of verb class: - *kathaa balaa_1* ‘to talk’, *laRaa_1* ‘to fight’, *khelaa_1* ‘to play’, *laagaa_3* ‘to occur or break out’, *gaRaa_2* ‘to develop’(human relations), *jaaoyaa_3* ‘to match’, *jamaa_4* ‘to have a chat’, *milaa_1* ‘to meet’, *shoyaa_2* ‘to have sex’

This is one interesting verb class. These verbs all represent human interactions, i.e., various kinds of interactions between humans in different social situations. Verbs class 11 is the group of intransitive human interaction verbs like *kathaa balaa_1* ‘to talk’, *laRaa_1* ‘to

fight’ *prem karaa_1* ‘to have a romantic relationship’ etc. Whereas verb class 17 represents transitive human interaction verbs like *khela_1* ‘to play’, *gaRaa_2* ‘to develop (human relations)’. The following facts should be understood regarding the human interaction verbs.

Features:

- i. cok1 is the coordinating *karta*.
- ii. Interaction verbs typically have two participants k1 and cok1, and both participants have an equal role to play with respect to the activity in question. Both are logical subjects if we see them from the event structure perspective.
- iii. In Bengali, *saathe* ‘with’ post-position is used to denote coordination.

The transitive human interaction verbs display the following verb frame.

VERB_FRAME

DEPENDENCY_RELATION	k1	cok1	k2
LEXICAL_TYPE	N	N	N
DEPENDENCY_MARKER	0	r/er+saathe	0
ONTOLOGY	+human	+human	

Here is the prototypical verb-frame for the intransitive human interaction verbs.

VERB_FRAME

DEPENDENCY_RELATION	k1	cok1
LEXICAL_TYPE	N	N
DEPENDENCY_MARKER	0	r/er+saathe
ONTOLOGY	+human	+human

5.11.7 Verb Class 8 [k1,k7p]

This class consists of intransitive verbs. These are motion verbs invariably attached to a location.

Members of the verb class: - *Dobaa_1* ‘to sink’, *aasaa_3* ‘to belong to (place)’, *aasaa_4* ‘to appear’, *basaa_1* ‘to sit’, *basaa_2* ‘to establish’, *calaa_3* ‘to run (transport)’, *Daa.NRaano_1* ‘to stand up’, *Daa.NRaano_2* ‘to contest in an election’, *dhaakkaa khaaoyaa_1* ‘to collide’, *ghoraa_3* ‘to visit’, *jamaa_3* ‘to be accumulated’, *jvalaa_1* ‘to burn’, *kaaTaa_3* ‘to dig’, *khaaTaa_1* ‘to toil’, *laagaa_4* ‘to be in touch of something’, *metaa_1* ‘to be completed’, *neoyaa_3* ‘to take shelter’, *phaa.Nsaa_1* ‘to be trapped’, *phoTaa_1* ‘to get pricked’, *thaakaa_4* ‘to have a quality’, *thaakaa_6* ‘to stay’, *thaamaa_1* ‘to stop’, *uthaa_10* ‘to board on a vehicle’.

Features: -

- i. Since these are locational verbs, k1 cannot be abstract because abstract karta does not need a location.
- ii. The second argument is +reg

VERB_FRAME

DEPENDENCY_RELATION	k1	k7p
LEXICAL_TYPE	N	N
DEPENDENCY_MARKER	0	e/te
ONTOLOGY	-abs	+reg

5.11.8 Verb Class 16 [k1im,k2]

Enigmatic impersonal verbs are still largely uncharted territory in Bengali grammar. In the present work, the impersonal subject is marked as 'k1im'. Impersonal verbs belong to the verb class 16.

Members of the verb class: - aasaa_2 'to be able to', aasaa_6 'to start to feel', baa.Ndhaa_2 'to get restricted or prevented (emotionally)', caaoyaa_2 'to need', calaa_5 'to be habituated', haoyaa_4 'to be in a state', haoyaa_5 'to happen', haoyaa_7 'to possess', kaaje laagaa_1 'to be useful', laagaa_7 'to feel', paaoyaa_2 'to feel', thaakaa_3 'to have an attribute'.

These verbs are not a homogeneous semantic class. But they have semantic commonalities. The common features of these verbs are like the following

Features:

- i. They seem to represent only non-volitional action (Klaiman, 1980).
- ii. In a sentence, the existence of an impersonal subject, i.e., the subject that does not agree with the verb, defines them as the impersonal verb.
- iii. The subject always appears in genitive form.
- iv. The impersonal sentence focuses on either the event or the theme. Agent or the logical subject is not the focus.

VERB_FRAME

DEPENDENCY_RELATION	k1im	k2
LEXICAL_TYPE	N	N/INF
DEPENDENCY_MARKER	r/er	0
ONTOLOGY	+human	

5.11.9 Verb Class 5 [pk1,jk1,k2]

Causative verbs belong to this group. Verbs belonging to this group take three arguments: the causer, causee and the object.

Members of the verb class: - *banaano_1* ‘make’, *dekhaano_1* ‘to show’, *dekhaano_3* ‘to bring someone for check or test’, *dekhaano_5* ‘to make someone experience’, *khaaoyano_1* ‘to feed’

These are causative verbs formed from the base transitive verbs like *banaa_1* ‘to be’, *basaa_1* ‘to sit’, *dekhaa_1* ‘to see’, *khaaoyaa_1* ‘to eat’ etc. Verbs take three arguments causer, causee and object.

Features:

- i) Causer is always +human in case of these verbs
- ii) The second argument causee generally manifests animate entities. But there are exceptions. Inanimate objects can be causee too. For example, the verb *baanaano_1* takes an inanimate object as its second argument.

5.11.10 Verb Class 30 [pki1,mk1,jk1,k2]

Verb class 30 represents the di-transitive causative verb, i.e., the causative verbs formed from the base di-transitive verbs. Since di-transitive verbs are very few in Bengali, this verb class is very small.

Members of the verb class: - *balaano_1* ‘to cause someone talk’, *aanaano_1* ‘to make someone bring’, *deoyaano_1* ‘to cause someone give’ etc.

Features:

- i) These verbs take a mediator causer which is tagged as mk1, along with the other arguments. The mk1 is followed by the post-position ‘*diye*’, an instrumental post-position in Bengali.
- ii) ‘jk1’ or the causee can be silent or non-represented in the surface structure of the sentence.
- iii) Both the pk1 and jk1 have to be human. The mediator causer mk1 can be a human or an instrument.

We have discussed in detail 11 verb classes in this section. Some of the verb classes are further sub-classified. Deep semantic generalisation shared by the members of those verb classes has been unveiled and enumerated. The verb classes were discussed and analysed to demonstrate that dependency-based verb classification is perfect mechanism for reaching a semantic verb classification. We can reach similar semantic generalisation if we analyse the other verb class classes as well. The semantic similarity is often clear enough to see even with just one glance. Sometimes, a deep investigation is required. The verb classification is complex and full of nuances. The complexity of the pattern is expected, considering the arbitrariness of language. But we are able to reach a structured pattern in the end.

5.11.11 Significance of Verb Classification

Levin’s classification of English verbs (Levin, 1993) is considered a seminal work in the verb semantics of English. Many well-known computational resources have later been developed based on Levin’s work. It indicates the importance of semantically well-defined and syntactically idiosyncratic verb classes. The present work has been attempted in a similar spirit.

One possible use of verb classes is the extrapolation of frames for new verbs, thus populating more verbs into the Bengali Verb Frame Network. The aforementioned strategy

has been taken in the development of the Hindi Verb Frame (Begum, 2017b). The strategy works on the basis of prior knowledge regarding the class membership of a verb. For example, we already know from the corpus observation and semantic commonality that the verb *pou.Nchono_1* ‘to reach’ belongs to the verb class 7 just like the verb *aasaa_1* ‘to come’ or *jaaoyaa_1* ‘to go’. The verb-frame for the verbs *aasaa_1* and *jaaoyaa_1* has been created and are included in the BVFN. But the verb-frame for the verb *pou.Nchono_1* has not been created. We can simply extrapolate the generalised verb-frame of the verb class 7 to populate the verb-frame for the verb *pou.Nchono_1*. In this way, we can populate an unknown number of new verbs in the Bengali Verb Frame Network. Here two things need to be understood. For this to work, i) generalised verb-frame for a verb class needs to be existed prior to the extrapolation. ii) A specific verb may slightly deviate from the generalised verb-frame with regard to ontology structure especially. Though we need not work from zero, we need to check the argument structure of the specific verb and make necessary changes in the generalised proto-typical verb-frame whenever we try to extrapolate it for a new verb.

Verb classification is based on the dependency structure of arguments. Dependency relations capture that level of semantics which is well connected with the verb’s syntactic behaviour and is reflected in the surface level of a sentence. There are probably many ways and approaches to have classified the verbs out of the database of the Bengali Verb Frame Network. The goal was to prove that meaningful verb classification can be attained through verb-frame analysis, which is also useful for many other purposes. This objective has been achieved and has been shown in this section.

5.12 Corpus Observations

During development, a large Bengali corpus representing both serious and casual language use has been utilised. Bengali Verb Frame Network has come into existence through

deep corpus analysis. In this section, general observations from the corpus analysis will be presented. Many of these observations are previously noted in the scholarly literature on Bengali. At the same time, the others may be regarded as new additions to knowledge.

1. Lexically unique verbs are more or less a closed class in Indian languages, including Bengali. One study has previously shown that the number of finite verbs is only 5% of the lexicon, against a whopping 62% of the nouns (Biswas, 1992, quoted in Dash, 2015, p 87). Dash (2015) has made an interesting comparative analysis which manifests that the frequency of verbs is much higher in the corpus than in the dictionary.

Parts-of-speech	Dictionary (1992)	Corpus (1995)
Nouns	55%	62%
Adjectives	34%	18%
Finite Verbs	5%	11%
Indeclinable	2%	1%
Adverbs	2%	4%
Pronouns	1%	2%
Non-finite Verbs	1%	2%

Table 5.5: Frequency of Different Lexical Types in Corpus and Dictionary

Dash's finding is in line with our observation. Bengali indeed uses a small number of verbs in many different contexts, resulting in higher lexical ambiguity among verbs.

2. The reason is that there is almost no morphological process of creating new lexical verbs to speak of. As a consequence, the number of lexically unique verbs is quite small if we compare it with a language like English. High degree of lexical

ambiguity in Bengali verbs (not complex predicates) can be explained as a result of a relative verb shortage in the language.

3. New verbs are usually created as conjunct verbs (N+V sequence). The most productive process is to use the verb *karaa* ‘to do’ after a noun in order to verbalise a nominal. We have observed much less ambiguity among the conjunct verbs and very little ambiguity among compound verbs. As a result, only a few conjunct verbs made their way into Bengali Verb Frame Network.
4. Double-object construction is not a commonplace phenomenon in Bengali, which also implies there are only a few di-transitive verbs in the language. Only verbs of transfer and verbs of communication participate in this kind of construction.
5. Bengali does not distinguish between direct and indirect objects. The only distinction Bengali has is between animate and inanimate objects. More specifically, Bengali treats human objects with a specific marker in contrast to non-human objects, sometimes including animals as well. Human entity (sometimes animal too) is marked with accusative marker *-ke*. On the other hand, no suffix is assigned to an inanimate object. The language employs ontological criteria to distinguish between objects instead of syntactic one.
6. Among the Bengali intransitive verbs, the unaccusatives are more numerous than the unergatives. Unaccusativity or unergativity, in itself, does not correlate well with other important semantic properties. Their importance is in the fact that many semantic features are more represented among unaccusatives than unergatives and vice-versa. Hence, the property of unaccusativity or unergativity can be a clue to many semantic features.

