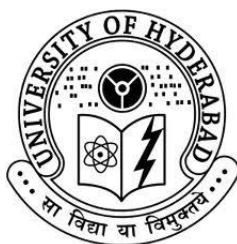


Verb Sense Disambiguation in Nepali Using Verb Frames

*A thesis submitted during 2023 to the University of Hyderabad in partial
fulfillment of the award of a Ph.D. degree in Applied Linguistics from the
Center for Applied Linguistics and Translation Studies*

By:
Krishna Maya Manger
Reg. No. 15HAPH04



Centre for Applied Linguistics and Translation Studies
School of Humanities

University of Hyderabad
P.O. Central University, Gachibowli,
Hyderabad – 500046
Telangana, India

Verb Sense Disambiguation in Nepali Using Verb Frames

*A thesis submitted during 2023 to the University of Hyderabad in partial
fulfillment of the award of a Ph.D. degree in Applied Linguistics from the
Center for Applied Linguistics and Translation Studies*

By:

Krishna Maya Manger

Reg. No. 15HAPH04

Supervisors

Dr. K. Parameswari

Assistant Professor, CALTS, UoH

Dr. S. Arulmozi

Professor, CALTS, UoH



Centre for Applied Linguistics and Translation Studies
School of Humanities

University of Hyderabad
P.O. Central University, Gachibowli,
Hyderabad – 500046
Telangana, India

DECLARATION

I, **Krishna Maya Manger** hereby declare that this thesis entitled *Verb Sense Disambiguation in Nepali using Verb Frames* submitted by me under the guidance and supervision of Dr. K. Parameswari and Dr. S. Arulmozi is a bonafide research work. I also declare that it has not been submitted previously in part or in full to this University or any other University or Institution for the award of any degree or diploma.

Date: 22/09/2023

Name : **Krishna Maya Manger**



Signature of the Student:

Reg. No.: 15HAPH04



CERTIFICATE

This is to certify that the thesis entitled *Verb Sense Disambiguation in Nepali* using *Verb Frames* submitted by Mrs. Krishna Maya Manger, bearing Registration Number 15HAPH04 in partial fulfillment of the requirements for award of Doctor of Philosophy in Applied Linguistics from the School of Humanities is a bonafide work carried out by her under our supervision and guidance.

This thesis is free from Plagiarism and has not been submitted previously 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(s) 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:

- i) 'Argument Structure of Nepali Verbs: A study on Lexico-Semantic Divergence' in *Journal of Advanced Linguistic Studies*, Vol.7:1-2, Jan-Dec 2018 (ISSN 2231-4075) which is appeared in Chapter 4 of her dissertation.
- ii) 'Development of Verb Frames for Nepali' in *Language in India*, Vol. 23:6, June 2023 (ISSN 1930-2940), which is appeared in Chapter 4 of her dissertation.

And she has made presentations in the following conferences:

- i) 11th Student's Conference of Linguistics in India (National) at School of Languages and Linguistics, 28-30th January 2017.
- ii) 39th Annual Conference of Linguistic Society of Nepal (International) held at CEDA hall, Tribhuvan University, Kathmandu, Nepal, 26-27th November 2018

Further, the student has passed the following courses towards fulfilment of coursework requirement for Ph.D:

Course Code	Name	Credits	Pass/Fail
1. AL-801	Research Methodology	4	Pass
2. AL-702	Current Trends in Applied Linguistics	4	Pass
3. AL-721	Research Oriented Readings	4	Pass
4. AL-722	Machine Translation	4	Pass

Supervisors

Head of the Department

Dean of School

ACKNOWLEDGEMENTS

First and foremost, I wish to express my profound gratitude to my esteemed supervisor, Dr. K. Parameswari, for her unwavering guidance and support throughout my research journey, spanning from 2015 to the present day. I am deeply indebted to her for making my research endeavors smoother with her patient and composed demeanor, addressing every facet of my thesis. Without her kind and diligent guidance on the computational aspects of my research, achieving my research goals would have been an insurmountable task.

I would also like to extend my appreciation to my supervisor Dr. S. Arulmozi for his invaluable assistance in identifying the shortcomings in my thesis and enlightening me with his constructive comments and suggestions.

I convey my heartfelt thanks to the esteemed Prof. G. Uma Maheswar Rao, a member of my doctoral committee, for his generous support and recommendations in selecting my research area and title. I am equally appreciative of his insightful comments on my research at various junctures.

My gratitude extends to all the faculty members at CALTS, UoH, and the dedicated non-teaching staffs for their support in both teaching and administrative matters. I also want to acknowledge the invaluable assistance provided by the administrative staffs in the UoH administrative section who have guided me at different stages of my research.

I hold in high regard to my revered guru, Prof. Ghanashyam Nepal, from the Department of Nepali at the University of North Bengal. His insightful guidance before the commencement of my research was instrumental in my decision to pursue linguistics and embark on a PhD outside my previous university. His visionary direction and steadfast support have been transformative in my academic journey, and I owe him a profound debt of gratitude.

I extend my heartfelt thanks to my colleagues and mentors at the Department of Nepali, University of North Bengal. In particular, I would like to express my gratitude to the department head, Dr. Pushkar Parajuli, for his constant encouragement and guidance throughout my research. I also appreciate the understanding and cooperation of esteemed professors such as

Prof. Sanjay Rai, Prof. K.R. Ghatani, and my dear colleague, Dr. Dup Tshering Lepcha, during my study leave.

Furthermore, I wish to acknowledge Dr. Archit Yajnik, Associate Professor at the Department of Mathematics, Sikkim Manipal Institute of Technology, for his assistance with the Nepali parser he developed, which was instrumental in validating my data.

I extend my sincere thanks to Mr. Ashish Tamang, a Research Scholar at the Nepali Department, University of North Bengal, for his unwavering support and invaluable assistance throughout the process of drafting and submitting my thesis.

I am also deeply grateful for the contributions of my friend, Mr. Khemraj Timsina, who played a crucial role in reviewing and rectifying the English glosses in my verb frames database.

My heartfelt appreciation goes to Mr. Nagaraju at IIIT, Hyderabad, for his assistance in the evaluation process of my research.

Lastly, I want to express my gratitude to my husband Mr. Amber Rai, my sister Uden, and my entire family especially my younger brother Bimal Manger for their enduring patience and unwavering moral support at every stage of my academic journey and research. Without their steadfast backing, embarking on this extensive educational and research odyssey would not have been possible.

Abbreviations

ACC - Accusative	M – Masculine
AI – Artificial Intelligence	MT – Machine Translation
ASSO – Associative	NEGP - Chunk for Negatives
BLK - Miscellaneous	NEG - Negative
CAUS - Causative	NF – Non Feminine
CCD - Coordinator	NLP – Natural Language Processing
CCP – Conjuncts	NN – Common Noun
CCS – Subordinator	NNP – Proper Noun
CL - Classifier	NNV - Verbal Noun
CONT - Continuous	NOM – Nominative
DMD – Deictic Demonstrative	NST – Noun of Space and Time
DMI - Indefinite Demonstrative	NP – Noun Phrase
DMQ - Interrogative Demonstrative	NPST – Non Past
DMR – Relative Demonstrative	PASS - Passive
ECH - Echo words	PERF – Perfective
ERG – Ergative	PL - Plural
F – Feminine	POS – Parts of Speech
FRAGP - Chunk Fragments	PRC – Reciprocal Pronoun
GEN - Genitive	PRF – Reflexive Pronoun
HAB - Habitual	PRI – Indefinite Pronoun
HH – High Honorific	PRP – Personal Pronoun
ILCI – Indian Languages Corpora	PRQ – Interrogative Pronoun
Initiative	PRL - Relative Pronoun
i.e. – that is	PSP – Post Position
INJ - Interjection	QTC - Cardinals
INTF - Intensifier	QTF - General
JJ - Adjective	QTO - Ordinals
JJP - Adjectival Chunk	RBP - Adverb Chunk
LOC - Locative	RB - Adverb

RDF - Foreign word	VGINF - Infinitival Verb Chunk
RPD - Default	VNG - Gerund
SID - Sense ID	VM – Main Verb
S - Sentence	VF – Finite Verb or Verb Frame
SYM - Symbol	VINF – Infinite Verb
TDIL – Technology Development for Indian Languages	VNF - Non-finite
UNK - Unknown	VP – Verb Phrase
VAUX – Auxiliary Verb	VSD - Verb Sense Disambiguation
VGF - Finite Verb Chunk	WSD – Word Sense Disambiguation
VGNN - Verb Chunk (Gerund)	
VGNF - Non-finite Verb Chunk	

Transliteration Schema

1. Vowels

IPA	Devanagari
ʌ	अ
ɑ	आ
i	इ
u	उ
e	ए
o	ओ

2. Consonants

IPA	Devanagari	IPA	Devanagari
k	क्	d	द्
kh	ख्	dh	ध्
g	ग्	n	न्
gh	घ्	p	प्
ŋ	ङ्	ph	फ्
c	च्	b	ब्
ch	छ्	bh	भ्
ɖ	ज्	m	म्
dʒh	झ्	j	य्
t	ट्	r	र्
tʰ	ठ्	l	ल्
d	ड्	v	व्
dʰ	ढ्	s	स्
t	त्	h	ह्
th	थ्	ʈ	ड्

ABSTRACT

This research delves into the intricate landscape of natural language processing (NLP) for low-resourced languages, with a focus on Nepali. It highlights the pressing need for the development of linguistic resources to empower language technologies for such languages. Through the lens of computational linguistics, this study navigates the challenges and complexities of linguistic ambiguity, emphasizing the pivotal role of linguistic knowledge in achieving precision in various NLP applications. A pivotal contribution of this research is the construction of a comprehensive knowledge base comprising 482 verb frames for 200 ambiguous Nepali verbs. These verb frames are meticulously enriched with lexical, syntactic, and semantic information, offering valuable assets to enhance the efficacy of NLP tasks and applications.

Ambiguity is a pervasive obstacle in natural language which is recognized as a formidable barrier to achieving accuracy in NLP applications. This research confronts the complexity of ambiguity and explores strategies to effectively address it, acknowledging it as an 'AI-complete problem.' The adoption of a knowledge-based approach, centered on verb frames, is showcased as a promising technique for Verb Sense Disambiguation. Leveraging the knowledge resource such as verb frames, this approach enhances the accuracy of sense disambiguation. The evaluation of the Verb Sense Disambiguation system offers critical insights into its strengths and weaknesses. A spectrum of identified errors, ranging from preprocessing challenges to database limitations and complexities arising from multiple matches, serves as a compass for future refinements and enhancements.

In summary, this research represents a significant stride toward advancing the capabilities of natural language processing for Nepali and similar underrepresented languages. It underscores the pivotal role of linguistic resources, contextual understanding, and systematic approaches in addressing linguistic intricacies and achieving precision in NLP applications. The journey continues, with promising opportunities for further research and advancements in the realm of computational linguistics for marginalized languages.