7. Unlike Hindi, there is a clear-cut distinction between *vibhaktis* (case affixes) and post-positions in Bengali. *Vibhaktis* are always attached to the nominals and post-positions are always used freely.
8. Post-positions are free forms in Bengali that usually have many functions. Marking dependency role is only one such function. They are not purely functional items like *vibhaktis*. Most post-positions are derived from the content word, thereby retaining lexical meaning even when acting as dependency markers.
9. Bengali verbs rarely have more than three essential arguments, as it is attested in verb-frames, and never have more than four. Only some di-transitive and causative verbs have four arguments.
10. Non-essential arguments are not considered for inclusion in verb-frames (see section 5.3.1). There is no grammatical or semantic restriction on the number of non-essential arguments. Still, corpus analysis suggests there are usually not more than four arguments in a clause, including essential as well as non-essential arguments.
11. The Bengali verb *karaa*, including all its conjugational forms comes out to be the most frequent verb (verb senses are not differentiated, it includes all occurrences of *karaa*). It occurs a whopping 100458 times in the corpus. It is also one of the three most frequent lexical items in the corpus. But most instances of the verb *karaa* are attached with conjunct verbs, as *karaa* is the most common light verb in Bengali. Next four most frequent verbs are *haoyaa*, *jaaoyaa*, *balaa*, *ache/thaakaa* in respective numerical order. An earlier study done by Niladri Shekhar Dash (2007) found that there are five verbal forms exist among the most frequent 20 words in the Bengali Text Corpus of five million words. The five verbal forms are *kare*, *hay*, *karaa*, *jaay* and *haye* in descending order. *karaa* and *haye* are two different non-

finite verbal forms. *kare* can both be finite verb or a non-finite verbal form (The form is homophonous. When it is used as a non-finite verbal form, the pronunciation would be different). *hay* and *jaay* are finite verbs. The form *haye* can be a non-finite form of the verb *haoyaa*, or it can be a post-position. Evidently, the result we have got is statistically consistent with this earlier study.

12. If we go by verb-sense, then *balaa_1* ‘to tell’ has the highest number of occurrences in corpus sentences, closely followed by *haoyaa_1* ‘to be’. The reason is the verb *karaa*, though the most frequent verb form, occurs within conjunct verb pairs most of the time. Its use as a single independent verb is rather insignificant compared to its total number of occurrences.

5.13 Final Remarks

Thus, a structured, multi-layered, interconnected, annotated bank of arguments has been developed in the form of verb-frames. One distinguishing feature is that it does not represent lexical entry verbs but rather abstract verb senses. Verb’s meaning is conceptualised as a spectrum. The semantic spectrum has then been segregated into verb-senses. This argument structure bank demonstrates the semantic potential and syntactic usage of verb senses. Another notable distinction is that the verb-frame is essentially made of two different but essentially complementary frameworks: Dependency based framework and ontology. The addition of ontology creates a feeding loop of supplementary information to the dependency-based framework. It provides certain other kinds of information which cannot be obtained from dependency analysis.

Bengali Verb Frame Network is developed as an application-neutral resource. Verb-frame captures the syntactic distribution in which a verb is most likely to be used and identifies the sense it will acquire in doing so. The verb-frames are saved in .txt files so that the

computation tools can use them. We also used many such tools and programs to analyse the data, get the statistics and display the visualisation of the data. We hope that the resource thus created and presented here will be helpful for future researchers.

CHAPTER 6

CONCLUSION AND FUTURE RESEARCH

This dissertation has so far provided an in-depth analysis of the development process, starting from the corpus development stage to the final evaluation of the Bengali Verb-frame Network. In this concluding chapter, we summarise the key contributions and future directions of the research. The chapter is organised as follows: section one describes the accomplishment of the research, section two evaluates what we have done in the earlier chapters, and section three suggests future research and possible application of this resource. Finally, section four concludes the chapter with a critical overview.

6.1 Accomplishment of the Research

As we conclude this thesis on the development of Bengali verb frames, we can say that the research has contributed to a deeper understanding of the Bengali language and its grammatical structures. This study has shed light on the complex system of verb frames in Bengali and the various factors that influence their usage. It has also highlighted the role of lexical and syntactic factors in determining the choice of verb-frame, including the semantic properties of the verb, the argument structure of the sentence, and the syntactic context in which the verb appears. Through careful analysis and interpretation of these factors, we have developed a comprehensive model of Bengali verb-frames that captures the richness and complexity of this aspect of the language. The research on Bengali verb frames has important implications for our understanding of language structure and processing. By revealing the intricate patterns of verb-frame usage in Bengali, it contributes to a broader understanding of how languages encode and convey meaning. This knowledge can be useful for language learners, linguists, and natural language processing researchers alike, as it provides a detailed

roadmap for understanding and analysing this important aspect of Bengali grammar. It opens up further avenues of future research and practical application.

This dissertation also provides empirical evidence to the central hypothesis that different senses of a verb always produce different verb-frames. It has been established through the discussion and presentations in chapter 5 that verb-frames of different senses of the same verb are always distinct. Therefore, verb-frame approach is proven to be useful in verb-sense disambiguation.

6.2 Aims and Objectives

Aims and objectives were introduced in the introductory chapter. The main aim of the study is to develop the Bengali Verb-frame Network which is an interconnected repository of argument structure of Bengali verbs annotated at various levels. Along with the verb-frame, we have also classified verbs based on the similarity of verb-frames. Apart from these two principal objectives, we had to develop a novel methodology and two different annotation schemes in order to annotate the verbal arguments. In chapter 2, we have elaborated on the existing research in the relevant field. Chapter 3 discusses the methodology. Four sections of a verb-frame, i.e., Dependency Relation, Lexical Type, Dependency Marker, Ontology, each of which represents a different layer of information, are detailed in Chapter 4. There are two elaborate annotation schemes used to annotate the dependency roles and ontology, respectively. They are explained in this chapter. Verb-frames were presented and analysed as well as the verb classification is established in chapter 5. Chapter 5 also includes a section in which some complex Bengali constructions were analysed through the verb-frame approach. Total 300 verb-frames were created through the course of the research and 32 verb classes were identified.

6.3 Scope of Future Research

The knowledge generated from the research on Bengali verb frames can be expanded in several ways. We plan to add more verb-frames to the existing repository so that, eventually, it comprises all lexical verbs and the most common conjunct and compound verbs used in Bengali.

One promising avenue for future research is to investigate the role of discourse and pragmatic factors in shaping the use of verb frames in Bengali. One such interesting research question is -how do speakers use different verb frames to signal information structure, such as topic and focus, in discourse? How do pragmatic considerations influence the choice of verb frames in different speech registers or genres, such as formal versus informal speech or written versus spoken discourse?

Another area for further investigation is the cross-linguistic variation in verb frames across different languages. How do verb frames in Bengali compare to those in other languages, both within and outside the Indo-European language family? What are the underlying factors that explain the similarities and differences in verb-frame usage across languages, and how can we use this knowledge to improve our understanding of language typology and linguistic universals?

Finally, it is worth noting that this study provides only a snapshot of the development of Bengali verb frames at a particular point in time. As with any living language, Bengali continues to evolve and change over time, and the patterns of verb-frame usage may shift or evolve in response to cultural, social, and linguistic factors. Future research could benefit from a longitudinal approach that tracks changes in verb-frame usage over time and examines the factors that drive these changes.

6.3.1 Possible Practical Applications

1) Bengali Verb-frame Network is tailor-made for verb-sense disambiguation. The approach taken here involves creating a separate verb-frame for each distinct meaning of a verb based on its usage in different contexts. It is created as a table and saved in a .txt file so that any computational system can easily employ this resource. It can directly be used in any NLP tool for verb-sense disambiguation in Bengali. It will definitely improve the accuracy of language-based processing.

2) It can be used for different purposes by different computational tools. For example, in machine translation, verb frames can be used to generate more accurate translations by disambiguating the meaning of words based on their associated verb frames. In text classification, verb frames can be used to identify the topic or intent of a text by analysing the semantic features of the verbs used in it. In information retrieval, verb frames can be used to improve the precision and recall of search results by matching the user's query with the relevant verb frames.

3) Verbal arguments are marked for ontological properties. It is already shown in various studies like Pazos et al. (2008) and Koeva and Doychev (2022) that ontological information can be used to automatically extract semantic frames and other tasks.

4) It can be made into a useful tool for foreign language learners of Bengali. Verbs are segmented into different senses, and the context in which a particular verb sense occurs is well-represented. Therefore, sense distinction can easily be inferred even by a person who doesn't know the language. Apart from it, verbs are classified according to their dependency structure. The verb classification is very much comparable to Levin's verb classes. Learners can learn a lot from it regarding the syntactic behaviour of Bengali verbs.

6.4 Final Words

Bengali verb-frame Network should be treated as an application-neutral resource created in the spirit of descriptive linguistics. Verb-frames were created for each verb-sense which distinguishes it from the Western attempts to create similar kinds of resources. Overall, the development of Bengali verb frames is a rich and complex area of research that has the potential to deepen our understanding of the structure and function of verb semantics in Bengali and beyond. A dependency framework influenced by the Paninian approach is at the core of this research. So, it constitutes an amalgamation of Western and Indian traditions of linguistics.

It has certain limitations too. One major limitation is that our corpus represents the Standard Colloquial Bengali. The other dialects of Bengali have not been included to preserve the consistency and homogeneity of the verb-usage pattern. I expect future researchers to come up and explore the dialects so that we get a comparative picture of the verb-frames between SCB and the dialects of Bengali.

APPENDIX I

VERBS AND THEIR PRIMARY DEPENDENCY STRUCTURE⁶

Sense_ID	Sense	Primary Dependency Structure
aanaa_1	'to bring'	k1 k2 k5 k7p
aanaano_1	'to cause to bring'	pk1 mk1 k2 k5
aasaa_1	'to come'	k1 k2p
aasaa_2	'to be able to'	k1im k2
aasaa_3	'to belong to'	k1 k7p/k7
aasaa_4	'to appear'	k1 k7p
aasaa_5	'to get involved'	k1 k7
aasaa_6	'to start to feel'(sleep, weep etc)	k1im k2
aasaa_7	'to be raised'(topic, concept etc)	k1
baa.Ndhaa_1	'to tie up'	k1 k2 k3
baa.Ndhaa_2	'to get restricted or prevented'(emotionally)	k1im k2
baanaano_1	'to change into'	pk1 jk1 k2
baanaano_2	'to accept as something'	k1 k2 k2s
baanaano_3	'to create'	k1 k2
baanaano_4	'to make a story'	k1 k2
baaRaa_1	'to go up'	k1
bakaa_1	'to speak incoherently'	k1 k2
bakaa_2	'to scold'	k1 k2
balaa_1	'to tell'	k1 k2 k4
balaa_2	'to chant'	k1 k2
balaa_3	'to be called/pronounced as'	k1 k4 k2
balaa_4	'to speak as representative'	k1 k5a k7p
balaano_1	'to cause to speak'	pk1 mk1 k2
banaa_1	'to possess'	k1 k2
banaa_2	'to become'(by having some attributes)	k1 k1s
banaa_3	'to act or play a part'	k1 k1s k7
banaa_4	'to become something else'	k1 k1s
basaa_1	'to sit'	k1 k7p
basaa_2	'to establish'	k1 k7p
basaa_3	'to engage in some activity'	k1 k2
basano_1	'to cause to sit'	pk1 jk1 k2

⁶ Some verbs have more than one verb-frames. In such cases, primary dependency structure refers to the dependency structure of the most common verb-frame. Most verbs have only one verb-frame. Hence, in case of most verbs, the term 'Primary Dependency Structure' do not imply anything else other than the dependency structure of the verb-frame.