TABLE OF CONTENTS

	Page No.
Declaration	i
Certificate	ii
Acknowledgements	iv
Abbreviations	vi
Transliteration Schema	ix
Abstract	x
Table of Contents	xi
List of Figures	xv
List of Tables	xvi
CHAPTER 1: INTRODUCTION	
1.1 Background	1
1.2 Aim and Scope of the Study	1
1.3 Review of Literature	2
1.3.1 Nepali Language Processing	2
1.3.2 Nepali Verbs	3
1.3.3 Verb Frames	5
1.3.4 Word Sense Disambiguation	6
1.3.5 Verb Sense Disambiguation	8
1.4 Significance of the study	9
1.5 Limitations of the Study	10
1.6 Organization of the Study	11
CHAPTER 2: METHODOLOGY	
2.1 Introduction	13
2.2 Collection of Data	13
2.2.1 The corpus	13
2.2.2 Extraction of Verbs	15
2.2.3 Selection of Verbs	15
2.3 Theoretical Framework	16
2.3.1 Verb Frame	16

2.3.2 Knowledge Based Approach for Verb Sense Disambiguation	22
CHAPTER 3: NEPALI VERBS	
3.1 Introduction	23
3.2 Verb Morphology	23
3.2.1 Finite verbs	25
3.2.2 Non Finite Verbs	26
3.2.3 Agreement System	26
3.2.4 Tense, Aspect and Mood	28
3.3 Verb Phrases	33
3.3.1 Compound Verb	33
3.3.2 Conjunct Verb	33
3.3.3 Transitive Construction	34
3.3.4 Passive Construction	35
3.3.5 Causative Construction	36
3.3.6 Negative Construction	37
3.4 Semantic Classification of Verbs	37
CHAPTER 4: ANALYSIS OF VERB FRAMES DATABASE	
4.1 Introduction	42
4.2 What is Verb Frame?	42
4.3 Importance of Verb Frame	43
4.3.1 A Knowledge Base	43
4.3.2 Linguistic Analysis of Verbs	44
4.3.3 Annotation of Verbs	44
4.3.4 A Resource for Verb Sense Disambiguation	44
4.4 Components of Verb Frames	44
4.4.1 Description	45
4.4.1.1 Verb Name	45
4.4.1.2 Sense ID	45
4.4.1.3 Verb Sense	46
4.4.1.4 English Gloss	46
4.4.1.5 Verb Class	46

4.4.1.6 Verb in Same Class	46
4.4.1.7 Example	48
4.4.1.8 Semantic Roles	48
4.4.1.9 Frame ID	50
4.4.2 Verb Frame	50
4.4.2.1 Arc Label	51
4.4.2.2 Necessity of the argument	51
4.4.2.3 <i>Vibhakti</i>	53
4.4.2.4 Lexical Category	54
4.4.2.5 Ontology	54
4.5 Statistics of Verb Frames	58
4.6 Classification of Nepali Verbs based on Verb Frames	58
CHAPTER 5: VERB SENSE DISAMBIGUATION SYSTEM	
5.1 Introduction	64
5.2 Nature and Scope of Ambiguity	65
5.3 Types of Ambiguity	66
5.3.1 Lexical Ambiguity	66
5.3.2 Syntactic Ambiguity	67
5.3.3 Semantic Ambiguity	69
5.3.4 Pragmatic Ambiguity	69
5.4 Word Sense Disambiguation and its Approaches	69
5.5 Verb Sense Disambiguation System	71
5.5.1 System Architecture	72
5.5.2 Input Sentence	73
5.5.3 Pre-processing	74
5.5.3.1 Parts of Speech Tagging	75
5.5.3.2 Chunking	77
5.5.3.3 Parsing	78
5.5.4 Conversion of Text into VSD system Format	78
5.5.5 Verb Frames Database	79
5.5.6 Verb Sense Disambiguation	80

5.5.6.1 Sense Identification	81
5.5.6.2 Match with Karaka Label	81
5.5.6.3 Match with Ontology	83
5.5.6.4 Match with Match with Verb Class	84
CHAPTER 6: EVALUATION AND ERROR ANALYSIS	
6.1 Introduction	85
6.2 Evaluation	85
6.3 Error Analysis	86
6.3.1 Pre-processing Error	87
6.3.2 Out of Verb Frame Database	88
6.3.3 Out of Verb Class Lexicon	90
6.3.4 Out of Ontological Lexicon	91
6.3.5 Error due to multiple matches in each level of information	92
CHAPTER 7: SUMMARY AND CONCLUSION	
7.1 Summary	93
7.2 Conclusion	98
REFERENCES	101
APPENDICES	
Appendix I: List of Ambiguous Nepali Verbs taken for building Verb Frames	106
Appendix II: Semantic Classification of Nepali Verbs based on Levin (1993)	108
Appendix III: Unique Verb Frames with their Sense IDs	132
Appendix IV: Ambiguous Nepali Verbs with respective number of senses	142
Appendix V: Sample Verb Frames	145
Appendix VI: Research Papers Published	163

List of Figures

Figure 2.1: Verb Frame Developed by Begum (2017)

Figure 2.2: Modified Verb Frame for Nepali

Figure 3.1: Components of Verb Phrase in Nepali

Figure 4.1: Verb Frame for the first sense of verb ‘au’

Figure 4.2: Verb Frame for the verb second sense of verb ‘au’

Figure 4.3: Verb Frame for the second sense of verb ‘dʒan’

Figure 4.4: Verb Frame for the third sense of the verb ‘ukas’

Figure 4.5: Tag Set for Ontology

Figure 4.6: Verb Frames for the verb ‘ukas’ having two different senses

Figure 5.1: System Architecture

Figure 5.2: Parser Output

Figure 5.3: Conversion of Text Format

Figure 5.4: Schema of Nepali Verb Frame

Figure 5.5 (a): System Output for First Sense of the Verb ‘ukel’

Figure 5.5 (b): System output for Second Sense of the Verb ‘ukel’

Figure 5.6: System output for two senses of the verb ‘markau’

Figure 5.7: System output for the two senses of the verb ‘ukhel’

Figure 6.1: Preprocessing Error

Figure 6.2: Error due to lack of Verb Frame Database

Figure 6.3: Error due to lack of Verb Class Information

Figure 6.4: Error due to lack of Ontological Information

Figure 6.5: Error due to multiple matches in each level of information

List of Tables

- Table 2.1: Corpus for Developing Verb Frames
- Table 2.2: Corpus for Testing
- Table 2.3: Number of Verbs with Respective Number of Senses
- Table 3.1: Nepali Verb Alternation based on Gender, Number, Person and Honorific Status
- Table 3.2: Tense and Aspect System in Nepali
- Table 3.3: Examples of Causative Formation in Nepali
- Table 3.4: Negativization Process in Nepali
- Table 4.1: *karakas* and dependency relations
- Table 4.2: Argument Structure and Ontology for different senses of *ukas*
- Table 4.3: Number of verbs with respective number of senses
- Table 5.1: BIS Tag Set for Indian Languages (2010)
- Table 5.2: Chunk Tag Set for Indian Languages
- Table 6.1: Test Result

CHAPTER 1

INTRODUCTION

1.1 Background

The ambiguity inherent in any natural language is the primary challenge when attempting to automate natural language processing through computers. Natural Language Processing (NLP), which aims to achieve near-perfect accuracy in interpreting human language, often encounters difficulties during computer programming. In this context, ambiguity has been a widely discussed problem in the field since its early stages of development.

In this study, we make an attempt to build a verb sense disambiguation system for Nepali using verb frames as a knowledge resource. Our goal is to initially create a database of Verb Frames for the Nepali language, which will be useful for different NLP applications along with sense disambiguation of Nepali verbs. The Verb Frames presented in this study offer comprehensive information about specific verbs and their unique behavior when combined with different nouns. Furthermore, this system incorporates ontological information to comprehend the underlying semantic features of a noun associated with a particular verb sense.

1.2 Aim and Scope of the Study

Machine learning without linguistic knowledge can be extremely challenging. To enable computers to understand natural language phenomena and produce accurate outputs, linguistic analysis of text becomes essential. Therefore, having a knowledge base that offers insights into strategies for handling ambiguity in any natural language proves invaluable.

Hence, the primary objective of this research is to establish a knowledge base for Nepali, which will aid in determining the correct interpretation of verbs within their given context. We have developed “a database comprising 482 Verb Frames for 200 ambiguous Nepali verbs. This database provides comprehensive information about each verb's lexical, syntactic, and semantic behaviours” (Manger, 2023, p.1).

The main scope of this study is primarily in developing and analyzing Verb Frames for ambiguous Nepali verbs to resolve Nepali verb ambiguity. However, this study also describes Nepali verbs' formal and functional attributes in general. On these bases, the research problems which this study aims to address are formulated as follows:

- i. To examine formal and functional attributes of Nepali verbs.
- ii. To build Verb Frames for ambiguous Nepali verbs.
- iii. To identify dependency relations of ambiguous verbs.
- iv. To provide ontological requirements of dependencies of ambiguous verbs.
- v. To classify verbs into their semantic types based on unique frames.
- vi. To provide linguistic cues for disambiguating ambiguous Nepali verbs based on verb frames

1.3 Review of Literature

The Nepali language lags behind in terms of NLP-focused research compared to many other languages. Only a few sporadic researches are available on Nepali NLP. However, many works are available on Nepali verbs from traditional and descriptive perspectives. A handful of works are also found on Nepali verbs from the computational perspective. To some extent, the literature on these aspects contributes to understanding the research problems formulated in the previous sections. The literature related to the research problems are thematically reviewed in five groups as follows:

1.3.1 Nepali Language Processing

NLP research from India and Nepal is being conducted on individual and institutional efforts for Nepali. They include creating and annotating various types of corpus and developing NLP tools and applications such as POS Tagger and MT system. The passages below summarise those efforts:

- Linguistic Data Consortium for Indian Languages (LDCIL), Central Institute of Indian Languages (CIIL), Mysuru, has developed POS Tagger for Nepali using a hybrid approach, which claims to have F-Score of 91.99% (Yoonus and Sinha, 2011). It also has created text corpora of 7, 057, 524 words for Nepali as of July 2014 (LDCIL, 2014) and collected comparable text corpora of 2632256 – 202157 words

for English-Nepali (LDCIL, 2014). The LDCIL, Mysuru, has also collected speech corpora of 88:55:04 hours as on July, 2014 (LDCIL, 2014).