	basaano_2	'to impose or install'	pk1 jk1 k7
	bhaabaa_1	'to think'	k1 k2
	bhaabaa_2	as' 'to perceive or take someone'	k1 k2 k2s
1	bhaalobaasaa_	'to love someone'	k1 k2
2	bhaalobaasaa_	'to have a fondness to'	k1 k2
3	bhaalobaasaa_	'to have a passion or liking'	k1 k2
	bhaangaa_1	'to break'	k1 k2 k3
	bhaangaa_2	'to break up'	k1 k2
	bhaangaa_3	'to be sad'	k1 k2
	bhaangaa_4	'to break record, barrier etc'	k1 k2
	bhaangaa_5	'to disintegrate'(organisation)	k1 k2
	bharaa_1	'to fill'	k1 k2 k7p
	bharaa_2	'to be filled with crowd'	k1
	bharaa_3	'to earn'	k1 k7
	bikri karaa_1	'to sell'	k1 k2 k4
	bojhaa_1	'to understand'	k1 k2
	bojhaa_2	'to feel'	k1 k2
	bojhaa_3	'to figure out or fathom'	k1 k2
	bojhaa_4	'to perceive'	k1 k2 k2s
	caaoyaa_1	'to want'	k1 k2
	caaoyaa_2	'to need'	k1im k2
	caapaa_1	'to press'	k1 k2 k3
	caapaa_2	'to suppress'	k1 k2
	caapaa_3	'to board a vehicle'	k1 k2 k5
	caapaa_4	'to be burdened'	k1 k7
	caapaa_5	'to physically throw oneself on someone's body'	k1 k7
	calaa_1	'to walk'	k1
	calaa_10	'to allow'	k1 rkc
	calaa_2	'to go'	k1 k2p
	calaa_3	'to run'(transport)	k1 k7p
	calaa_4	'to have a particular time'	k1
	calaa_5	'to be able to take or be habituated'	k1im k2
	calaa_6	'to continue'	k1
	calaa_7	'to carry on in a certain way'	k1 k2
	calaa_8	'to last'	k1 k7t
	calaa_9	'to happen or occur'	k1 k7p
	cenaa_1	'to recognise'	k1 k2
	cenaa_2	'to know someone'	k1 k2
	cenaa_3	'to learn'	k1 k2
	chaar		
deoyaa_1		'to give discount'	k1 k7

chaar deoyaa_2	'to exempt'	k1 k7
daa.NRaano_1	'to contest in an election'	k1 k7p
daa.NRaano_2	'to stand'	k1 k7p
Daakaa_1	'to call'	k1 k2
Daakaa_2	'to make a noise'(animals)	k1
Daakaa_3	'to designate as'	k1 k2 k2s
dekhaa_1	'to see'	k1 k2
dekhaa_2	'to watch'	k1 k2
dekhaa_3	'to check/test'	k1 k2
dekhaa_4	'to monitor'	k1 k2
dekhaa_5	'to have a viewpoint'	k1 k2
dekhaa_6	'to consider'	k1 k2 k2s
dekhaano_1	'to show'	pk1 jk1 k2
dekhaano_2	'to look'	k1im k2
dekhaano_3	'to bring someone for check or test'	pk1 jk1 k2
dekhaano_4	'to show feelings'	k1 k2
dekhaano_5	'to be made to experience'	pk1 jk1 k2
dekhaano_6	'to prove'	k1 k2
deoyaa_1	'to pay'	k1 rt k4 k2
deoyaa_2	'to put'	k1 k2 k7p
deoyaa_3	'to give'	k1 k4 k2
deoyaa_4	'to assign or impose'	k1 k4 k2
deoyaa_5	'to cause an emotion, sensation etc'	k1 k4 k2
deoyaa_6	'to convey'	k1 k4 k2
dhaakkaa khaoyaa_1	'to collide'	k1 k7p
dhaakkaa khaoyaa_2	'to have a shock'	k1 k7
dharaa_1	'to catch'	k1 k2
dharaa_2	'to get a transport'	k1 k2
dharaa_3	'to have an abstract state'	k1 k7
dharaa_4	'to grab'	k1 k2
dharaa_5	'to capture'(abstract)	k1 k2
dharaa_6	'to hire or employ'	k1 k2 rt
Dobaa_1	'to drown or sink'	k1 k7p
Dobaa_2	'to disappear beyond the horizon'	k1
Dobaa_3	'to be ruined'	k1
galaa_1	'to melt'	k1
galaa_2	'to have positive feelings'	k1 k7
gaRaa_1	'to build'	k1 k2 k7p
gaRaa_2	'to develop'(human relations)	k1 k2 cok1
gaRaa_3	'to organise'	k1 k2
gaRaa_4	'to create or make'	k1 k2
ghoraa_1	'to revolve'	k1

ghoraa_2	'to turn'	k1 rd
ghoraa_3	'to walk around/visit'	k1 k7p
ghoraano_1	'to make something or someone revolve'	pk1 jk1
ghoraano_2	'to prepare travel and visit for someone'	pk1 jk1 k2p
ghoraano_3	'to harass'	k1 k2 rt
ghumono_1	'to sleep'	k1
ghumono_2	'to be inactive'	k1
haa.NTaa_1	'to walk'	k1 k7p
haa.NTaa_2	'to go on'	k1 k7
haoyaa_1	copula	k1 k1s
haoyaa_2	'to get changed into'	k1 k2
haoyaa_3	'to attain some (social) role'	k1 k1s
haoyaa_4	'to be in a state'(feelings, sickness etc)	k1im k2
haoyaa_5	'to occur or happen'	k1im k2
haoyaa_6	'to be created'	k1
haoyaa_7	'to possess'	k1im k2
jaagaa_1	'to be awoken'	k1 k7t/rsd
jaagaa_2	'to have sudden outbursts of desire, expectation etc'	k1 k7/k7t
jaanaa_1	'to be able to'	k1 k2
jaanaa_2	'to know'	k1 k2
jaanaa_3	'to feel'	k1 k2
jaanaa_4	'to know someone'	k1 k2 rsd
jaanaa_5	'to regard'	k1 k2 k2s
jaaoyaa_1	'to go'	k1 k2p
jaaoyaa_2	'to loss'	k1 k7
jaaoyaa_3	'to match'	k1 cok1
jaaoyaa_4	'to stretch'(mental activity)	k1 rd
jaaoyaa_5	'to have lost'(Honour)	k1
jamaa_1	'to be deposited'	k1 k7
jamaa_2	'to have an enjoyable occasion'	k1
jamaa_3	'to be accumulated'	k1 k7p
jamaa_4	'to be in good terms'(with someone)	k1 cok1
jharaa_1	'to drop'	k1 k5
jho.Nkaa_1	'to lean forward'	k1 k5
jho.Nkaa_2	'to favour a side'(in a conflict situation)	k1 k7 rd
jvalaa_1	'to burn'	k1 k7p
jvalaa_2	'to be jealous'	k1 rt
jvalaa_3	'to feel hot or itching'	k1
kaaje laagaa_1	'to be useful'	k1im k2
kaamaano_1	'to earn'	k1 k2 k7

kaamaano_2	'to shave'	k1 k2
kaaTaa_1	'to cut'	k1 k2 k3
kaaTaa_2	'to deduct'(amount of money)	k1 k2 k5
kaataa_3	'to dig'	k1 k7p
kaaTaa_4	'to recover or be okay'	k1
kaaTaa_5	'to mark with a cross or a line'	k1 k2
kaaTaa_6	'to pass the time'	k1
kenaa_1	'to buy'	k1 k2 k5
khaaTaa_1	'to work'	k1 k7p
khaaTaa_2	'to put effort'	k1 rt
khaaTaa_3	'to be applicable'	k1 k7
khaoyaa_1	'to eat'	k1 k2
khaoyaa_2	'to drink'	k1 k2
khaoyaa_3	'to consume'	k1 k2
khaoyaa_4	'to take bribe'	k1 k2
khaoyaano_1	'to feed'	pk1 jk1 k2
khelaa_1	'to play'	k1 k2 k7p
khelaa_2	'to conceive an idea'	k1 k2
khelaa_3	'to work'(used for supernatural and larger than life figures)	k1 k2
khola_1	'to open an institution at a particular time'	k1 k7t
khola_10	'to open'(abstract entity)	k1 k2
khola_2	'to start'	k1 k2
khola_3	'to unstitch'	k1 k2
khola_4	'to dismantle or uncover a machine'	k1 k2
khola_5	'to open a document'	k1
khola_6	'to unlock and open'	k1 k2 k3
khola_7	'to undress'	k1 k2
khola_8	'to unravel'	k1
khola_9	'to untie or uncuff'	k1 k2
balaa_1	kothaa 'to talk'	k1 cok1
laagaa_1	'to be in use'	k1
laagaa_2	'to require'(money)	k1 rt
laagaa_3	'to occur or break out'	k1
laagaa_4	'to be in touch of something'(mainly liquid or dust)	k1 k7p
laagaa_5	'to get touched'	k1 k7p
laagaa_6	'to look'	k1im k2 vmod
laagaa_7	'to feel' (usually temperature, pain etc)	k1im k2
laagaa_8	'to have caught attention or get shocked'	k1 k2
laagaano_1	'to attach'	k1 k2 k7p
laagaano_2	'to spend'(money)	pk1 jk1 rt

laagaano_3	'to cause a fight'	k2 pk1 cok1 jk1
laagaano_4	'to put a vehicle at some place'	pk1 jk1 k7p
laagaano_5	'to touch'	k1 k2 k7p
laagaano_6	'to paint'	pk1 jk1 k7p
laagaano_7	'to shift the blame on someone else'	k1 k2 k7
laagaano_8	'to do sexual intercourse'	k1 k2
laRaa_1	'to fight'	k1 cok1
maar khaoyaa_1	'to get beaten'	k1 rkc
maar khaoyaa_2	'to lose'	k1
maaraa_1	'to kill'	k1 k2
maaraa_2	'to beat'	k1 k2 k3
mane karaa_1	'to mind'	k1 k7
mane karaa_2	'to recall'	k1 k2
mane karaa_3	'to consider as'	k1 k2 k2s
mane karaa_4	'to suppose'	k1 k2
maraa_1	'to die'	k1
maraa_2	'to be reduced'	k1 rt
melaa_1	'to spread out'	k1 k2 k7p
meTaa_1	'to be completed'	k1 k7p
meTaa_2	'to meet some expectations, desire, demand etc'	k1
milaa_1	'to correspond'	k1 cok1
milaa_2	'to be able to get'	k1 k7
naamaa_1	'to get down'	K1 k5 k7p
naamaa_2	'to start or be involved in some activity'	k1 k7
naamaa_3	'to reduce or decrease'	k1
neoyaa_1	'to borrow'	k1 k2 k5
neoyaa_2	'to recruit'	k1 k2 k7
neoyaa_3	'to take shelter'	k1 k7p
neoyaa_4	'to take'(abstract)	k1 k2
neoyaa_5	'to assume a role'	k1 k2 k7
paaoyaa_1	'to get'	k1 k2 k5
paaoyaa_2	'to feel'	k1 k2
paaoyaa_3	'to achieve or attain'	k1 k2 k7
paaoyaa_4	'to receive'	k1 k2 k5
paaoyaa_5	'to have or possess'	k1 k2
paaoyaa_6	'to be able to perceive'	k1 k2 k4
paaraa_1	'to be able to'	k1 k2
paataa_1	'to lay out'(for construction)	k1 k2 k7p
paataano_1		
paaThaano_1	'to send'	k1 k2 k4

paRaa_1	'to read'	k1 k2
paRaa_2	'to understand'(mind, situation etc)	k1 k2
paRaa_3	'to study'	k1 k7
phaa.Nsaa_1	'to be trapped'	k1 k7p
phaa.Nsaa_2	'be in trouble'	k1 rt
pheraa_1	'to return'	k1 k5 k7p
pheraa_2	'to turn around'	k1 rd
phoTaa_1	'to get pricked'	k1 k7p
phoTaa_2	'to appear'	k1
phoTaa_3	'to boil'	k1 k7p
raakhaa_1	'to put'	k1 k2 k7p
raakhaa_2	'to express an opinion'	k1 k2 k7p
raakhaa_3	'to keep'(something in certain order)	k1 k2
raakhaa_4	'to keep'	k1 k2 rkc
raakhaa_5	'to employ'	k1 k2
raakhaa_6	'to have'(abstract)	k1 k2 k7
saaraa_1	'to recover'	k1 rt
saaraa_2	'to do or execute something'	k1 k2
shekhaa_1	'to learn'	k1 k2
shonaa_1	'to hear'	k1 k2
shonaa_2	'to be aware of, get to know'	k1 k2
shonaa_3	'to obey'	k1 k2
shonaa_4	'to listen to'	k1 k2
shonaa_5	'to endure insult'	k1 k2
shoyaa_1	'to lie down'	k1
shoyaa_2	'to have sex'	k1 cok1
Taanaa_1	'to attract'(gravitational)	k1 k2
Taanaa_2	'to attract'(personal)	k1 k2
Taanaa_3	'to pull'	k1 k2
Taanaa_4	'to draw'	k1 k2 k5
Taanaa_5	'to be burdened with responsibility, liability etc,	k1 k2
Taanaa_6	'to draw a line or curve'	k1 k2
tairii karaa_1	'to make'	k1 k2
tairii karaa_2	'to create'(abstract)	k1 k2
thaakaa_1	existential be	k1 k7/k7p
thaakaa_2	'to have an attribute'	k1 k1s
thaakaa_3	'to have'	k1im k2
thaakaa_4	'to have a quality'	k1 k7p
thaakaa_5	'to last'	k1 k7t
thaakaa_6	'to stay'	k1 k7p
thaakaa_7	'to represent'	k1 k5a k7p
thaamaa_1	'to stop'	k1 k7p
thaamaa_2	'to stop a process or activity'	k1
Thelaa_1	'to push'(concrete entity)	k1 k2 rd