- The North East Indo WordNet for North Eastern Indian languages including Nepali, has been developed by IIT, Bombay (Bhattacharya, 2010).
- The Indian Languages Corpora Initiative Project, conducted collaboratively under the leadership of Jawaharlal Nehru University, has produced annotated monolingual and parallel corpora comprising approximately two hundred thousand sentences in Nepali (Jha, 2010).
- NeLRaLEC project led by Open University, UK and funded by Asia IT & C Programme of the European Commission has developed “a Nepali National Corpus of 13 million words which contains written texts from 15 different genres with 2000 words each published between 1990 and 1992 and texts from a wide range of sources such as the internet webs, newspapers or books” (Bal, 2009).
- Kevin Hendricks has built Nepali Thesaurus using MyThes framework 6, which contains 5,500 entries with information such as POS Tag, meaning, and synonym (Bal, 2009).
- Andrew Hardie has developed POS Tagger for Nepali, also known as Unitag. It is a unified tagging system based on Unicode and a language-independent tagging system. It claims to have 93% accuracy (Bal, 2009).
- “English to Nepali Machine Translation System, also known as Dobhase, has been developed by a collaboration between Kathmandu University and Madan Puraskar Pustakalaya” (Bal, 2009). It is a rule and transfer-based Machine Translation System that can present gist translations to simple declarative sentences.

1.3.2 Nepali Verbs

- Adhikary (1980) describes '-cha' as a non-past marker in Nepali. He also formally classifies the Nepali tense as past and non-past against a threefold traditional classification: past, present, and future. He has defined '-ne' in Nepali as a prospective marker rather than a future tense marker. He argued that only definiteness cannot be the criterion for the classification of tenses.
- Pant (1990) analyzes and classifies Nepali verbs based on their semantic and

syntactic properties. Pant argues that the presence and absence of inherent properties like telic situation and postpositions like dative, locative, ergative case markers, transitivity, volitionally, etc., help determine the Nepali verb class.

- Acharya (1991) is a corpus-based descriptive grammar of Nepali. He has presented a detailed description of verb phrases and regular inflectional suffixes of Nepali finite verbs.
- Pokharel (1991) is a study of Nepali compound verbs, and it exemplifies Nepali compound verbs with a maximum of ten vectors in a string.
- Pokharel (1992) presents the four matrices for the classification of moods in Nepali: Imperative, Optative, Potential, and Indicative.
- Pokharel (1997) is the first to define Nepali verbs syntactically. He presents the classification of Nepali verbs based on post position with a syntactic approach.
- Lohani (1999) is an analysis of permissive construction in Nepali within the framework of Lexical-Functional Grammar. The main argument of his study is that Nepali permissive construction is a complex predicate in which the argument structure is complex, and composition is in the syntax.
- Lohani (2001) is about the valency of Nepali predicates. He describes the process of transitivization and negativization as the valency-increasing factors in Nepali.
- Pant (2001) explains the logical structure of Nepali verbs. He has tested 20 different verbs based on the absence or presence of Inherent Terminal Points and Process of Action.
- Sharma (2005) is a detailed study of Nepali nominal and verbal agreement patterns. This study shows that agreement between subject and verb regarding person, number, and honorifics is found in Nepali.
- Prasain (2008) is “a computational analysis of Nepali basic verbs (written form)”. He categorizes Nepali basic verbs based on their structure into underived, derived, and compound.
- Pokharel (2010) searches for “the general formula of verb root derivation in Nepali, and it finds a mathematical strategy of root derivation, which is commonly adopted to discover an underlying form in classical generative phonology, is general and more reliable to derive roots in Nepali”.

1.3.3 Verb Frames

- Levin (1993) introduced an extensive classification system for English verbs. This classification is rooted in the correlation between syntactic and semantic aspects of verb behavior. It was created as part of the Lexicon project within the Centre for Cognitive Science at MIT. Levin's framework posits that a verb's behavior, especially in terms of expressing and interpreting its arguments, is predominantly shaped by its meaning. In essence, the meaning of a verb plays a pivotal role in dictating its behavior.
- WordNet stands as a lexical database for the English language. This resource organizes various word types, including nouns, adjectives, verbs, and adverbs, into sets of cognitive synonyms referred to as "synsets." Each synset represents a unique concept and is interconnected through conceptual semantic and lexical resources. "Users can navigate these connections through a browsing interface. While synonymy serves as the primary relationship among words in WordNet, it also encodes sense relations such as hyperonymy, meronymy, troponymy, and antonymy" (Miller, 1995).
- FrameNet is a machine-readable lexical database for English. This project emerged from the International Computer Science Institute at Berkeley and focuses on semantic role labeling within frames. "FrameNet comprises over 13,000 word senses, each accompanied by annotated examples. Additionally, it contains a repository of more than 200,000 manually annotated sentences, linked to over 1,200 semantic frames" (Baker et al., 1998).
- PropBank is an abbreviation of "Proposition Bank." This resource features a corpus annotated with verbal propositions and their associated arguments. "PropBank provides a wealth of sentences annotated with semantic roles, each defined in relation to a specific sense of an individual verb" (Palmer et al., 2005). Each sense of a verb within PropBank is associated with a distinct set of roles, including Arg0, Arg1, Arg2, and so forth.
- VerbNet is an online lexicon of English verbs. This resource was developed as part of a research project and aims to offer a hierarchical, domain-independent, and comprehensive coverage of English verbs. VerbNet provides mappings to other

lexical resources like WordNet, Xtag, and FrameNet. “Its primary goal is to refine and augment Levin's verb classes to achieve greater syntactic and semantic coherence among class members” (Schuler, 2005). Each class within VerbNet provides two types of information: syntactic descriptions illustrating thematic roles and selectional restrictions of a verb, and syntactic frames that depict the verb's syntactic structure along with associated semantic predicates indicating temporal function.

- Ghosh (2015) has constructed verb frames for Bengali vector verbs derived from 15 Bengali compound verbs. These frames capture information regarding the number of arguments and case markings these verbs exhibit when used in both simple and compound verb constructions.
- Begum (2017) developed a database containing 486 Verb Frames for 300 verbs in Hindi. Begum's model of verb frames is grounded in the Paninian Grammatical Framework, which represents linguistic information about a verb, including its description and dependency relations, in a tabular format.

1.3.4 Word Sense Disambiguation

Its inception can be traced back to the emergence of Machine Translation (MT), a field confronted with the persistent challenge of ambiguity right from its early stages. The pivotal recognition of WSD as a central endeavor within MT can be attributed to the pioneering work of Weaver (1949) and Hillel (1960). Presently, three predominant WSD methodologies endeavor to resolve this ambiguity by harnessing knowledge bases or corpora.

- **Knowledge-based Approach** - Halliday and Hasson (1976), Lesk (1986), Gale, Church, and Yarowsky (1992), Pustejovsky (1991, 1995), Brin and Page (1998), Buitelaar (1998), Kilgarriff and Rosensweig (2000), Magnini and cavalier (2000), Veronis (2004), Khapraet *al* (2010) etc. are based on a knowledge-based approach, which relies on machine-readable language resources such as dictionaries, thesaurus, and WordNet.
- **Supervised Approach**- Leacock et al. (1993), Ng and Lee (1996), Mooney (1996), Voorhees (1998), Mihalcea and Moldovan (1999), Pedersen et al. (2000), Towell and

Montoyo *et al.* (2005), Dang *et al.* (2003), Carpuat and Wu (2007), Basile (2007), Zampieri (2010), Singh *et al.* (2014), Menai (2014), etc. are based on a supervised approach which seeks resolution using machine learning algorithms modeled upon trained corpus.

- **Unsupervised Approach** - Dagan and Itai (1994), Schütze (1998), Windows *et al.* (2003), Purandare and Pedersen (2004), Razon and Cheng (2011), Pal *et al.* (2013), etc. use the raw corpus for machine learning which deploy clustering algorithms.

Word Sense Disambiguation in Indian Languages

- Malarkodi and Balamurugan (2013) provide an extensive overview of Word Sense Disambiguation (WSD) approaches tailored specifically for Indian languages. Their comprehensive analysis encompassed supervised, unsupervised, and knowledge-based methodologies. Additionally, they conducted assessments of WSD systems within the context of the Senseval/Semeval campaigns
- Agarwal and Jain (2015) present a new approach for WSD in Hindi utilizing HAL vectors and Fuzzy-Means clustering.
- Bhingardive and Bhattacharyya (2016) present the usage of Indo WordNet in performing WSD. They have used linked WordNets and lexico-semantic relations for disambiguation tasks following the unsupervised approach of WSD.
- Sheth and Vyas (2018) discuss all the approaches for WSD and apply a Graph-Based unsupervised approach for Indian languages using Hindi WordNeT.
- Pal *et al.* (2021) provide a detailed analysis of various WSD approaches in Bengali and show the effectiveness of combining multiple methods to achieve significant improvement in accuracy compared to the baseline strategies.
- Maurya and Bahadur (2022) emphasize the importance of WSD in computational etymology and its relevance to MT and AI. It highlights the ongoing efforts to tackle ambiguity and presents a detailed examination of the different approaches and algorithms employed in WSD research for various Indian languages.

Word Sense Disambiguation in Nepali

- Shrestha et al. (2008) delved into the realm of ambiguity resolution, specifically examining the influence of Natural Language Processing (NLP) resources such as the Morphological Analyzer and Machine Readable Dictionary. Their research also introduced a novel adaptation of the Lesk Algorithm for Word Sense Disambiguation (WSD) in the Nepali language, utilizing the Nepali WordNet, a resource containing several noun sets. Their findings indicated that the Lesk Algorithm, when applied with Nepali WordNet, achieved an accuracy range of 50-70%, a notable contribution. However, it's worth noting that this accuracy fell short compared to the 80% accuracy achieved through manual morphological analysis, highlighting the ongoing challenges in automating ambiguity resolution in the Nepali language
- Dhungana *et al.* (2014, a) proposes a new knowledge-based approach for disambiguating Nepali words, i.e., using clue words defined as 'related context words for each word.' It claims 91% accuracy for disambiguating small-scale corpora, which contain 201 sentences with 201 polysemous words.
- Dhungana *et al.* (2014, b) is a similar study to the previous one. It is an organized projection of the same WSD approach using clue words for Nepali.
- Dhungana *et al.* (2014, c) propose a modified adapted Lesk Algorithm for WSD in Nepali. It includes synset, gloss, example, hypernym, and a context window containing all the words except articles, prepositions, and pronouns. It claims 88% accuracy while testing 201 sentences using a small-scale WordNet of 348 words.
- Roy *et al.* (2014) propose the knowledge-based approach, especially an overlapped, conceptual distance, and semantic graph-based approach to Nepali WSD using Indo WordNet. It presents test data with a corpus containing Nepali documents from the Indian Language Technology Proliferation and Development Centre under the Ministry of IT, Govt. of India. They take a total of 1663 with 912 nouns and 751 adjectives for disambiguation, and an accuracy of 54% for nouns and 42% for adjectives using an overlapped-based approach is claimed. Likewise, using conceptual distance and a semantic graph-based approach, an accuracy of 62% for nouns and 58% for adjectives is achieved.

- Singh et al (2021) undertook an evaluation of the streamlined Lesk algorithm for Word Sense Disambiguation (WSD) within the context of the Nepali language. This algorithm operates by quantifying the similarity between sense definitions and the contextual usage of a word with multiple meanings.