Thelaa_2	'to go forward through pushing'	k1 k2
Thelaa_3	'to push'(in an abstract sense)	k1 k2 rd
tolaa_1	'to lift'	k1 k2 k3
tolaa_2	'to bring a topic'	k1 k2
tolaa_3	'to wake'	k1 k2 k5
tolaa_4	'to rescue from'(a mental situation)	k1 k2 k5
tolaa_5	'to raise money'	k1 k2 k5
tolaa_6	'to build'	k1 k2 k7p
tolaa_7	'to pluck'	k1 k2
tolaa_8	'to voice'(voice)	k1 k2
uthaa_1	'leave one's seat'	k1 k5
uthaa_10	'to board on a vehicle'	k1 k7p
uthaa_11	'to arrive in a market'	k1 k7
uthaa_2	'to raise'(price, income etc)	k1 k7
uthaa_3	'to be raised' (question, complaint etc)	k1 k7
uthaa_4	'to get promoted'	k1 k2
uthaa_5	'to grow or advance'(one's social-economic situation)	k1
uthaa_6	'to grow' (any socio-economic entity itself)	k1
uthaa_7	'to rise'(astronomical body)	k1
uthaa_8	'to wake up'	k1 k5
uthaa_9	'to go up/climb up'	k1 k3 k7p
uthaano_1	'to lift'	pk1 jk1 k2p
uthaano_2	'to raise'(complaint, question etc)	k1 k2

APPENDIX II

VERB CLASSES AND THEIR MEMBERS

Class	Verb	Dependency Structure	Members of the Verb Class
1		$V_{int}[k1]$	baaRaa_1, calaa_1, calaa_4, calaa_6, Daakaa_2, Dobaa_2, Dobaa_3, jaaoyaa_5, jamaa_2, jvalaa_3, kaaTaa_4, kaaTaa_6, maraa_1, naamaa_3, phoTaa_2, galaa_1, laagaa_1, ghumono_1, ghoraa_1, uthaa_5, uthaa_6
2		$V_{tr}[k1,k2]$	Daakaa_1, bakaa_1, bakaa_2, balaa_2, bojhaa_1, basaa_3, calaa_7, kaaTaa_5, paaraa_1, paaraa_2, saaraa_2, tairii karaa_1, bojhaa_3, dekhaa_1, dekhaa_3, dekhaa_4, dekhaa_5, dekhaano_6, shonaa_2, jaanaa_1, dharaa_1, khaoyaa_1, khaoyaa_2, Taanaa_1.txt, Taanaa_3, Taanaa_6, Thelaa_2, tolaa_7, tolaa_2 , uThaa_4, cenaa_1, bhalobaasaa_1, bhalobaasaa_2, neoyaa_4, khelaa_2, khelaa_3,

		laagaano_8, haoyaa_2, jaanaa_3, shekhaa_1
3	$V_{dtr}[k1,k2,k4]$	balaa_1, deoya_1, deoya_3, deoya_4, deoya_5, deoya_6, paaThaano_1
4	$V_{Caus}[pk1,jk1]$	ghoraano_1
5	$V_{Caus}[pk1,jk1,k2]$	banaano_1, dekhaano_1, dekhaano_3, dekhaano_5, khaaoyaa_1
6	$V_{tr}[k1,k2,k3]$	baa.Ndha_1,kaaTaa_1, bharaa_1, caapaa_1, kholaa_6, tolaa_1, maaraa_2
7	$V_{int}[k1,k2p]$	aasaa_1, calaa_2, jaaoyaa_1
8	$V_{int}[k1,k7p]$	Dobaa_1, aasaa_3, aasaa_4, basaa_1, basaa_2, calaa_3, calaa_9, Daa.NRaano_1, Daa.NRaano_2, dhaakkaa khaaoyaa_1, haoyaa_5, ghoraa_3, haa.NTaa_1, jamaa_3, jvalaa_1, kaaTaa_3, khaaTaa_1, laagaa_4, metaa_1, neoyaa_3, phaa.Nsaa_1, phoTaa_1, thaakaa_4, thaakaa_6, thaamaa_1, uthaa_10
9	$V_{tr}[k1,k2,k2s]$	bhaabaa_2, Daakaa_3, banaano_2

10	$V_{tr}[k1,k2,k7p]$	bharaa_1, bikri karaa_1, khelaa_1, laagaano_1, paataa_1, tolaa_6
11	$V_{int}[k1,cok1]$	jaaoyaa_3, jamaa_4, laaRaa_1, milaa_1, shoyaa_2
12	$V_{int}[k1,rt]$	maraa_2, khaaTaa_2, jvalaa_2, ph.Naasaa_2, laagaa_2, saaraa_1
13	$V_{int}[k1,k7]$	naamaa_2, jamaa_1, khaaTaa_3, aasaa_5, milaa_2
14	$V_{int}[k1,k5,k7p]$	naamaa_1, pheraa_1
15	$V_{int}[k1,rd]$	pheraa_2, jaaoyaa_4, ghoraa_2
16	$V_{tr}[k1im,k2]$	aasaa_2, aasaa_6, baa.Ndhaa_2, caaoyaa_2, calaa_5, haoyaa_4, haoyaa_5, haoyaa_7, kaaje laaga_1, laagaa_7, paaoyaa_2, thaakaa_3
17	$V_{tr}[k1,k2,cok1]$	gaRaa_2, laagaa_3,
18	$V_{tr}[k1,k2,k7]$	kenaa_1 , kaamaano_1, laagaano_7, neoyaa_2, neoyaa_5, paaoyaa_3
19	$V_{int}[k1,vmod]$	pheraa_2
20	$V_{tr}[k1,k2,vmod]$	laagaa_6
21	$V_{tr}[k1,k2,k5]$	kaaTaa_2, kenaa_1
22	$V_{int}[k1,k1s]$	haoyaa_1, haoyaa_3, banaa_2, banaa_4, thaakaa_2

23	$V_{int}[k1,k5]$	jharaa_1, jho.Nkaa_1, uThaa_8
24	$V_{int}[k1,k5a,k7p]$	balaa_4, daa.Nraano_1, thaakaa_7
25	$V_{int}[k1,k7t]$	basaa_2, calaa_8, jaagaa_1, jaagaa_2 kholaa_1, thaakaa_5
26	$V_{tr}[k1,k2,rd]$	Thelaa_1, Thelaa_3, deoyaa_2
27	$V_{int}[k1,rkc]$	calaa_10, maar khaaoyaa_1, shekhaa_1
28	$V_{int}[k1,k5prk]$	haoyaa_2
29	$V_{caus}[pk1,mk1,k2]$	balaano_1
30	$V_{caus}[pk1,mk1,jk1,k2]$	deoyaano_1
31	$V_{caus}[pk1,mk1,k5,k2]$	aanaano_1
32	$V_{caus}[pk1,jk1,k7p]$	basaa_1, laagaano_4,

REFERENCES

- Agnihotri, Ramakant. (2007). *Hindi: An Essential Grammar*. Routledge.
- Baker, C., & Ruppenhofer, J. (2002). FrameNet's Frames vs. Levin's Verb Classes. In J. Larson and M. Paster (Eds), *Proceedings of the 28th Annual Meeting of the Berkeley Linguistics Society*, (p. 27-28). Berkeley Linguistics Society
- Bayer, J. (2010). Two Grammars in One: Sentential Complements and Complementizers in Bengali and other South Asian Languages. In R. Singh & P. Dasgupta (Ed.), 2001 (pp. 11-36). Berlin, New York: De Gruyter Mouton.
<https://doi.org/10.1515/9783110245264.11>
- Begum, R., & Sharma, D. M. (2017a). Development and Analysis of Verb Frame Lexicon for Hindi. *Linguistics and Literature Studies* 5(1), 1-22. DOI: 10.13189/lis.2017.050101
- Begum, R. (2017b). *Developing a pilot Hindi Treebank based on Computational Paninian Grammar*. [Doctoral Dissertation, International Institute of Information Technology, Hyderabad]. Hyderabad.
- Berlin, B., & Kay, P. (1969). Basic Color Terms: Their Universality and Evolution. University of California Press.
- Bharati, A., Sharma, D. M., Bai, L., & Sangal, R. (2006). AnnCorra: Annotating Corpora Guidelines for POS And Chunk Annotation for Indian Languages. *LTRC, IIIT Hyderabad*.
- Bharati, A., Husain, S., Vijay, M., Deepak, K., Sharma, D.M., & Sangal, R. (2009). Constraint Based Hybrid Approach to Parsing Indian Languages. *Pacific Asia Conference on Language, Information and Computation*, 614-621
- Bharati, A., & Sangal, Rajeev. (2010). *Computational Paninian Grammar Framework*.

- Bharati, A., Chaitanya, V., & Sangal, R. (1995). *Natural Language Processing: A Paninian Perspective*. New Delhi, India: Prentice Hall of India
- Bharati, A., Sharma, D. M., Hussain, S., Bai, L., Begum, R., & Sangal, R. (2012). *Ancorra: TreeBanks for Indian Languages*.
- Blevins, J. (2003). Passives and impersonals. *Journal of Linguistics* 39. 473-520.
- Bloomfield, L. (1933). *Language*. Holt.
- Busa, F. (1996). *Compositionality and the Semantics of Nominals*. [Doctoral Dissertation, Brandeis University]. Waltham, MA.
- Brachman, Ronald J. (1979). On the Epistemological Status of Semantic Networks. In N. Findler (ed.) *Associative Networks: Representation and Use of Knowledge by Computers*. Academic Press.
- Carnap, Rudolf (1956). *Meaning and Necessity*. University of Chicago Press.
- Chakrabarti, D., & Bhattacharyya, P. (2004). Creation of English and Hindi Verb Hierarchies and their Application to Hindi WordNet Building and English-Hindi MT. *Proceedings of the International Conference on Global Wordnet (GWC 04), Brno, Czech Republic*. 83-90.
- Chatterji, S.K. (1968). *সরল ভাষা-প্রকাশ বাংলা ব্যাকরণ*. Baksahitya.
- Chomsky, N. (1957). *Syntactic structures*. Mouton.
- Chomsky, N. (1981). *Lectures on Government and Binding*. Foris Publications.
- Comrie, B. (1985). Causative verb formation and other verb-deriving morphology. T. Shopen (ed.), *Language typology and syntactic description, Vol 3*, (pp. 309-348.) Cambridge: Cambridge University Press
- Collins, A., and Quillian, M. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*, 9, 240-247.
- Creissels, D. (2008). Impersonal and related constructions: a typological approach.

- Dasgupta, P. (2004). Some non-nominative subjects in Bangla. In P. Bhaskararao & P. B. Subbarao, *Non-nominative Subjects* (pp. 129-140). John Benjamins Publishing Company.
- Dasgupta, P. (2006). On Bangla Impersonals, Including Passives. *Deccan College Working Papers in Linguistics* 7.1-18.
- Dash, N.S. (2015). *A Descriptive Study of Bengali words*. Cambridge University Press.
- Dowty, D. (1991). Thematic Proto-Roles and Argument Selection. *Language*, 67(3), 547–619. <http://www.jstor.org/stable/415037>
- Fillmore, C. (1976). Frame semantics and the nature of language. Conference on the origin and development of language and speech. *Annals of the New York Academy of Sciences*, Volume 280, 20–32.
- Fillmore, C. J. (1977a). The Case for Case Reopened. In P. Cole (ed.), *Syntax and Semantics 8: Grammatical Relations* (pp. 59-81). Academic Press.
- Fillmore, C. J. (1977b). Topics in Lexical Semantics, In R. Cole (ed.), *Current Issues in Linguistic Theory* (pp. 76-138). Indiana University Press.
- Fillmore, C. J. (1977c). Scenes-and-frame Semantics. In Antonio Zampolli (ed.), *Linguistic Structures Processing* (pp. 55-82). North Holland Publishing Company.
- Fillmore, C. J. (1982a.) Frame Semantics. In *Linguistics in the Morning Calm*. Hanshin, 111-138.
- Fillmore, C. J. (1982b). Toward a Descriptive Framework of Spatial Deixis. In R. J. Jarvella and W. Klein (eds.), *Speech, Place and Action*. John Wiley.
- Fillmore, C. J. (1985). Frames and the Semantics of Understanding. *Quaderni di Semantica* 6(2): 222-53.