1.3.5 Verb Sense Disambiguation

- Ye (2004) investigates the performance of semantic role labeling systems for verb sense disambiguation and explores alternative ways of applying selectional preferences by deploying WordNet noun synonym sets and hypernymy sets.
- Buscaldi et al. (2006) propose the disambiguation of verbs utilizing Support Vector Machines and using WordNet-extracted features based on the hypernyms of context nouns.
- Ye and Baldwin (2006) delved into a study exploring the enhancement of supervised verb sense disambiguation performance through the utilization of multi-semantic-role (MSR) based selectional preferences.
- Dligach and Palmer (2008) introduced a novel method for extracting semantic information related to the arguments of verbs and subsequently applying this information to the task of verb sense disambiguation. Their approach aimed to enhance the accuracy of distinguishing different senses of verbs in context.
- Sudarikov et al. (2016) experiment the verb sense disambiguation using verbal patterns based on corpus pattern analysis and verbal word senses from valency frames.
- Gella and Elliott (2019) propose “the cross-lingual verb sense disambiguation using the Multi-Sense dataset of images annotated with English, German, and Spanish verbs”.
- Chen and Palmer (2009) emphasize “the effectiveness of the high-performance WSD system for English verbs using linguistically motivated features and a smoothed maximum entropy machine learning model”.
- Dutta and Borgohain (2021) propose the Latent Semantic Analysis for verb sense disambiguation by “applying a prediction algorithm to derive the most appropriate semantic neighbors for the target polysemous verb from the semantic space”.

- Dey and Maringanti (2022) examine the sense disambiguation of Odia verbs by annotating them with morphological features.

1.4 Significance of the Study

Verb sense disambiguation involves determining the correct sense or meaning of an ambiguous verb within a given context. We have used verb frames as an approach to address this challenge.

The verb sense disambiguation helps NLP system to understand the verb senses in better way. Since verbs play a central role in sentence structures, disambiguating their meanings accurately contributes to more accurate and meaningful language understanding.

Verb sense disambiguation holds significant importance across a spectrum of NLP applications, including but not limited to “machine translation, information retrieval, semantic role labelling, question answering, and parsing”. In machine translation system, ambiguity in verb senses can lead to incorrect translations and by utilizing verb sense disambiguation system, machine can choose the appropriate translations based on the context which results in more accurate translations.

Likewise, in semantic role labelling, it helps determine the correct semantic roles of the arguments which a verb can take.

The verb sense disambiguation is also useful for the task of information retrieval. It helps to retrieve more relevant information by ensuring that the retrieved results are in line with the intended meaning of the query.

Similarly, in question answering systems, accurately disambiguating verb senses can lead to more precise answers. It ensures the system understands the exact meaning of the question.

1.5 Limitations of the Study

The study has the following limitations:

- There are various grammatical categories which encounter the problems of ambiguity like noun, pronoun, adjective and adverb but this study deals with ambiguity in one grammatical category only i.e. verb.

- Among various approaches for verb sense disambiguation like supervised, semi-supervised, unsupervised etc., the knowledge-based approach is utilized for the sense disambiguation of the Nepali verbs in this study.
- With limited resource and time constraint, we could develop “small scale database of verb frames for Nepali” (Manger, 2023) which represents the model of verb sense disambiguation system using verb frames. However, by developing large scale database and incorporating deep learning techniques, one can enhance the performance of the system.

1.6 Organization of the Study

Chapter 1: Introduction

This chapter sets the foundational framework for the thesis, introducing key elements such as the study's Aim and Scope. It also includes a comprehensive Literature Review, providing an overview of existing research in the field. Additionally, the chapter explores the Significance and Limitations of the research.

Chapter 2: Methodology

It delves into the methodological aspects of the research, divided into two sections: Corpus Collection and Theoretical Background.

Chapter 3: Nepali Verbs

This chapter provides a detailed morphological description of Nepali verbs, encompassing two broad topics: Verb Morphology and Verb Phrases.

Chapter 4: Analysis of the Verb Frames Database

It is dedicated to the analysis of the verb frames database developed as part of this research.

Chapter 5: Verb Sense Disambiguation System

It elaborates on the stepwise architecture of the verb sense disambiguation system, which was constructed using the verb frames as a foundational knowledge resource for disambiguating Nepali verbs.

Chapter 6: Evaluation and Error Analysis

In this Chapter, the study evaluates the performance of the verb sense disambiguation system and discusses its results. This chapter also provides an in-depth analysis of any errors that occurred during the disambiguation process.

Chapter 7: Summary and Conclusion

This chapter summarizes and concludes the whole research.

CHAPTER 2

METHODOLOGY

2.1 Introduction

This chapter delves into the methodological aspects of the research, which are categorized into two main sections: Collection of Data and Theoretical Framework. In Section 2.2, the source and characteristics of the data employed in constructing verb frames for Nepali are discussed. Section 2.3 provides an in-depth exploration of the theoretical framework underpinning the creation of verb frames and outlines the approach for tackling the task of verb sense disambiguation through the utilization of these frames.

2.2 Collection of Data

In this study, “a total of 482 verb frames for each sense of 200 ambiguous verbs” have been developed. The verb senses are verified using the corpus of 200K sentences and four standard Nepali dictionaries available in print form like *Nepali kriyahaaruko kosh* (Poudel, 2015), *Nepali brihat shabdakosh* (Parajuli, 2010), *Ekta Concise Nepali-to-English Dictionary* (Lohani & Adhikari, 2010) and *Pragya Nepali-to-English Dictionary* (Upreti et al., 2013). The following subsections describe the source and nature of data collected for this study:

2.2.1 The Corpus

The corpus for two different purposes was collected in this study. The first is for developing verb frames, and the another is for testing the data for the verb sense disambiguation task. Firstly, the corpus of 200K sentences was collected using various written sources from personal effort as well as from various NLP projects like the Indian Languages Corpora Initiative (ILCI) project, websites of Technology Development for Indian Languages (TDIL), Government of India and Samakalin Sahitya, an online Nepali literary journal and the <https://github.com/sharad461/nepali-translator>. The written corpus of different domains, like entertainment, health, sports, news, literature, tourism, agriculture, etc., was collected from the above sources. Table 1.1 presents detailed information about the collection of the corpus.

Table 2.1 Corpus for Developing Verb Frames

Source	Sentences	Domain	Tokens	Types
ILCI	100K	Tourism, Health, Entertainment, Literature, Science and Technology	357151	50310
TDIL	1K	Tourism	35716	5031
www.samakalinsahitya.com	20K	Literature	71430	10062
https://github.com/sharad461/nepali-translator	70K	General	250005	35217
	Total = 200K		Total = 714302	Total = 100620

Secondly, the corpus of 1K sentences was collected for testing the data. As mentioned above, the corpus sources were ILCI, TDIL, Samakalin Sahitya, and the GitHub account. However, the corpus used for testing differed from that used for developing verb frames. Table 2.2 demonstrates the source and nature of the corpus used for testing.

Table 2.2 Corpus for Testing

Source	Sentences	Domain	Tokens	Types
ILCI	500	Tourism, Health, Entertainment, Literature, Science and Technology	6812	2431
TDIL	100	Tourism	1362	486
www.samakalinsahitya.com	100	Literature	1362	486
https://github.com/sharad461/nepali-translator	300	General	4088	1461
	Total = 1000		Total = 13624	Total = 4867

2.2.2 Extraction of Verbs

The data for developing verb frames were extracted from the corpus of 200K sentences of different domains of language use. The first step of extracting the data was to sort out all verb occurrences from the corpus. Secondly, unique verb forms looking at their roots were separated. Thirdly, looking at the frequency of occurrences, 200 verbs were chosen for the study.

2.2.3 Selection of Verbs

Ambiguity is the primary criterion for selecting data for this study, i.e., Nepali verbs with more than one sense have been taken for developing verb frames. Two hundred verbs with two to fifteen different senses are chosen for verb frames. Frequently used verbs with more obvious ambiguous senses are preferred for making verb frames. Table 2.3 demonstrates the number of verbs with a respective number of senses.

Table 2.3 Number of verbs with the respective number of senses

Number of Verbs	Number of Senses
166	2
16	3
9	4
2	5
4	6
1	6
1	11
1	15

Since there are various kinds of verbs based on their morphology, viz. simple, complex, and compound, etc., only simple verbs with one root are incorporated in this study. This kind of selection is because complex and compound verbs are made up of two or more roots or stems, which show complexity in their morphological behaviour that would result in the different meanings of a particular verb in different contexts.

2.3 Theoretical Framework

In this section, we delve into the theoretical underpinnings, comprising two distinct sub-sections. The initial sub-section delves into the exposition of the theoretical model employed in the creation of a verb frames database. Conversely, the subsequent sub-section offers an in-depth exploration of the methodology used in crafting a verb sense disambiguation system, leveraging the power of verb frames.

2.3.1 Verb Frame

In this study, a Verb Frame serves as a knowledge repository, furnishing lexico-syntactic and semantic insights concerning a specific verb. These Verb Frames encapsulate a linguistic analysis of verbs presented in a structured tabular format. Our development of these Verb Frames is geared toward the objective of verb sense disambiguation, drawing inspiration from various NLP resources as our theoretical foundation. The notion of constructing verb frames in this tabular format is indebted to the pioneering work of Begum (2017), who crafted verb frames for the Hindi language, encompassing both syntactic and semantic information pertaining to verbs. Her verb frames provide two pieces of information about a particular verb: Description of a verb and Verb Frame. The first component contains the following information:

- i. Verb name;
- ii. Sense-id;
- iii. Verb Sense;
- iv. English gloss;
- v. The verb in the same class;
- vi. Verb Frame Populated;
- vii. Example sentence;
- viii. Theta roles; and
- ix. Frame ID.

On the other hand, the second component of her verb frame has the following information:

- i. Dependency relations of the mandatory and desirable arguments;
- ii. Number of vibhakti or case markers;

- iii. Part of Speech information;
- iv. Position of a verb to its arguments; and,
- v. Relation of arguments to a verb.

The model of the verb frame developed by Begum (2017) is represented in Figure 2.1.

Figure 2.1 Verb Frame developed by Begum (2017)

Verb:: aa					
SID:: aa%VI%S1					
Verb_Sense:: HWN (1)					
Eng_Gloss:: come					
Verbs_in_Same_Class:: Synonyms → pahuMcha, padhaara%VI%S1%FID1;					
Verb frames populated → pahuMcha					
Frames::					
Frame_Name_1::					
Example:: raam aksara haidaraabaad aataa hai					
ram often Hyderabad come-PRES be-PRES					
"Ram often comes to Hyderabad"					
Theta_Roles:: AGENT DESTINATION V					
Demand_Frame:: FRAME_ID_1					
FRAME_ID:: aa%VI%S1%FID1					
arc-label	necessity	vibhakti	lextype	posn	reln
k1	m	0	n	1	c
k2p	m	0 ke paasa para taka	n	1	c

However, our verb frame has some modifications in terms of components of the verb frame, viewing their necessity for the sense disambiguation task. We added two more pieces of information about a verb in our verb-frame. They are – ‘verb class’ and ‘ontology’. The verb class is depicted following the semantic classification of English verbs by Levin (1993), while the ontological features of arguments are given following the tag set developed by CALTS (2017). These two pieces of information are crucial in determining the proper sense of a verb in the given context.

Another modification in our verb frame is removing information like ‘position’ and ‘relation’ incorporated in Begum’s verb frame. The reasons behind this exclusion are: i) the prevailing SOV (Subject-Object-Verb) dominant phrase order in Indian languages

and the consistent position of a verb relative to its arguments in nearly all sentences and ii) our approach aligns with Paninian grammar, which views a verb as the parent to its arguments, establishing a parent-children relationship among them. Given the uniformity of argument relations across sentences, we have opted to exclude this information from our verb frames. Figure 2.2 represents a sample of our verb frame.

Figure 2.2 Modified Verb Frame for Nepali

```
*****
Verb::ukhel
SID::ukhel%VT%S1
Verb_Sense::nikal
Eng_Gloss::Uproot
Verb_class::Verb of removing
Verbs_in_Same_Class::Synonyms>karh%VT%S4%FID4
Example::us-le    ghās    ukhel-jo
           he-ERG  grass.ACC uproot-3.SG.M.PST
           'He uprooted the grass.'
Theta_Roles::AGENT MATERIAL VERB
Frame_ID::ukhel%VT%S%FID1
-----
arc-label  necessity  vibhakti  lextype    ontology
-----
k1          m          le/0      p/n        [+ani/+hum]
k2          m          0        n          [+con]
-----
*****
```

The components of our verb frame can be divided broadly into i) Description of the verb and ii) Verb Frame. Information like verb name, sense ID, verb sense, English gloss, verb class, verbs in the same class, example, theta role, and frame ID, is included in the description part. In contrast, the verb frame consists of information like arc-label, the necessity of the arguments, *vibhakti*, lexical type, and ontology.

Figure 2.2 above represents each component of the verb frame for the first sense of the verb 'ukhel.' It has 'ukhel' as the *verb name*, which is the root form of the verb in Nepali. 'ukhel%VT%S1' is *sense ID*, comprised of the information separated by a percentage symbol: verb name, verb type (transitive or intransitive or causative verb) and the sense number based on the number of senses a verb may have. Likewise, the *verb sense* of 'ukhel' is given as 'nikal,' a synonymous sense of 'ukhel' in Nepali. *English gloss* of the verb 'ukhel' in the given frame is 'uproot,' and the *verb in the same class* for the

given verb is ‘kɑɾh%VT%S4%FID4’, which is a frame ID for the synonymous verb of 'ukhel.'

Likewise, within our verb frames, we incorporate the concept of verb class, which signifies the semantic category of the verb. This classification aligns with Levin's categorization of English verbs, as outlined in her work from 1993. We have integrated 38 out of the 49 verb classes delineated by Levin (1993) into our frames. Here is the list of the 38 verb classes we have employed:

- 1) Verbs of Putting
- 2) Verbs of Removing
- 3) Verbs of Sending and Carrying
- 4) Verbs of Exerting Force Push/Pull Verbs
- 5) Verbs of Change of Possession
- 6) Learn Verbs
- 7) Hold and Keep Verbs
- 8) Verbs of Concealment
- 9) Verbs of Throwing
- 10) Verbs of Contact by Impact
- 11) Verbs of Cutting
- 12) Verbs of Combining and Attaching
- 13) Verbs of Separating and Disassembling
- 14) Verbs of Coloring
- 15) Verbs of Creation and Transformation
- 16) Verbs of Perception
- 17) Psych-Verbs Verbs of Psychological State
- 18) Verbs of Desire
- 19) Verbs of Searching
- 20) Verbs of Social Interaction
- 21) Verbs of Communication
- 22) Verbs of Sounds Made by Animals
- 23) Verbs of Ingesting
- 24) Verbs Involving the Body

- 25) Verbs of Killing
- 26) Verbs of Emission
- 27) Destroy Verbs
- 28) Verbs of Change of State
- 29) Lodge Verbs
- 30) Verbs of Existence
- 31) Verbs of Appearance, Disappearance, and Occurrence
- 32) Verbs of Body-Internal Motion
- 33) Verbs of Assuming a Position
- 34) Verbs of Motion
- 35) Verbs of Lingering and Rushing
- 36) Measure Verbs
- 37) Aspectual Verbs
- 38) Weather Verbs

Another vital component integrated into our verb frames is the inclusion of *examples*. These examples consist of sentences in the Nepali language, illustrating the specific sense of the verb. To ensure accurate pronunciation, these examples are transcribed in the International Phonetic Alphabet (IPA), facilitating proper enunciation of the sentence.

Theta roles are semantic roles played by arguments of the verb, which are represented with the help of semantic role labelling made by VerbNet (Schuler, 2005). Following theta roles are captured in our verb frames:

- a) Actor
- b) Agent
- c) Asset
- d) Beneficiary
- e) Destination
- f) Source
- g) Location
- h) Experiencer
- i) Instrument

- j) Patient
- k) Recipient
- l) Stimulus
- m) Theme
- n) Time
- o) Topic

The *frame ID* is a unique identification number given to each verb frame for each sense of a particular verb. It consists of a sense ID and a frame ID number. The frame ID for the verb 'ukhel' in Figure 2.2 is given as 'ukhel%VT%S%FID1'.

The second part of the verb-frame is *verb frame*, a specific term for the tabular representation of the argument structure of a verb that contains information like arc-label, necessity, *vibhakti*, *lextype*, and ontology. The *arc label* is the label for the dependency relation of each verb argument. The dependency relation in our verb frames is represented following the Paninian Grammatical Framework, expressed through the *karaka* roles of arguments in a sentence. The annotation guideline prepared by Bharati et al. (2012) has been devised for annotating *the karaka* relations in our verb frame.

Necessity is the information regarding requirement types of arguments in a sentence. It can be 'mandatory' or 'desirable' depending upon the requirement of the proposition made by a predicate in a sentence. Figure 2.2 represents two mandatory arguments in the verb frame, represented by the abbreviation 'm'.

vibhakti is information about the case marker or any postpositional element that arguments of a verb may take. Figure 2.2 shows that the first argument of the verb 'ukhel' sometimes takes an ergative case marker (*karta*) 'le,' and sometimes a zero '0' marker depending upon different TAM. However, the second argument of the verb 'ukhel' in Figure 2.2 has zero '0' *vibhakti*.

Lextype gives parts of speech information about the arguments in the example sentence. The first and second arguments of the verb 'ukhel' in Figure 2.2 are pronoun ('p') or noun ('n') and noun, respectively.

Finally, the ontological features of each argument are represented in the component of the verb frame called *Ontology*. As we see in Figure 2.2, the ontology for the first and second arguments are animate (+ani) or human (+hum) and concrete (+con), respectively.

The tag set prepared by CALTS (2017) has been devised for annotating ontological features of arguments in our verb frame.

A detailed description of each component of the verb frame is provided in chapter 4 of this thesis.

3.3.2 Knowledge-Based Approach for Verb Sense Disambiguation

Verb Sense Disambiguation (VSD) stands as a pivotal sub-task within the domain of natural language processing. Its primary objective is to discern the most appropriate sense or meaning of a verb within a provided context, a fundamental endeavour in language understanding and computational linguistics.

Several approaches are employed for VSD, broadly categorized into Supervised Machine Learning, Unsupervised, Semi-Supervised, and Knowledge-Based approaches.

The knowledge-based approach to VSD leverages external linguistic resources such as dictionaries, thesauri, and WordNet to provide valuable insights into verb senses and their interrelationships. In our research, we have adopted a knowledge-based approach for VSD, as our system utilizes verb frames as a linguistic knowledge resource for disambiguating verb senses. As discussed in the previous section, our verb frames encompass comprehensive information about a verb, ranging from its sense to its argument structure. Details like karaka relations, verb classes, and ontologies have proven to be beneficial for the disambiguation task. Chapter 5 of this thesis presents an in-depth analysis of our VSD system, shedding light on its intricacies and effectiveness.

CHAPTER 3

NEPALI VERBS

3.1 Introduction

This chapter presents a general overview of the morphological and semantic attributes of Nepali verbs. The chapter is divided broadly into three sections: Section 3.2 presents the description of Nepali Verb Morphology which incorporates the description of Finite and Non Finite verbs, Agreement System and Tense, Aspect, and Mood in Nepali. Section 3.3 is about the verb phrases with due focus on compound and conjunct verbs as well as transitive, passive, causative and negative constructions in Nepali. Section 3.4 presents the semantic classification of Nepali verbs.

3.2 Verb Morphology

Nepali is one of the widely spoken languages in South-East Asia “that belongs to the northern group of Modern Indo-Aryan languages” (Nepal, 2009: 97). It has “proximity in grammar as well as in lexicon to other cognate languages like Kumauni and Garwali. The other names for this language are ‘Khas Kura,’ ‘ParbateBhasa,’ ‘Gorkhali’, and ‘Dzongkha Lhotshammikha.’ It is spoken widely in Nepal, India, Bhutan and Myanmar” (Manger, 2020, p.215).

In terms of linguistic typology, Nepali can be categorized as an agglutinating language. It exhibits a Subject-Object-Verb (SOV) phrase order as its dominant syntactic structure. Furthermore, Nepali features a robust nominal agreement system, which involves inflections for gender, number, person, and case.

In addition to its nominal agreement system, Nepali also boasts a complex verbal system. This verbal system encompasses various agreement patterns, including gender, number, person, tense, aspect, mood, and honorific status. These linguistic features collectively contribute to the rich and intricate nature of the Nepali language

Most of the verb roots in Nepali are monosyllabic. For example, [kha] 'eat,' [dʒa] 'go,' [sut] 'sleep,' [hĩɽ] 'walk' etc. The language tends to use the [nu] suffix with verb root

to create its infinitive form, and the same is considered as a lexical unit in the Nepali Lexicon viz. [kha-nu] 'to eat,' [dʒa-nu] 'to go, [sut-nu] 'to sleep, [hĩt-nu] 'to walk' etc.

Morphologically, Nepali verbs are inflectional as well as derivative in nature. The formation of the Nepali verbs is sometimes due to the derivational suffixes to nominal stems. For example, [dʌr-au-nu] 'to be frightened,' [bhok-au-nu] 'to be hungry' are the derivative verb forms made out of noun stems [dʌr] 'fear' and [bhok] 'hunger' respectively. Likewise, [rat-i-nu] 'to be reddish' and [moʈ-au-nu] 'to be fatty' are derivative verb forms made out of adjectival stems [rato] 'red' and [moʈo] 'fat'. Similarly, the adverbial stems like [bahirʌ] 'outside' and [ʈaʈa] 'far' have their derivative verbal form [bahir-i-nu] 'to go outside' and [ʈar-i-nu] 'to be far' respectively.