- Fodor, J. (1970). Three Reasons for Not Deriving 'Kill' from 'Cause to Die'.
Linguistic Inquiry 1, 429-438.
- Fodor, J (1975). *The Language of Thought*. Harvard University Press.
- Fodor, J. A. (1980). 'Methodological Solipsism Considered as a Research
Strategy in Cognitive Psychology. *Behavioral and Brain Sciences* 3, 63-73.
- Ghosh, S. (2014). Making Verb Frames for Bangla Vector Verbs. *Proceedings of the 11th
International Conference on Natural Language Processing*. Goa, India. 305–314
- Goldberg, A. E. (1992). The Inherent Semantics of Argument Structure: The Case of the
English Ditransitive Construction. *Cognitive Linguistics* 3(1): 37-74.
- Goldberg, A. E. (1995) *Constructions: A Construction Grammar Approach to Argument
Structure*. University of Chicago Press.
- Goldberg, A. E. (1998). Semantic Principles of Predication. In Jean-Pierre Koenig (ed.),
Discourse and Cognition: Bridging the Gap (pp. 41-55). CSLI Publications.
- Goldberg, A. E. (2005). Constructions, Lexical Semantics and the Correspondence
Principle: Accounting for Generalizations and Subregularities in the Realization
of Arguments. In N. Erteschik-Shir & T. Rapoport (eds.), *The Syntax of Aspect*.
Oxford University Press.
- Goldberg, A. E. (2006). *Constructions at Work: The Nature of Generalization in
Language*. Oxford University Press.
- Goldberg, A. E., and Jackendoff, R. (2004). The Resultative as a Family of
Constructions. *Language* 80(3), 532-568.
- Goldberg, A. E. (2010). Verbs, Frames and Constructions. In M. Rappaport Hovav, E.
Doron and I. Sichel (eds.). *Syntax, Lexical Semantics and Event Structure*. Oxford
University Press. 39-58.
- Gruber, J. S. (1965) *Studies in Lexical Relations*. PhD Dissertation. MIT.

- Gruber, T. R. (1993). Toward principles for the design of ontologies used for knowledge sharing. *International Journal Human-Computer Studies* 43(5-6), pp. 907-928.
<https://doi.org/10.1006/ijhc.1995.1081>.
- Jackendoff, R. (1985). Multiple Subcategorization and the Theta-Criterion: The Case of Climb. *Natural Language and Linguistic Theory* 3: 271-295.
- Jackendoff, R. (1972). *Semantic Interpretation in Generative Grammar*. MIT Press.
- Jackendoff, R. (1990). *Semantic Structures*. Massachusetts: MIT Press
- Joshi, S.D. (2009). Background of the Aṣṭādhyāyī. In Amba Kulkarni and Gérard Huet (Eds.), proceedings of the *Third International Sanskrit Computational Linguistics Symposium*, Hyderabad, India, pp. 1-5.
- Katz, J. J. (1972). *Semantic Theory*. New York: Harper and Row.
- Katz, J. J. and Fodor, J. (1963). The structure of a semantic theory. *Language* 39: 170–210.
- Katz, J. J., & Postal, P. M. (1964). *An integrated theory of linguistic descriptions*.
M.I.T. Press.
- Karim, Rezaul & Chakravarthi, Bharathi & Arcan, Mihael & McCrae, John & Cochez, Michael. (2020). *Classification Benchmarks for Under-resourced Bengali Language based on Multichannel Convolutional-LSTM Network*.
- Kingsbury, P., & Palmer, M. (2002). From Treebank to PropBank. *Proceedings of the 3rd International Conference on Language Resources and Evaluation, Las Palmas, Canary Islands, Spain*, pp. 1989–1993.
- Kipper, K., Dang, H., & Palmer, M. (2000) Class-based construction of a verb lexicon. *AAAI/IAAI*, 691-696.
- Kipper, K. (2005). *VerbNet: A broad-coverage, comprehensive verb lexicon*. [Ph.D. thesis, Computer and Information Science Dept., University of Pennsylvania] Philadelphia, PA.

- Kiparsky, P. (2009). On the Architecture of Pāṇini's Grammar. In: Huet, G., Kulkarni, A., Scharf, P. (eds) Sanskrit Computational Linguistics. ISCLS 2007 2008. Lecture Notes in Computer Science(), vol 5402. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-00155-0_2.
- Kiparsky, P., & Staal, J.F. (1969). Syntactic and Semantic Relations in Panini. *Foundations of Language*. 5, 83–117.
- Klaiman, M.H. (1980). Bengali Dative Subjects. *Lingua* 51. pp 275–295. North Holland Publishing Company.
- Klaiman, M.H. (1981). *Volitionality and Subject in Bengali: A Study of Semantic parameters in Grammatical Process*. Indiana
- Koeva, S and Doychev. E. (2022). Ontology Supported Frame Classification. *Proceedings of the 5th International Conference on Computational Linguistics in Bulgaria (CLIB 2022)*, pp 203–213, Sofia, Bulgaria.
- Lakoff, G. (1971). On generative semantics. In D. Steinberg and L. Jakobovits (eds.), *Semantics: An Interdisciplinary Reader*. Cambridge University Press.
- Lakoff, G. (1987). Cognitive models and prototype theory. In U. Neisser (Ed.), *Concepts and conceptual development: Ecological and intellectual factors in categorization* (pp. 63–100). Cambridge University Press.
- Lenat, D., Prakash, M., & Shepherd, M. (1986). CYC: Using common sense knowledge to overcome brittleness and knowledge acquisition bottlenecks. *AI Magazine*, 6(4), 65–85.
- Levin, B., & Rappaport Hovav, M. (1995). *Unaccusativity: At the Syntax-Lexical Semantics Interface*. Retrieved September 2, 2021, from <https://hdl.handle.net/2027/heb.08443>.

- Levin, B. (1993). *English Verb Classes and Alternations: A Preliminary Investigation*. Chicago: University of Chicago Press.
- Matilal, B. K. (1991). BHAVĀNANDA ON WHAT IS KARAKA. In M.M. Deshpande and S. Bhate (Eds.), *Paninian Studies: S. D. Joshi Felicitation Volume* (pp. 263-282).
- Miller, G. A., Beckwith, R., Fellbaum, C. D., Gross, D., & Miller, K. (1990). WordNet: An online lexical database. *International Journal of Lexicography*, 3(4), 235–244.
- Mishra, A. (2009). Modelling the Grammatical Circle of the Paninian system of Sanskrit Grammar. In A. Kulkarni and G.Huet (Eds.), proceedings of the *Third International Sanskrit Computational Linguistics Symposium*, Hyderabad, India, pp. 40-55.
- Mithun, M. Forthcoming. (2008). The emergence of agent/patient systems. In Donohue, M. & S. Wichmann (eds.), *The typology of semantic alignment systems*. Oxford: Oxford University Press.
- Mitkovska, Liljana and Bužarovska, Eleni. (2021). Quasi-passive reflexive constructions: Bridging autonomous and passive situations. *Linguistic Typology*, 25(2), 303-343.
<https://doi.org/10.1515/lingty-2020-2062>
- Moens, M., & Steedman, M. (1988). Temporal Ontology and Temporal Reference. *Computational Linguistics*, 14(2), 15-28.
<http://www.anthology.aclweb.org/J/J88/J88-2003.pdf>
- Moerdler, G.D, McKeown, K & Ensor. J.R. (1987). Building a Natural Language Interface to Expert Systems. *Proc. of the 10th International Conference on Artificial Intelligence*.
- Ovu, Basil F. (2011). Lexical Sources of Ambiguity in English and Daily Communication. *Oluoha Journal of Languages*, 1(1), 15-26
- Palmer, M. (1985). *Drivillg Semantics for a Limited Domaill*. [Doctoral Dissertation, University of Edinburg]. Edinburg.

- Palmer, M., Gildea, D., & Kingsbury, P. (2005) The Proposition Bank: An Annotated Corpus of Semantic Roles. *Computational Linguistics*, 31(1). 71-105.
- Pazos R, R.A., Martínez F, J.A., González B, J., Morales-Rodríguez, M.L., Galiana B, G.M., Castro H., A. (2008). Ontology-Based Approach for Semi-automatic Generation of Subcategorization Frames for Spanish Verbs. In Corchado, E., Abraham, A., Pedrycz, W (eds), *Hybrid Artificial Intelligence Systems. Lecture Notes in Computer Science*, vol 5271. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-87656-4_69
- Pease, A., Niles, I., & Li, J. (2002). *The Suggested Upper Merged Ontology: A Large Ontology for the Semantic Web and its Applications*. Retrieved September 2, 2021, from <https://www.aaai.org/Papers/Workshops/2002/WS-02-11/WS02-11-011.pdf>
- Pease, A. (2011, Aug 30). *Formal Ontology and the Suggested Upper Merged Ontology (SUMO)*[Video File]. Retrieved September 2, 2021, from <https://www.youtube.com/watch?v=EFQRvyyv7Fs>
- Pinker, P. (1989). *Learnability and Cognition: The Acquisition of Argument Structure*. MIT Press, Cambridge, Massachusetts.
- Pustejovsky, J. (1991a). The Generative Lexicon. *Computational Linguistics*, 17(4), 409–441.
- Pustejovsky, J. (1991b). The syntax of event structure. *Cognition*, 41(1-3), 47–81. [https://doi.org/10.1016/0010-0277\(91\)90032-Y](https://doi.org/10.1016/0010-0277(91)90032-Y).
- Pustejovsky, J. (1995). *The Generative Lexicon*. Cambridge, MA.: The MIT Press.
- Pustejovsky, J & Boguraev, B (1993). Lexical knowledge representation and natural language processing. *Artificial Intelligence* 63(1-2), 193-223.
- Ray, P. Sloka. Ray, Lila. Hai, Mohammed Abdul. (1966). *Bengali language handbook*. Washington, D.C. Center for Applied Linguistics.

- Quillian, M. R (1968). Semantic memory. In M. Minsky (ed.) *Semantic Information Processing*. Cambridge, MA: The MIT Press.
- Schalley, A. (2019). *Ontologies and ontological methods in linguistics*. Retrieved September 2, 2021, from <https://onlinelibrary.wiley.com/doi/10.1111/lnc3.12356>.
- Schalley, A. & Zaefferer, D. (2007). Ontolinguistics – An outline, in Schalley, A. & Zaefferer, D. (eds), *Ontolinguistics. How ontological status shaped the linguistic coding of concepts* (3-22). Mouton de Gruyter.
- Schank, R (1975). *Conceptual Information Processing*. North-Holland Publishing.
- Schulte im Walde, S. (2009). The induction of verb frames and verb classes from corpora. In Anke Lüdeling and Merja Kytö (eds), *Corpus Linguistics. An International Handbook*. Mouton de Gruyter.
- Srsihti, Richa. (2011). *Hindi Verb Classes and Their Argument Structure Alternations*. Cambridge Scholars Publishing.
- Sowa, J. (2010, Nov 30). *Ontology*. Retrieved September 2, 2021, from <http://www.jfsowa.com/ontology>.
- Tesnière Lucien. (1959). *Éléments de syntaxe structurale*. C. Klincksieck.
- Thompson, H.-R. (2003). *Toward a Definitive Grammar of Bengali- A Practical Study and Critique of Research Selected Grammatical Structures*. [Doctoral dissertation. South Asia Department, School of Oriental and African Studies]. London.
- Thompson, H.-R. (2010). *Bengali: A Comprehensive Grammar* (1st ed.). Taylor and Francis.
- Trask, R.L. (1993). *A Dictionary of Grammatical Terms in Linguistics*. Routledge. London
- Van Hout, A. (2000). Event semantics in the lexicon-syntax interface: Verb frame alternations in Dutch and their acquisition. In C. Tenny & J. Pustejovsky (eds.), *Events as grammatical objects*, 239-282. CSLI, Stanford.

Vendler, Z. (1967). *Linguistics and Philosophy*. Cornell University Press.

Vossen, P. (1998). Introduction to EuroWordNet. *Computers and the Humanities*, 32(2/3), 73-

89

Wilks, Y (1975). A preferential pattern seeking semantics for natural language inference.

Artificial Intelligence, 6, 53-74



Ministry of Education
Government of India



The 44th International Conference of Linguistic Society of India



Centurion
UNIVERSITY
Shaping Lives...
Empowering Communities

icolsi 44
November 24 - 26, 2022

Certificate



This is to Certify that Ms/Mr/Dr. **ARUNAVA KAR**
of **CALTS, UNIVERSITY OF HYDERABAD** has participated / presented a paper/
chaired a session in the 44th International Conference of the Linguistic Society of India, organized by
Department of English, Centurion University of Technology and Management (CUTM), Odisha and
Linguistic Society of India (LSI), Pune from 24 to 26 November 2022.