Morpho-Syntactically, the verb in Nepali inflects for tense, aspect, mood, and voice as well as negative and causative constructions. It agrees with the subject for gender, number, person, case and honorifics.

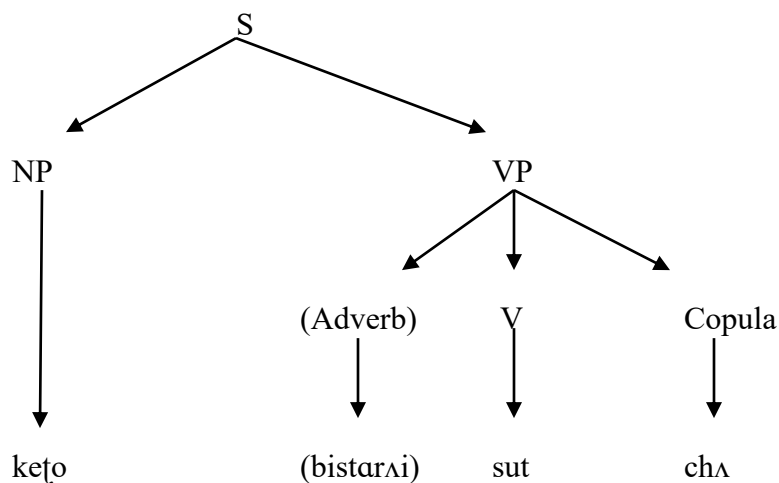
Syntactically, the distribution of verbs in Nepali shows the subject (S)-(Object (O))–Verb (V) pattern as in the sentences below:

- | | | | |
|-----|--------------------------|------------------|-----------------|
| i) | keʈo | hās-jo | [SV] |
| | boy.NOM | smile-3.SG.M.PST | |
| | 'The boy smiled' | | |
| ii) | birala-le | musa | mar-jo [SOV] |
| | cat-ERG | rat.ACC | kill-3.SG.M.PST |
| | 'The cat killed the rat' | | |

The components of verb phrase in Nepali can either be {Root+Copula} or {Adverb+Root+Copula} as seen in the instances below:

- | | | | |
|-----|-----------------------------|-------------------|------------------------------|
| i) | keʈo | sut-chʌ | {Verb+Copula} |
| | boy.NOM | sleep-3.SG.M.NPST | |
| | 'The boy will sleep' | | |
| ii) | keʈo | bistarʌi | sut-chʌ {Adverb+Root+Copula} |
| | boy.NOM | slowly | sleep-3.SG.M.NPST |
| | 'The boy will sleep slowly' | | |

Figure 3.1: Components of Verb Phrase in Nepali



As seen in the above instances, Verb Phrase in Nepali constitutes the adverb optionally and at least one root mandatorily. No other grammatical categories can act as a member of a Verb Phrase in Nepali.

3.2.1 Finite Verb

The verbs in Nepali can be either finite or infinite based on finiteness. The formal structure of the Finite verb in Nepali is 'root+copula.' In another way, the Finite verb in Nepali ends in [chu/chau/chas/chau/cha/chan] or [ho/thiyo] copular suffixes. Instances for Finite Verbs in Nepali are as follows:

Finite Verb

- | | | | | | |
|-----|-------------------------------|-------|-------|---------------|----------------|
| i. | ma | adʒa | tjo | kitab | paʃ-chu |
| | I.NOM | today | that | book | read-1.SG.NPST |
| | 'I will read that book today' | | | | |
| ii. | mai-le | tyo | kitab | paʃeko | ho |
| | I-ERG | that | book | read | be-1.SG.PST |
| | 'I have read that book' | | | | |

The main verb in any natural language expresses the action. In another way, the main verb embraces the major information about the nature of an action. Whereas an auxiliary verb, popularly known as a 'helping verb,' does not carry major information about the action but provides knowledge about the time of that action. In linguistic terms, it carries information about tense, aspect, mood, modality, etc. In Nepali, the task of

distinguishing an auxiliary verb from the main verb is quite cumbersome since the orthographic system of Nepali tends to write the main verb and auxiliary together as bound morphemes. For example- Nepali verb forms like **sak-cha** 'can do,' **bhan-cha** 'will say,' **sak-jo** 'did,' and **bhan-jo** 'said' have two elements in them. They are –main verbs **sak**, **bhan** and auxiliaries **cha**, **jo**. However, the grammatical researches to date consider them as main verbs as a whole because they are written together as one lexical unit. Hence, there is confusion about the distinction of the main and auxiliary verbs in Nepali. Though based on the definition of a main and an auxiliary verb found in linguistic literature, viz. Crystal (1980), we can argue that **cha**, **jo**, and **ho** are the only auxiliaries in Nepali which exhibit various morphological forms according to tense, aspect, mood, and other inflectional categories.

3.2.2 Non-Finite Verb

The Non-finite verbs in Nepali have either of [dʌi/dʌi/erʌ] endings. Hence, the formal structure of Non-Finite Verbs in Nepali is 'root+dʌi/dʌi/erʌ' suffixes. The instances below exemplify it more clearly.

- i. mʌ tjo kitab **pʌʃ-dʌi** khana pʌkaũ~chu
I.NOM that book read-PROG food cook-1.SG.NPST
'I will cook food reading that book'
- ii. mai-le kitab **pʌʃ-da** u gʌe-cha
I-ERG book read-PROG he go-1.SG.PST
'He went while I was reading the book'

3.2.3 Agreement System

The verb is the most inflected grammatical category in any natural language. Gender, Number, Person, and Honorificity are the primary inflectional units that trigger a verb in Nepali to exhibit various kinds of morphological forms. Like other Modern Indo-Aryan Languages, the Nepali language has a rich linguistic system of distinguishing verbal forms according to various factors like gender, number, person and honorifics. Among these, gender and honorifics reveal an idea about the social structure of Nepali society as well.

Grammatically, the Nepali verb inflects for two kinds of gender. They are Feminine and Non-Feminine. The linguistic entity that denotes other than a human does not show grammatical gender distinction in Nepali. The human masculine entities, as well as non-human entities, including non-living things, fall under the Non-Feminine gender in Nepali.

- | | | | |
|------|-------------------------|-------------|------------------|
| i. | ramro | keṭo | α-jo |
| | good | boy.NOM | come-3.SG.NF.PST |
| | ‘The good boy came’ | | |
| ii. | ramro | gai | α-jo |
| | good | cow.NOM | come-3.SG.NF.PST |
| | ‘The good cow came’ | | |
| iii. | ramro | gaṛi | α-jo |
| | good | vehicle.NOM | come-3.SG.NF.PST |
| | ‘The good vehicle came’ | | |
| iv. | ramri | keṭi | α-i |
| | good | girl.NOM | come-3.SG.F.PST |
| | ‘The good girl came’ | | |

In the sentences (i, ii, and iii) above, the verb [**α-**] ‘come’ have the same endings, i.e., they are the examples of how Non-Feminine gender is marked in Nepali. Whereas in the sentence (iv), the verb [**α-**] has [**-i**] ending and hence is the example of a Feminine marker. It is interesting to see that the Nepali verb not only shows an agreement pattern for nouns but adjectives as well, like in the sentences above, viz. [**ramr-o**] ‘good’ is for Non-Feminine and [**ramr-i**] is for feminine gender. In Nepali, the verbal forms for the Feminine gender have [**-i**], [**-in**], [**-che**], and [**-chin**] endings.

First, Second and Third persons with no inclusive and exclusive distinctions in Second person pronominal are found in Nepali. However, the person is an important grammatical unit with due reference to the Subject-Verb agreement system in Nepali. In other words, the verb forms in Nepali are vastly affected by their referent pronominal. Table 3.1 demonstrates an overall idea of Nepali Verb Alternation based on Gender, Number, Person and Honorificity as follows:

Table 3.1 Nepali Verb Alternation based on Gender, Number, Person and Honorific Status

Honorific Status →		Low		Mid		High		Higher		Royal	
Per- Son	Num ber	Non-Feminine	Feminine	Non-Feminine	Feminine	Non-Feminine	Feminine	Non-Feminine	Feminine	Non-Feminine	Feminine
1	SG	ma aũ-chu 'I will come'	ma aũ-chu 'I will come'	---	---	---	---	---	---	---	---
	PL	hami aũ- chau 'We will come'	hami aũ-chau 'We will come'	---	---	---	---	---	---	---	---
2	SG	tã aũ-chas 'You will come'	tã aũ-ches 'You will come'	timi aũ-chau 'You will come'	timi aũ-chjav 'You will come'	tapã au-nu-huncha 'You will come'	tapã au-nu-huncha 'You will come'	jahã ai-sinu-huncha 'You will come'	jahã ai-sinu-huncha 'You will come'	hadjur ai-baksin-cha 'You will come'	hadjur ai-baksin-cha 'You will come'
	PL	tã-haru aũ- chas 'You will come'	tã-haru aũ- chas 'You will come'	timi-haru aũ-chau 'You will come'	timi-haru aũ-chau 'You will come'	tapã-haru au-nu-huncha 'You will come'	tapã-haru au-nu-huncha 'You will come'	jahã-haru ai-sinu-huncha 'You will come'	jahã-haru ai-sinu-huncha 'You will come'	hadjur-haru ai-baksin-cha 'You will come'	hadjur-haru ai-baksin-cha 'You will come'
3	SG	u/tjo aũ-cha 'He will come'	u/tjo aũ-che 'She will come'	umi/tini aũ-chan 'He will come'	umi/tini aũ-chin 'She will come'	uhã au-nu-huncha 'He will come'	uhã au-nu-huncha 'She will come'	uhã ai-sinu-huncha 'He will come'	uhã ai-sinu-huncha 'She will come'	hadjur ai-baksin-cha 'He will come'	hadjur ai-baksin-cha 'She will come'
	PL	umi/tini-haru aũ-chan 'They will come'	umi/tini-haru aũ-chan 'They will come'	umi/tini-haru aũ-chan 'They will come'	umi/tini-haru aũ-chin 'They will come'	uhã-haru au-nu-huncha 'They will come'	uhã-haru au-nu-huncha 'They will come'	uhã-haru ai-sinu-huncha 'They will come'	uhã-haru ai-sinu-huncha 'They will come'	hadjur-haru ai-baksin-cha 'They will come'	hadjur-haru ai-baksin-cha 'They will come'

Table 3.1 illustrates that in Nepali; verb forms are subject to modification based on Gender, Number, Person, and honorifics, although this modification is not universal across all grammatical units. The most noticeable variations occur predominantly in the cases of second-person singular and third-person singular forms. In contrast, the verb remains invariant and does not exhibit any alterations concerning gender or honorific status in cases involving the first-person singular and plural forms, as well as the second and third-person plural forms. However, it's important to note that variations are indeed present for both Number and Person in the Nepali verb, showcasing the language's flexibility and complexity in verb conjugation.

3.2.4 Tense, Aspect and Mood

There is a twofold distinction for tense in Nepali: Past and Non-Past in terms of morphology. Since there is no distinct marker for future tense, Nepali tends to include it

in the Non-Past tense along with the Present one. The examples below demonstrate the tense system in Nepali as follows:

- i. mΛ hiɖʒo bΛɖʒar gΛ-jē
 I.NOM yesterday market.LOC go-PST
 'I went to the market yesterday'
- ii. mΛ aɖʒΛ ghΛr ɖʒa-n-**chu**
 I.NOM today home.LOC go-NPST
 'I go/will go home today'

There are five aspects which trigger a Nepali verb to show different morphological forms in Nepali. The aspects in Nepali are Simple, Continuous, Perfective, Habitual and Unknown. Among these, Habitual and Unknown aspects are for Past tense only. Table 3.2 illustrates the tense and aspect system.

Table 3.2 Tense and Aspect System in Nepali

Tense	Aspect	Example
Non Past	Simple	mΛ khana kha-n- chu I.NOM food.ACC eat-1SG.NPST 'I eat food'
	Continuous	mΛ khana kha-i-rΛh-e- chu I.NOM food.ACC eat-CONT-1SG.NPST 'I am eating food'
	Perfective	mΛ-i-le khana kha-i-sΛk- ē I-ERG food.ACC eat-PERF-PST 'I ate food already'
Past	Simple	mΛ-i-le khana kha-ē I-ERG food.ACC eat-1.SG.SIM.PST 'I ate food'
	Continuous	ma khana kha-i-rΛh-e- thē I-NOM food eat-CONT-1.SG.PST 'I was eating food'
	Perfective	mΛ-i-le khana kha-i-sΛk-e- thē I-ERG food.ACC eat-PERF-1.SG.PST 'I ate food already'
	Habitual	mΛ khana kha-i-rΛhΛ-n- thē I.NOM food.ACC eat-HAB-1.SG.PST 'I used to eat food'
	Unknown	mΛ-i-le khana kha-n- the-chu I-ERG food.ACC eat-UNK-1.SG.PST 'I don't know, but I used to eat food'

Table 3.2 shows that the Nepali verb most often alternates its form according to different tenses and aspects. The following points were analyzed to observe the same:

- a. Among five types of aspects, the Non Past tense does mark for three aspects. They are- Simple, Continuous, and Perfective. Habitual and Unknown aspects are seen in the Past tense only.
- b. **-chu** and **-ẽ** are simple aspect markers for Non-Past and Past Tense, respectively, and the verb form for simple aspect in both cases is of one root. Besides this, morphological forms made up of two roots express all the other aspects in Nepali that can be considered as a compound verb. For instance, the morpheme **-raɦ+e**, along with the main verb, expresses the aspectual meaning of continuity, and **-saɦ+e**, along with the main verb, expresses the meaning of perfective aspect. Whereas the morpheme **-raɦɦ+ɦ** and **-the+chu**, along with the main verb, carries the meaning of habitual and unknown aspects, respectively.

There are five types of grammatical moods in Nepali. They are Assertive, Imperative, Interrogative, Indicative and Presumptive.

The following examples demonstrate the cases of verbal inflections in terms of mood in Nepali:

i. **Assertive Mood**

The verbal form that simply asserts one's idea or makes a statement about something is considered an Assertive mood. In Nepali, this kind of mood is expressed through the verb with simple tense, such as:

- a. u me-ro bhai ho
He.NOM I-GEN brother-ACC be-3.SG.NPST
'He is my brother'
- b. hidzo mɦ ghɦr gɦ-eko thi-ẽ
yesterday I.NOM home-LOC go-NF be-1.SG.NF.PST
'I went to my home yesterday'

ii. **Imperative Mood**

It represents the meaning of command, request, or desire made by the speaker of a particular expression as in the following sentences:

- a. tɦ ghɦr dʒɦ
you.NOM home-LOC go-2.SG.NPST.LH

'You go home' (command)

- b. timi ghar ga-erA **sut-A**
you.NOM home-LOC go-NF sleep-3.SG.NPST.MH

'You go to home and sleep' (command/request)

- c. tapaĩ jahũ na-bas-nu-**hos**
you.NOM here NEG-sit-INF-be-3.SG.NPST.HH

'You please don't sit here' (request)

- d. ma jas-pali pas **ho-ũ**
I-NOM this-time pass be-1.SG.NPST

'I wish I pass this time' (desire)

In the examples above, the verb forms for imperative mood are in bold letters. It indicates that there are different forms for the same mood according to different grammatical properties because the verb in Nepali always agrees with the subject in a sentence.

iii. Interrogative Mood

It carries the meaning of asking questions, and it is interesting to see that most often, this kind of mood is expressed by increasing the intonation of an assertive sentence in Nepali, such as:

- a. timi baḍgar ḍan-chau (Assertive mood)
you.NOM market-LOC go-2.SG.NPST
'You will go to the market'
- b. timi baḍgar ḍan-chau? (Interrogative mood)
you.NOM market-LOC go-2.SG.NPST
'Will you go to the market?'

The sentence with the same sequence of words in a low intonation scale like sentence (a) is an assertive sentence, and the same sentence with the same sequence of words but on high intonation scale like sentence (b) is interrogative in Nepali.

Another method for making an interrogative sentence in Nepali is by positioning interrogative pronouns in a sentence like-

- a. **ke** timi bholi aũ-chau?
What you.NOM tomorrow come-be-2.SG.NPST

‘Will you come tomorrow?’

- b. sita **kinΛ** rojeki?
Sita.NOM why cry-3.SG.NPST
‘Why is Sita crying?’

iv. **Indicative Mood**

It conveys the meaning of indicating something that is going to happen in the near future. Examples in Nepali are as follows:

- a. ram dʒΛn-chΛ bhΛne **dʒa-vos**
Ram go-3.SG.NPST if go-3.SG.NPST
‘Ram can go if he wants’
- b. pani pΛr-jo bhΛne mΛ ghΛr **phΛrkin-chu**
rain fall-PST if I-NOM home return-1.SG.NPST
‘If rain falls, I will return home’

v. **Presumptive Mood**

It expresses the meaning of possibility. Examples in Nepali are as follows:

- a. bholi pani pΛr-chΛ **ho-la**
Tomorrow rain fall-NPST be-may
‘It might rain tomorrow’
- b. meri bΛhini ghΛr gΛi **ho-li**
My sister's home go-3.SG.PST be-may
‘My sister might go home’
- c. abΛ tΛ manche-hΛru gΛ-je **ho-lan**
now PAR people-PL go-3.PL.PST be-may
‘People might have left now’

3.3.3 Transitive Construction

Verbs in the Nepali language are of two types based on number of arguments they take in a sentence. They are Intransitive and Transitive Verbs.

- | | | |
|------|--------------------|--------------------|
| i. | keṭo | sut-jo |
| | Boy.NOM | sleep-3.SG.M.PST |
| | 'The boy slept' | |
| ii. | kukkur | mar-jo |
| | Dog.NOM | die-3.SG.M.PST |
| | 'The dog died' | |
| iii. | sikchak | a-je |
| | teacher.NOM | come-3.SG.M.MH.PST |
| | 'The teacher came' | |

Most often, an intransitive construction with the subject in a simple past tense tends to be in Nominative case in Nepali. However, the intransitive verb with the inherent meaning of some kind of physical movement can have a subject with an Ergative, as seen in the instances below:

- | | | |
|-----|-------------------|------------------|
| iv. | bhai-le | khok-jo |
| | brother-ERG | cough-3.SG.M.PST |
| | 'Brother coughed' | |
| v. | mai-le | thuk-ē |
| | I-ERG | spit-1.SG.PST |
| | 'I spat' | |

The other instances of intransitive verb roots in Nepali are **hās** 'laugh,' **bas** 'sit,' **cal** 'move,' **dʒa** 'go,' **baɽ** 'increase,' **laɽ** 'fall,' **luk** 'hide,' **ru** 'weep,' **dʒhar** 'fall,' **uɽ** 'fly' etc.

The Transitive Verb in the Nepali can have up to three objects in a sentence. Thus, the transitive verbs in Nepali are mono-transitive, di-transitive, or tri-transitive. The process of narrativization increases the valence of a verb resulting in the formation of tri-transitive construction. The following instances clarify the status of transitive constructions in the Nepali:

Mono-Transitive Verb:

- i. keti-haru-le git **ga-je**
girl-PL-ERG song.ACC sing-3.PL.PST
'The girls sang a song'
- ii. ma-sāga pustak **cha**
I-ASS book.ACC be-1.SG.NPST
'I have a book'

Di-Transitive Verb:

- i. ram-le us-ko bhai-lai miṭhai **di-jo**
Ram-ERG he-GEN brother-DAT sweet.ACC give-3.SG.M.PST
'Ram gave sweets to his brother'
- ii. us-le ma-lai phul **bec-jo**
He-ERG I-DAT flower.ACC sell-3.SG.PST
'He sold the flower to me'

Tri-Transitive Verb:

- i. mai-le us-dvara parjatāk-lai darjiling **ghum-a-ē**
I-ERG he-by tourist-ACC Darjeeling visit-CAUS-1.SG.PST
'I made him the tourist to visit Darjeeling'
- ii. ama-le susare-dvara chora-lai khana **khu-va-in**
Mother-ERG nurse-by son-DAT food.ACC feed-CAUS-3.SG.F.PST
'Mother made the nurse feed her son'

3.3.4 Passive Construction

Changing an active voice sentence into a passive one is referred as the passivization process. It is a kind of alternation process for Nepali verbs since the verbs in Nepali inflects for this process as in other Indian languages like Hindi. In Nepali, the morpheme **-i** is suffixed to a verb to make passive construction. For example –

- i. guru-le kitab paṭ-na **lga-je** (Active voice)
teacher-ERG book read-INF make-3.SG.PST
'The teacher asked (someone) to read the book'
- ii. guru-dvara kitab paṭ-na **lga-i-jo** (Passive voice)

teacher-ERG book read-INF make-3.SG.PST

'(Someone) was asked by the teacher to read the book'

Unlike the English language, the subject need not change its position while making a passive construction in Nepali. Rather the morph –i renders the idea about the voice of a particular expression.

3.3.5 Causative Construction

The Nepali language expresses the causation process morphologically because the process makes the verb alternate its form by affixing or changing the sound of a morpheme in between a verb root and copula. Prasain (2011: 124-127) has identified ‘five methods of causative formation’ in Nepali. They are as follows:

- By using-**a** suffix;
- By using-**a/al** suffix;
- By changing [ʌ → [a];
- By changing [u] → [o]; and,
- By -**a** insertion.