The title of the Sole/Co-author's paper is **DEVELOPING BENGALI FRAME NETWORK**

Biswanandan Dash
Organising Secretary, ICOLSI 44

G. Umamaheshwar Rao
President, LSI, Pune

D. P. Pattanayak
Chairperson, ICOLSI 44

Supriya Pattanayak
Vice Chancellor, CUTM



**Centre for Outreach and Digital Education
Indian Institute of Technology Madras**

This is to certify that

Arunava Kar

has presented a paper titled

Positioning Bengali Passives within the Framework of Impersonal Construction: A Verb Frame Approach

in the Seminar "All about 'Voice': A Crosslinguistic Perspective"

organized by the Department of Humanities and Social Sciences

from 14th to 15th April, 2023

Dr. Anindita Sahoo
Seminar Coordinator
Department of
Humanities and Social Sciences

Prof. Jyotirmaya Tripathy
Head
Department of
Humanities and Social Sciences

Prof. Devendra Jalihal
Chairman
Centre for Outreach and
Digital Education



Graduate Research Meet

Department of Humanities and Social Sciences
Indian Institute of Technology, Guwahati



This is to certify that.....*Arunava Kar*.....

PhD/ MPhil. Scholar,*University of Hyderabad*.....

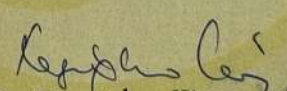
presented a paper titled,

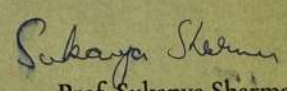
.....*The Riddle of Impersonal Subject Construction in*.....
.....*Bengali : An Alternative Descriptive Approach*.....

at the Graduate Research Meet held on 6 and 7 January 2023 organised by Department of Humanities and Social Sciences, IIT Guwahati.


Meledathu Thomas Kuriakose
Student Convener




Dr. Ngamjahao Kipgen
Faculty Convener


Prof. Sukanya Sharma
Head of Department
Humanities and Social Sciences
IIT Guwahati



**Volume 4
IC3C PROCEEDING**

**Proceeding
of
International Conference
on
Communication, Computation, And Cognition**

**Edited
By
Soumya Sankar Ghosh & Samir Karmakar**

© 2021 by the author(s) of the respective articles

No part of this material protected by this copyright notice may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying, recording or by any information storage and retrieval system, without written permission from the copyright owner(s).

**Editorial Board:
Atanu Saha, Jadavpur University
Indranil Dutta, Jadavpur University
Samir Karmakar, Jadavpur University**

**Managing Editor:
Samir Karmakar, Jadavpur University**

TABLE OF CONTENTS

Content	Page Number
Preface	i
Steering Committee	ii
About VIT Bhopal	iv
About Jadavpur University	iv
Acknowledgments	v
Messages	vi

Sl. No.	Paper Name	Page
1	A Study on the Affective Factors in Writing Skills of the Second Language Learners at Tertiary Level. <i>K. Jefferson and V. Radhakrishnan</i>	1
2	Metaphors of Meat: An Eco-Critical Discourse Analysis <i>Diva Patel and Mona Parakh</i>	8
3	Hindi and Magahi: A Comparative and Quantitative Analysis <i>Supriya Chauhan and Prarabdh Mani Tiwari</i>	22
4	Teaching Pronunciation to the ESL learners - Tasks and Methods <i>J. Rajini</i>	34
5	Analysing Readability and CEFR Level of English School Textbooks in India <i>Princy Pappachan</i>	40
6	Comprehension of Concepts in Children Below the Age of Three: A Study Based on Malayalam Speaking Community. <i>Sreeja T D and Mythila Raja</i>	55
7	Teaching English as a Second Language: A Study of Pedagogical Approaches Adopted in U.P. Board Schools <i>Syed Abid Zaki and Usman Khan</i>	62
8	Taste-Terms in Gujarati: A Cognitive Semantic Analysis <i>Avani Wakhale and Ajay Sarvaiya</i>	76
9	Why case <i>Sayantani Banerjee</i>	86
10	Ordering of Events in Mental Space of the Reader <i>Sayani Banerjee and Samir Karmakar</i>	104
11	How Well Corpus based Measures Reflects Human Perception of Semantic Similarity: An Experiment in Bangla <i>Rakesh Dutta, Manjira Sinha and Tirthankar Dasgupta</i>	119
12	Numerical Classifier '-t ^h e' or '-t ^h o' or '-go' in Bhojpuri <i>Rupkatha Mukherjee and Salman Alam</i>	130
13	Evaluating English Language Teaching (ELT) and Learning in Odia Medium Educational Institutions in Odisha in the Light of Emerging Models of Second Language Pedagogy <i>Pratap Kumar Dash</i>	141
14	Bhojpuri WordNet : A Preliminary Investigation <i>Imran Ali</i>	152
15	Use of Ontology in Bengali Verb Frame and its importance in Verb Sense Disambiguation <i>Arunava Kar</i>	161
16	Plural Formation in the Telangana Telugu: A Contrastive Study <i>P. Prakash</i>	173
17	Challenges of English Language: A Ditch between the Contemporary High School Education and Higher Education in India. <i>Fazil</i>	182



Use of Ontology in Bengali Verb Frame and its importance in Verb Sense Disambiguation

Arunava Kar

University of Hyderabad, India

Keywords:

Ontology, Verb Semantics, Bengali Verb Frame, Verb – Sense Disambiguation

Abstract

In simplest terms, verb frame is the systematic representation of syntactic preferences of a verb. Verbs frames are used as lexical resources for various NLP applications such as parser, machine translation, language generation etc. We start from the assumption that verb's syntactic behavior is determined by its meaning (Levin, 1990). We are developing a Bengali Verb Frame Network which can potentially be used for many NLP purposes, especially verb sense disambiguation. Our verb frame captures the relevant structural context in which verb occurs in a sentence. For an ambiguous verb, we hope to get different verb frames for each verb sense and thereby map meaning difference to structural difference. Each verb sense is treated as a separate entry in our verb frame. Our verb frame represents the relevant features of verbal arguments. Ontology, in information science, is usually referred to any hierarchical organization of concepts which is used to classify, analyse or annotate any data. Our verb frame has four columns Dependency Roles, Argument Types, Vibhakti/Post-positions and Ontology. We have developed an upper ontology scheme for our verb frame. Though we have not directly adapted any existing ontology model, we have consulted Suggested Upper Merged Ontology (SUMO) for the background idea as it is already widely used in NLP research and linked to many databases like EuroWordNet. Ontology shows what the verbal argument we are trying to annotate, means in real world thereby incorporate real world knowledge with structural information. This kind of real world knowledge is very important for meaning disambiguation. This paper describes the role of ontology in our verb frame and how it helps in verb sense disambiguation. From machine learning perspective, we argue here that ontology is the most suitable way of providing real world information.

1 Introduction: -

The term ontology has different connotation in different domain of knowledge. The term has originated in philosophy where it means the study of being in general. From knowledge representation context, Sowa defines ontology as “a catalog of the types of things that are assumed to exist in a domain of interest D from the perspective of a person who uses a language L for the purpose of talking about D” (Sowa, 2010). Similarly, in linguistics, ontology is usually thought of an interconnected network of concepts. Hence, it is quite clear that ontology is regarded as a way of representing human conceptualization in all these domains. In artificial intelligence, Gruber defined ontology as an “explicit specification of a conceptualization” (Gruber, 1993, p. 908) where conceptualisation “is an abstract, simplified view of the world that we wish to represent for some



purpose”. In other words, from machine learning point of view, ontology is usually a formal structured schema which organises real world objects and abstractions in a hierarchical order.

Here, we will consider ontology as a technology resource, in contrast to that of a philosophical concept. In this regard, ontologies are usually of two types. First one is called **domain-specific ontology**. It represents hierarchy of concepts in a particular area of knowledge like physics or biology. Second is known as **upper ontology** or universal ontology or top ontology. These are general purpose ontologies seeking to represent a vast area of knowledge. We are developing ontological framework in order to use it in Bengali verb frame which, in turn, will be used for Verb Sense Disambiguation. Ontology is now becoming an important aspect of knowledge-engineering and information science research. Most WordNets either directly use some kind of ontology framework or is mapped to any established upper ontology. Most noted ontological models are developed in some sort of machine readable format. For exploitation of language data for computational processing purpose, ontology is absolutely important. Ontological tagging shows what a linguistic item refers to in the real world. Whenever we use language, we always process ontological information without really noticing it. For, it is the part of our innate world knowledge. But computer do not possess any such information. The point of ontology, especially in NLP, is to provide some kind of world knowledge so that the machine is able to connect the real world entity with the linguistic item. We are developing Bengali Verb Frame Network and the ontological format which we are discussing here is a part of this Bengali Verb Frame Network. In this paper, I will describe our verb frame model with special focus on the ontology section.

When a verb has multiple possible senses, the system has to disambiguate between all of them and identify its correct meaning as per the context. We will start from the basic assumption that a verb's syntax is determined by its meaning (Levin, 1993). Hence, we expect to get a different verb frame for each of the senses of a particular verb. The proposed Bengali verb frame lexicon is planned to capture fine-grained, context -sensitive information associated with each sense of a verb. Purpose of this Bengali Verb Frame Network is to help NLP tools in sense disambiguation.

In simplest terms, verb Frame is a systematic representation of syntactic preferences of a verb. Verb frame portrays the syntactico-semantic behavior of a verb. Our verb frame is divided in four sections which are a) Dependency Roles b) Argument Type c) Vibhakti/Post-position d) Ontology. This paper will present development, analysis and critical estimation of our ontology model within the complete verb frame model. Verbs frames are used as lexical resources for various NLP applications such as parser, machine translation, language generation etc. This work will be an important database for NLP tools like parsers, question answering system etc.

2 Framework: AnnCorra and Paninian Computational Grammar:-

This section seeks to elaborate the theoretical framework which is Paninian dependency grammar. Three sections of our verb frame namely Dependency Roles, Argument Types and Vibhakti/Post-positions is directly related to Paninian dependency grammar. Forth and last section is Ontology which is the focus of this paper is not derived from Paninian Dependency Grammar. We have built a separate framework for Ontology which will be discussed in the later sections. Paninian Grammar (PG) is a dependency grammar that treats sentence as a series of modifier-modified relations. It was developed by Panini (4th century BC) for Sanskrit language in ancient India. Panini produced a linguistic theory for analysis of sentences (Kiparsky and Staal, 1969). In Paninian dependency, verb is in the center of all dependency relations. Most important part of Paninian

grammar is its karaka relations. Karaka is the structural-semantic relation that connects a verb with the nominal elements (nouns and pronouns) in a sentence. Paninian Grammar specifies 6 karaka roles i.e,

1. Karta
2. Karma
3. Karana
4. Sampradana
5. Apadana
6. Adhikarana

These karaka relations described by Panini, to some extent, correspond to the thematic roles. We can map them to thematic roles in this way

- (1) karta (k1) 'agent/theme/experience/actor'
- (2) karma (k2) 'theme/patient'
- (3) karana (k3) 'instrument'
- (4) sampradāna (k4) 'recipient'
- (5) apādāna (k5) 'source'
- (6) adhikarana (k7p) 'location'

Karaka roles are not completely semantic in nature. There can be large number of high level semantic relations. But the karaka roles are fixed number of relations which capture only a certain level of semantics. Though we can roughly map the last four karakas to their thematic counterpart, the '*karta*' and '*karma*' are entirely different from agent and theme. These two karakas are roughly equivalent to syntactic criteria like sentential subject and sentential direct object respectively (Bharati, Chaitanya & Sangal, 1995). This is because thematic roles are purely semantic whereas karaka relations are syntactico-semantic in nature.

Based on Paninian Computational grammar, scholars from IIIT-Hyderabad and others has developed a multi level dependency based tagging scheme named AnnCorra (Bharati, Sharma, Bai, & Sangal, 2006). AnnCorra consists of three different tag sets 1. PoS tags 2. Chunk tags 3. Dependency tags. AnnCorra dependency tags have six karaka roles as well as many non-karaka roles. Our dependency tags are based on AnnCorra. Very limited modifications are done where it is absolutely necessary. We have followed AnnCorra tagging scheme for Dependency Roles annotation and Argument Type annotation in the development of our verb frame.

3 Methodology: -

Objective of our research, as already stated, is to create lexical knowledgebase in order to disambiguate among the many senses of an ambiguous verb. Our basic assumption is meaning determines syntactic behavior of a verb. Hence, difference in meaning suppose to cause difference in syntactic behavior. Our verb frame is designed to capture the totality of contextual information in which the given verb is likely to occur. Since, context is dependent on verb's meaning, for ambiguous verbs, we expect to get different verb frame for each of its meaning. In other words, difference in verb frame in different context reflects different meaning of the same verb.

3.1 The primary resources

The primary resources that will be used for this study are:-

1. Bengali monolingual Dictionary

2. A Bengali WordNet
3. Corpus
4. Verb Frame Model

Dictionary is needed to get the verbs. Since, some dictionaries enlist different senses for each lexical items, it can also be used in order to identify ambiguous verbs. I have selected one Bengali monolingual dictionary ‘Adhunik Bangla Obhidhan: Bangla Academy’ and collected more than 100 verbs which has more than one senses according to the dictionary. WordNet is required for sense information. I am using Indo-WordNet which includes Bengali WordNet as well. One simply has to search on Indo WordNet website with the verb already collected from dictionary to see the different meanings with example. Corpus is the most basic resource for this research. Corpus will be consulted to get the syntactic distribution. We have made use of TDIL corpus which comprises of over hundred thousand Bengali sentences. ‘Itrans’ scheme is used for transliteration.

3.2 Architecture of the Verb Frame:-

Each Verb frame includes a number of basic and necessary information related to that particular sense. In addition to the syntactico-semantic information provided in the verb frame other details like the following are also provided. Our verb frame consists of two sections 1.General Information 2.Verb Frame. General Information section includes some basic details about the verb sense, such as:

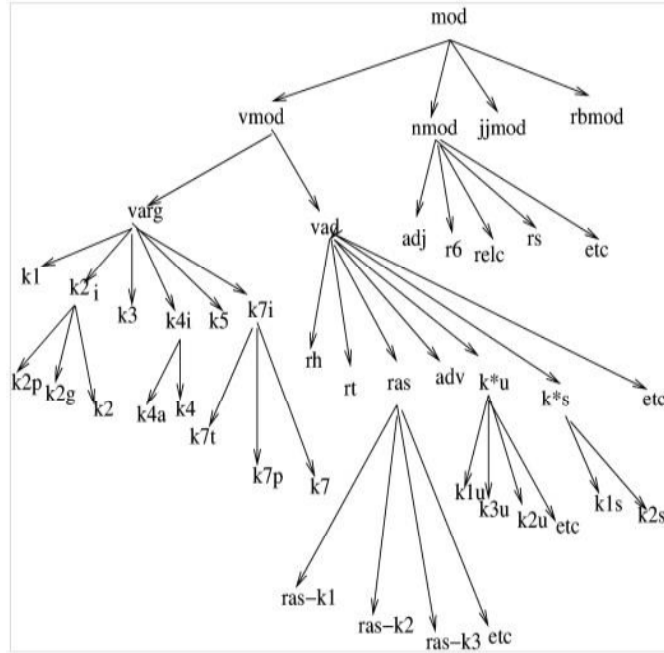
1. Sense ID
2. English Gloss
3. Verb Class
4. Verbs in the Same Class
5. Example Sentence

The second section is Verb Frame. It includes:

1. Dependency relations
2. Argument Type
3. Vibhakti/post-position
4. Ontology

The verb frame is developed in tabular format. It has four columns: Each one provides a different type of information. I have includes two complete examples of our verb frame at the end of this section.

First column is **Dependency relations**, the second column: **Argument type**, third column: **Vibhakti/Post-position**, fourth column: **Ontology**. We have mostly adopted AnnCorra (Bhrati et al 2006) dependency tags with little modification. Bengali has certain constructions which cannot be properly annotated by using existing AnnCorra tags. We have either modified existing tags or added new tags only in these cases. AnnCorra dependency tags are given below in a diagram.



AnnCorra Dependency Tags

Argument type is the second column of our verb frame model. Here, we have used AnnCorra chunk roles to tag argument types. We have just added one tag label i.e. ‘Cl’ for clause as AnnCorra do not have any tag for clause. Each chunk is named according to the corresponding phrasal category. For example, noun chunk will be tagged as NP. The tags which we have used to label Argument Type are NP (noun chunk), VGINF (infinitival verb chunk), VGNN (gerunds), JJP (adjectival chunk), RBP (adverb chunk) and CL (clause)

Vibhaktis or case endings are the suffixes which marks karaka. **Post-positions** perform many other functions apart from signaling dependency relations. Since numbers of vibhaktis are limited in Bengali, post-positions mark dependency roles where vibhakti is not available. Many karaka relations and all non-karaka relations are marked by post-positions.

Ontology is the forth column of our verb frame. We have developed a novel ontology schema for our verb frame. It will be described in detail in the very next section

```

SENSE_ID      deoya_1
SENSE        'to pay'
VERB_CLASS    4 (k1+k2+k4)
VERBS_IN_THE_SAME_CLASS  balaa_1, deoyaa_3,deoyaa_4,deoyaa_5,deoyaa_6

EXAMPLE_SENTENCE  o  ei  kajer  jonno  amay  tiris  haajaar  Taakaa  dichChe
                  he this work+gen for me thirty thousand Rs. give+imperfect
                  He is paying me thirty thousand rupee for this work.

VERB_FRAME
DEPENDENCY_RELATIONS      k1          rt          k4          k2
ARGUMENT_TYPE              np          np          np          np
VIBHAKTI/POST-POSITION    0          jonno      0          0
ONTOLOGY                   +human          +human      +cur
                           +org          +org

```

```
SENSE_ID          deoya_2
SENSE             'to put'
INDO_WORDNET_REFERENCE
VERB_CLASS        9 (k1+k2+k7p)
VERBS_IN_THE_SAME_CLASS    melaa_1

EXAMPLE_SENTENCE  egulo nice dao
                  these below put
                  Put these things below

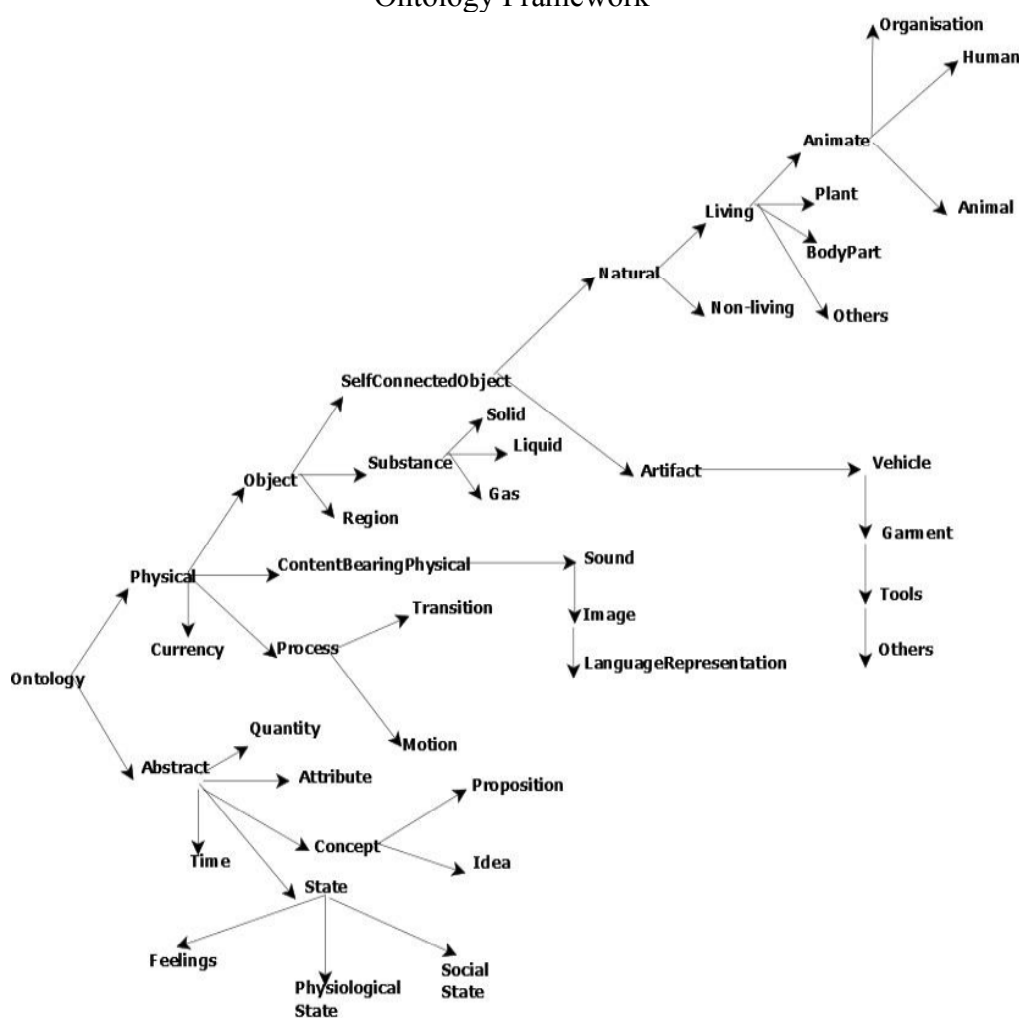
VERB_FRAME
    DEPENDENCY_ROLES          k1
    ARGUMENT_TYPE              NP
    VIBHAKTI/POST-POSITION    0
    ONTOLOGY                   +human
```

Our ontology is, for the most part, based on Suggested Upper Merged Ontology (SUMO). SUMO is an upper ontology, perhaps the largest upper ontology existing in the public domain. According to its developers, SUMO (Pease, Niles & Li, 2002) is created by merging a large number of other existing upper ontologies. Apart from the upper ontology itself, SUMO has a number of separate domain specific ontologies as well. SUMO along with its domain specific ontologies comprises of approximately 25,000 terms 80,000 axioms, 4730 rules and 1280 relations. SUMO is written in SUO-KIF language. Since it is completely written in formal language, it proves to be very useful for the machine learning purposes. SUMO is linked to the various WordNets in many different languages.

We have used the SUMO terms and followed SUMO definition of those terms as much as possible in our modeling. But SUMO is neither created specifically for language analysis nor is it practical to use thousands of terms in our ontology schema. We had to simplify it to suit our purpose. The

motivation behind SUMO is to represent all known concepts and objects in a formal, ordered hierarchical order, whereas we need ontology in order to make ontological generalisation of the verbal arguments. This will, in turn, facilitate us in Verb Sense Disambiguation. Our ontology has 43 terms altogether. We have here taken a bottom up approach in creating this resource. We started from the verb frame and find out what are the ontological concepts we need in order to reach the ontological generalisation regarding verbal arguments. Then we have developed an ontology model in which our necessary concepts are mapped on. SUMO has the concept ‘human’ as well as the concept ‘hominid’. We don’t need a concept like ‘hominid’, but we need a concept for ‘human’ entities. The reason behind this choice is that a large number of verbs like verbs of cognition, verbs of social interaction etc take ‘human’ entities as their agent. There is no such use for the term ‘hominid’. Hence, we have included ‘human’ and discarded most of the intermediate nodes. We are only annotating the Bengali nominal constructions which can act as an argument for the senses of an ambiguous Bengali verb. Hence, we don’t need all concepts. Through we made our ontology very simplistic, it is still an upper ontology and is internally consistent i.e. it still represent a wide range of object and concepts in a consistently way. It works well in consort with the other elements of our verb frame. Our ontology schema is presented below.

Ontology Framework



5 Importance and Issues:

This section will show the significance as well as issues regarding ontology in our verb frame.

5.1 Role of Ontology in Bengali Verb Frame:

Ontology is an added component in our verb frame which is otherwise built on Paninian Dependency Grammar. Task of ontology is to provide some other kind of information which cannot be obtained by structural-semantic analysis of Paninian dependency framework. Dependency roles show the semantic connection between the verb and the argument. On the other hand, ontology tells us what kind of entity the argument is in the real world. Ontological information is not any kind of relationship between verb and the argument, rather a property of the argument itself. Ontological features do not denote a linguistic entity, rather the real world object or concept that the linguistic item signals. In this way, ontology is language independent and universally applicable, at least with regard to the language. We need such language independent resource when we try to systematically decode the real world information packed within linguistic structures. Our basic idea is simple. If a given ambiguous verb has 7 different senses, we are supposed to get 7 different verb frames. Meaning difference must be reflected in the difference of verb frames.

Consider the following examples taken from corpus,

1. tini aamaake Daaklen
he/she me+acc call+per
He/she has called me
2. svaamiike daakla strii
husband+acc call+per wife
Wife called her husband
3. poShaa kukurTike daaklen bRRiddha
pet dog+acc call+per old man
The old man has called his pet dog

The verb sense used in above example is Bengali verb *Daakaa_1* 'to call' as per our Bengali Verb Frame Network, which implies - the activity of verbal calling meant to attract attention of the hearer. This activity needs two elements; the agent who enacts the activity of calling and a patient who is at the receiving end of the activity. As illustrated in the example, the agent or the first argument of the verb *Daakaa_1* is always a human i.e. it is always humans who calls someone (at least in the real world scenario). The second argument or patient can be a human or an animal. Hence, in our verb frame we mark the first argument of the verb sense *Daakaa_1* as '+human' and the second argument as '+human +animal'. This way it is able to capture the generalisation regarding the verb sense in question.

VERB_FRAME		
DEPENDENCY_RELATION	k1	k2
ARGUMENT_TYPE	NP	NP
VIBHAKTI/POST-POSITION	0	ke
ONTOLOGY	+human	+human +animal

Verb Frame for the verb *Daakaa_1*

Now, let us look at the corpus drawn sentences of another verb *Daakaa_2* ‘to make a noise’ which implies the activity of animals making a sound in Bengali.

4. gaadhaa theke-theke Daakchila
 donkey at times sound+past+pro
 The donkey was braying

5. kukur gheu gheu kare Daakche
 dog bark post-position sound+pro
 The dog is barking

Similarly, these examples clearly show that the agent of the verb *Daakaa_2* can only be an animal. Therefore, we have marked it as ‘+animal’ in our verb frame.

VERB_FRAME	
DEPENDENCY_RELATIONS	k1
ARGUMENT_TYPE	NP
VIBHAKTI/POST-POSITION	0
ONTOLOGY	+animal

Verb Frame for the verb *Daakaa_2*

In the above examples, the contextual difference between two verb senses is clearly revealed in the form of tabular verb frame. Machine can make use of such verb frame to identify the correct sense form the context. Here verb frame is different in terms of both the dependency relations as well as ontology. But, sometimes two senses of an ambiguous verb appear to have identical dependency structure. Then, ontology becomes the most important marker of contextual difference. Reverse is also true. In other words, ontological information compliments dependency information. Often, ontology presents novel aspects of the context which cannot otherwise be revealed.

5.2 Issues and Limitations:

Adam Pease in a 2011 lecture said “creating a large ontology does consist of a lot of trivial assertions about the world. Things that people already know. But, the fact is computers do not know them” (Pease, 2011). He was then taking about one of the primary criticisms directed at SUMO. Ontologies are usually built for machine learning purposes. The information that ontologies represent may not appear very unique to us, but computer needs this kind of world knowledge to perform NLP tasks. Our ontology is specifically developed to categorise nominal arguments of Bengali verbs. The objective is to arrive at some categorical generalisations regarding the argument themselves which is otherwise consistent with the dependency structure

information and compliments the same. Though our ontology performs well in its task it has some limitations. One of the limitations of our ontology is that it does not cover all scenarios. Some limitations are coming from the purpose of the ontology itself. For example, our ontology cannot account for a supposed fantasy fictional or mythological situation where everything, even a stone, can talk. We cannot include this kind of fantasy or mythology scenarios in our ontology because then it will be open-ended then and it will no longer be used for meaning disambiguation. For, in fantasy, everything is possible. We won't be able to find necessary constraints on which we can base our generalisation. There is another kind of limitations which is more subtle. Our ontology will not be able to take care of the scenario where a robot or speech generation system is talking. Our ontology is an upper ontology; it is not designed to take care of such specific situation like robotic speech. This information, if necessary, can be conveyed to the machine by other means. We can also develop a domain specific ontology for example, ontology of speech technology, to handle such scenarios. Purpose of this ontology is to formally represent the fact that when the activity of talking takes place in real world, it is the usually the human beings who are involved. Hence, the ontological information in Bengali verb frame provides categorical generalisation regarding the verbal arguments. But the generalisation assumes and projects real world context; it may not represent any specific utterance which represent a rare unusual context.

7 Conclusion:

The research initiative which is hereby being discussed is of applied nature. The motto of this initiative is to prescribe a framework aimed at resolving Verb Sense Disambiguation. As already discussed sense disambiguation is required for increasing the accuracy and precision of various NLP applications and it is most useful in machine translation. Though one of the main objectives of this project is to create resources for computational linguists, researchers of other areas like syntax, generative semantics, translation theory etc may also be benefitted from this resource. We believe this research can be equally beneficial for language teachers and second language learners of Bengali.

Ontology is a section in our verb frame. Our Bengali Verb Frame is still under development. Though it is aimed at resolving verb sense disambiguation issue which is mostly a NLP problem, it can offer fresh insight to the theoretical linguists as well. As per my knowledge, there was no previous initiative to create any such resource for Bengali. Verb frame has been developed in Hindi and few other Indian languages. But, those resources did not try to include any ontological framework. In this way, it is a novel development in verb frame research. Ontology is a section in our verb frame. Existing ontologies are not developed for this kind usage. Hence, we had to drastically reduce the ontological features and modify some of the features. Our ontology has become a novel ontological framework most suitable for semantic analysis. Here, in our research project, it reveals completely different aspect of verb semantics which is clearly not accessible through Paninian Dependency framework. Ontology has proved to be a very valuable section in our research. As ontology is language independent, it may be applied to similar projects on other languages with or without major modification.

References:

- Baker, C & Ruppenhofer, J. (2002). FrameNet's Frames vs. Levin's Verb Classes. In J. Larson and M. Paster (Ed). *Proceedings of the 28th Annual Meeting of the Berkeley Linguistics Society*, 27-28.
- Begum, R., & Sharma, D. M. (2017). Development and Analysis of Verb Frame Lexicon for Hindi. *Linguistics and Literature Studies* 5(1), 1-22. DOI: 10.13189/lis.2017.050101
- Bharati, A., Sharma, D. M., Bai, L., & Sangal, R., (2006). AnnCorra : Annotating Corpora Guidelines For POS And Chunk Annotation For Indian Languages. *LTRC, IIIT Hyderabad*.
- Bharati, A., Chaitanya, V., & Sangal, R. (1995). *Natural Language Processing: A Paninian Perspective*. New Delhi, India: Prentice Hall of India
- Chakrabarti, D., & Bhattacharyya, P. (2004). Creation of English and Hindi Verb Hierarchies and their Application to Hindi WordNet Building and English-Hindi MT. *Proceedings of the International Conference on Global Wordnet (GWC 04), Brno, Czech Republic*. 83-90.
- Comrie, B. (1985). Causative verb formation and other verb-deriving morphology. T. Shopen (ed.) *Language typology and syntactic description, Vol 3*, (pp. 309-348.) Cambridge: Cambridge University Press
- Gruber, T. R. (1993). Toward principles for the design of ontologies used for knowledge sharing. *International Journal Human-Computer Studies* 43(5-6), pp. 907-928. <https://doi.org/10.1006/ijhc.1995.1081>.
- Jackendoff, J. (1990). *Semantic Structures*. Massachusetts: MIT Press
- Kingsbury, P., & Palmer, M. (2002). From Treebank to PropBank. *Proceedings of the 3rd International Conference on Language Resources and Evaluation, Las Palmas, Canary Islands, Spain*, pp. 1989–1993.
- Kipper, K., Dang, H., & Palmer, M. (2000) Class-based construction of a verb lexicon.. *AAAI/IAAI*, 691-696.
- Kipper, K. (2005). *VerbNet: A broad-coverage, comprehensive verb lexicon*. Ph.D. thesis, Computer and Information Science Dept., University of Pennsylvania, Philadelphia, PA.
- Kiparsky, P., & Staal, J.F. (1969). Syntactic and Semantic Relations in Panini. *Foundations of Language*. 5, 83–117.
- Levin, B., & Rappaport Hovav, M. (1995). *Unaccusativity: At the Syntax-Lexical Semantics Interface*. Retrieved September 2, 2021, from <https://hdl.handle.net/2027/heb.08443>.
- Levin, B. (1993). *English Verb Classes and Alternations: A Preliminary Investigation*. Chicago: University of Chicago Press.

Miller, G. A., Beckwith, R., Fellbaum, C. D., Gross, D., & Miller, K. (1990). WordNet: An online lexical database. *International Journal of Lexicography*, 3(4), 235–244.

Palmer, M., Gildea, D., & Kingsbury, P. (2005). The Proposition Bank: An Annotated Corpus of Semantic Roles. *Computational Linguistics*, 31(1), 71–105.

Pease, A., Niles, I., & Li, J. (2002). *The Suggested Upper Merged Ontology: A Large Ontology for the Semantic Web and its Applications*. Retrieved September 2, 2021, from <https://www.aaai.org/Papers/Workshops/2002/WS-02-11/WS02-11-011.pdf>

Pease, A. (2011, Aug 30). *Formal Ontology and the Suggested Upper Merged Ontology (SUMO)* [Video File]. Retrieved September 2, 2021, from <https://www.youtube.com/watch?v=EFQRvyyv7Fs>

Pinker, P. (1989). *Learnability and Cognition: The Acquisition of Argument Structure*. MIT Press, Cambridge, Massachusetts.

Schalley, A. (2019). *Ontologies and ontological methods in linguistics*. Retrieved September 2, 2021, from <https://onlinelibrary.wiley.com/doi/10.1111/lnc3.12356>.

SCHALLEY, A. & ZAEFFERER, D., (2007), Ontolinguistics – An outline, in Schalley, A. & Zaefferer, D. (eds), *Ontolinguistics. How ontological status shaped the linguistic coding of concepts* (3–22). Berlin: Mouton de Gruyter.

Sowa, J. (2010, Nov 30). *Ontology*. Retrieved September 2, 2021, from <http://www.jfsowa.com/ontology>.

Vossen, P. (1998). Introduction to EuroWordNet. *Computers and the Humanities*, 32(2/3), 73–89.

Building Bengali Verb-frame Network: A Paninian Dependency Approach

by Arunava Kar



Librarian
Indira Gandhi Memorial Library
UNIVERSITY OF HYDERABAD
Central University P.O.
HYDERABAD-500 046.

Submission date: 30-Jun-2023 03:10PM (UTC+0530)

Submission ID: 2124704761

File name: Arunava_Kar.pdf (1.13M)

Word count: 33490

Character count: 174592

Building Bengali Verb-frame Network: A Paninian Dependency Approach

ORIGINALITY REPORT

3%

SIMILARITY INDEX

2%

INTERNET SOURCES

2%

PUBLICATIONS

1%

STUDENT PAPERS

PRIMARY SOURCES

1	pt.scribd.com Internet Source	<1 %
2	dokumen.pub Internet Source	<1 %
3	www.gabormelli.com Internet Source	<1 %
4	en.wikipedia.org Internet Source	<1 %
5	kipdf.com Internet Source	<1 %
6	verbs.colorado.edu Internet Source	<1 %
7	Lecture Notes in Computer Science, 2012. Publication	<1 %
8	www.lotpublications.nl Internet Source	<1 %
9	scholar.sun.ac.za Internet Source	<1 %

10

f.javier.io

Internet Source

<1 %

11

"Chinese Lexical Semantics", Springer Science and Business Media LLC, 2018

Publication

<1 %

12

Benjamin Lyngfelt. "Chapter1. Introduction", John Benjamins Publishing Company, 2018

Publication

<1 %

13

Pustejovsky, J.. "Lexical knowledge representation and natural language processing", Artificial Intelligence, 199310

Publication

<1 %

14

Martha Palmer, Daniel Gildea, Nianwen Xue. "Semantic Role Labeling", Springer Science and Business Media LLC, 2010

Publication

<1 %

15

"Ontologies", Springer Science and Business Media LLC, 2007

Publication

<1 %

16

Andrea C. Schalley. "Cognitive Modeling and Verbal Semantics", Walter de Gruyter GmbH, 2004

Publication

<1 %

17

Nikolaos Lavidas, Kiki Nikiforidou. "Studying Language Change in the 21st Century", Brill, 2022

Publication

<1 %

18	hdl.handle.net Internet Source	<1 %
19	Hooi Ling Soh, Hiroki Nomoto. "The Malay verbal prefix meN- and the unergative/unaccusative distinction", Journal of East Asian Linguistics, 2011 Publication	<1 %
20	doc.lagout.org Internet Source	<1 %
21	"Chinese Lexical Semantics", Springer Science and Business Media LLC, 2016 Publication	<1 %
22	"Semantics and the Lexicon", Springer Science and Business Media LLC, 1993 Publication	<1 %
23	Thierry Fontenelle. "Turning a Bilingual Dictionary into a Lexical-Semantic Database", Walter de Gruyter GmbH, 1997 Publication	<1 %
24	Submitted to Middle East Technical University Student Paper	<1 %
25	acl.ldc.upenn.edu Internet Source	<1 %
26	epdf.pub Internet Source	<1 %

27

theses.gla.ac.uk

Internet Source

<1 %

28

Submitted to University of Hyderabad,
Hyderabad

Student Paper

<1 %

29

www.aui.ma

Internet Source

<1 %

Exclude quotes On

Exclude matches < 14 words

Exclude bibliography On