Table 2.3 presents the example sentences for the process of causativization by the methods mentioned above.

Table 3.3 Examples of Causative Formation in Nepali

Non-Causative Construction	Causative Construction
keṭo hās-jo boy.NOM laugh-3.SG.NF.PST 'The boy laughed'	keṭi-le keṭa-lai hās-a-jo girl-ERG boy-ACC laugh-CAUS-3.SG.NF.PST 'The girl makes the boy to laugh'
bhai gaṭi-ma bas-jo brother.NOM vehicle-LOC sit - 3.SG.NF.PST 'The brother sat on the vehicle.'	ram-le bhai-lai gaṭi-ma bas-a-jo/bas-al-jo Ram-ERG brother-ACC vehicle-LOC sit-CAUS- 3.SG.NF.PST 'Ram made the brother to sit on the vehicle'
kukkur mar-jo dog-NOM die-3.SG.NF.PST 'The dog died away'	bagh-le kukur mar-jo Tiger-ERG dog-ACC kill-3.SG.NF.PST 'The tiger killed the dog.'
dailo khul-jo door-NOM open-3.SG.NF.PST 'The door opened.'	hava-le dailo kh-o-l-jo wind-ERG door-ACC open-CAUS-3.SG.NF.PST 'The wind made the door open.'
gaṭi bigri-jo vehicle-NOM break-3.SG.NF.PST 'The vehicle broke down.'	hari-le gaṭi big-a-r-jo Hari-ERG vehicle-ACC break-CAUS-3.SG.NF.PST 'Hari made vehicle to break down'

The instances given above in Table 3.3 make an obvious description of the methods of causative formation in Nepali. In Nepali, causativization and transitivization involve the same processes of verb alternation. The difference is only that “in transitivization, an argument is added irrespective of the role of the argument, but in causativization, the added argument is a causer” (Prasain, 2011, p.122).

3.3.6 Negative Construction

In Nepali, **-na** is a negative marker. It is either prefixed or suffixed to a verb stem to make a negative sentence. Hence, the process of negation involved the alternation of a verb compulsorily in Nepali. Let us illustrate the case with the help of examples given in table 3.4 below:

Table 3.4 Negativization Process in Nepali

Non-Negative Construction				Negative Construction			
mΛ	tjo	kam	gar-chu	mΛ	tjo	kam	gar-di-na
I-NOM	that	work.ACC	do-1.SG.NPST	I-NOM	that	work .ACC	do-1.SG.NPST-NEG
'I will do that work'				'I will not do that work'			
Timi	bholi	ghar	ɖʒa-u	Timi	bholi	ghar	na-ɖʒa-u
you-NOM	tomorrow	home	go-2.SG.NPST	you-NOM	tomorrow	home	NEG-go-.SG.NPST
'You go home tomorrow'				'You don't go home tomorrow'			

In the examples above in Table 3.4, the first sentence demonstrates the negation process by suffixation, whereas the second sentence is about the negation process by prefixation. Formation of negative construction by suffixation or prefixation depends upon the tense, aspect, and mood information about a verb in Nepali.

3.4 Semantic Classification of Verbs

The available literature reveals a limited number of studies that have undertaken the task of categorizing Nepali verbs according to their semantic characteristics. Within this domain, notable works include those by Basnyat (1986), Wallace (1979), Pokharel (2054 BS), Adhikari (2055 BS), Pokharel (1999), and Pant (2000). Among these contributions, Pant's study from the year 2000 stands out as particularly significant. It distinguishes itself by its attempt to systematically categorize Nepali verbs into semantic

classes, drawing inspiration from the work of Vendler (1967) and Dowty (1979). Pant's classification scheme encompasses various dimensions, including:

Temporal Attributes: Discriminating between Durative and Punctual verbs, investigating the Progressive Aspect, and classifying verbs into Telic (with a specific endpoint) and Atelic (lacking a defined endpoint) categories.

Event Structure: Categorizing verbs based on their temporal and event structure characteristics into Achievement Verbs, Activity Verbs, Accomplishment Verbs, and State Verbs.

Volition: Distinguishing between Volitional (actions performed willingly) and Non-Volitional (actions not under one's control) verbs.

Stativity and Dynamicity: Examining whether verbs convey states or dynamic actions.

Pant's work represents a notable endeavor in systematically characterizing Nepali verbs based on their semantic properties, contributing valuable insights to the understanding of verb semantics within the Nepali language context.

In this study, an attempt has been made to classify Nepali verbs based on Levin's classification of English verbs (1993). Levin (1993) is a large-scale classification of English verbs that has made a significant contribution towards the building of lexical databases like VerbNet, FrameNet, Propbank, WordNet, etc. The numerous NLP tasks like Computational lexicography, Language Generation, Machine Translation, Word Sense Disambiguation, Semantic Role Labelling, Sub Categorisation Acquisition, etc., are benefitted from the work of Levin (1993). Since this study aims to build a database of verb frames, particularly for Verb Sense Disambiguation, and since Levin (1993) provides a detailed approach for the classification of verbs, it is useful to adopt her approach. Also, the verb frames developed in this study incorporates the description of the semantic class of a verb which has proved to be helpful in sense disambiguation as well. Levin (1993) has classified English verbs broadly into 48 classes which are further categorized into various sub-classes as follows:

1. **Verbs of Putting:** Put Verbs, Verbs of Putting in a Spatial Configuration, Funnel Verbs, Verbs of Putting with a Specified Direction, Pour Verbs, Coil Verbs, Spray/Load Verbs, Fill Verbs, Butter Verbs, and Pocket Verbs.

2. **Verbs of Removing:** Remove Verbs, Banish Verbs, Clear Verbs, and Wipe Verbs, Verbs of Possessional Deprivation: Steal Verbs, Verbs of Possessional Deprivation: Cheat Verbs, Pit Verbs, Debone Verbs, Mine Verbs.
3. **Verbs of Sending and Carrying:** Send Verbs, Slide Verbs, Bring and Take, Carry Verbs, Drive Verbs.
4. **Verbs of Exerting Force: Push/Pull Verbs**
5. **Verbs of Change of Possession:** Give Verbs, Contribute Verbs, Verbs of Future Having, Verb of Fulfilling, Equip Verbs, Verbs of Obtaining, Verbs of Exchange, Berry Verbs.
6. **Learn Verbs**
7. **Hold and Keep Verbs:** Hold Verbs, Keep Verbs.
8. **Verbs of Concealment**
9. **Verbs of Throwing:** Throw Verbs, Pelt Verbs.
10. **Verbs of Contact by Impact:** Hit Verbs, Swat Verbs, Spank Verbs, Non-Agentive Verbs of Contact by Impact.
11. **Poke Verbs**
12. **Verbs of Contact: Touch Verbs**
13. **Verbs of Cutting:** Cut Verbs, Carve Verbs.
14. **Verbs of Combining and Attaching:** Mix Verbs, Amalgamate Verbs, Shake Verbs, Tape Verbs, and Cling Verbs.
15. **Verbs of Separating and Disassembling:** Separate Verbs, Split Verbs, Disassemble Verbs, Differ Verbs.
16. **Verbs of Coloring**
17. **Image Creation Verbs:** Verbs of Image Impression, Scribble Verbs, Illustrate Verbs, Transcribe Verbs.
18. **Verbs of Creation and Transformation:** Build Verbs, Grow Verbs, Verbs of Preparing, Create Verbs, Knead Verbs, Turn Verbs, and Performance Verbs.
19. **Engender Verbs**
20. **Calve Verbs**

21. **Verbs with Predicative Complements:** Appoint Verbs, Characterize Verbs, Dub Verbs, Declare Verbs, Conjecture Verbs, Masquerade Verbs, Orphan Verbs, and Captain Verbs.
22. **Verbs of Perception:** See Verbs, Sight Verbs, Peer Verbs, and Stimulus Subject Perception Verbs.
23. **Psych-Verbs Verbs of Psychological State:** Amuse Verbs, Admire Verbs, Marvel Verbs, Appeal Verbs.
24. **Verbs of Desire:** Want Verbs, Long Verbs.
25. **Judgment Verbs**
26. **Verbs of Assessment**
27. **Verbs of Searching:** Hunt Verbs, Search Verbs, Stalk Verbs, Investigate Verbs, Rummage Verbs, and Ferret Verbs.
28. **Verbs of Social Interaction:** Correspond Verbs, Marry Verbs, and Meet Verbs.
29. **Verbs of Communication:** Verbs of Transfer of a Message, Tell Verb, Verbs of Manner of Speaking, Verbs of Instrument of Communication, Talk Verbs, Chitchat Verbs, Say Verbs, Complain Verbs, Advice Verbs.
30. **Verbs of Sounds Made by Animals**
31. **Verbs of Ingesting** (Eat Verbs, Chew Verbs, Gobble Verbs, Devour Verbs, Dine Verbs, Gorge Verb, Verbs of Feeding)
32. **Verbs Involving the Body:** Verbs of Bodily Processes, Verbs of Nonverbal Expression, Verbs of Gestures/Signs Involving Body Parts, Snooze Verbs, Flinch Verbs, Verbs of Body-Internal States of Existence, Suffocate Verbs, Verbs of Bodily State and Damage to the Body.
33. **Verbs of Grooming and Bodily Care:** Verbs of Caring for the Whole Body, Verbs of Caring for a Specific Body Part, Verbs of Dressing.
34. **Verbs of Killing:** Murder Verbs, Poison Verbs.
35. **Verbs of Emission:** Verbs of Light Emission, Verbs of Sound Emission, Verbs of Smell Emission, Verbs of Substance Emission.
36. **Destroy Verbs**

37. **Verbs of Change of State:** Break Verbs, Bend Verbs, Cooking Verbs, Verbs of Entity-Specific Change of State, Verbs of Calibratable Changes of State.
38. **Lodge Verbs**
39. **Verbs of Existence:** Exist Verbs, Verbs of Entity-Specific Modes of Being, Verbs of Modes of Being Involving Motion, Verbs of Sound Existence, Verbs of Group Existence, Verbs of Spatial Configuration, Meander Verbs, Verbs of Contiguous Location.
40. **Verbs of Appearance, Disappearance, and Occurrence:** Verbs of Appearance, Verbs of Disappearance, Verbs of Occurrence.
41. **Verbs of Body-Internal Motion**
42. **Verbs of Assuming a Position**
43. **Verbs of Motion:** Verbs of Inherently Directed Motion, Manner of Motion Verbs, Verbs of Motion Using a Vehicle, Waltz Verbs, Chase Verbs, Accompany Verbs.
44. **Verbs of Lingering and Rushing:** Verbs of Lingering, Verbs of Rushing.
45. **Measure Verbs:** Register Verbs, Cost Verbs, Fit Verbs, Price Verbs, Bill Verbs.
46. **Aspectual Verbs:** Begin Verbs, Complete Verbs.
47. **Weekend Verbs**
48. **Weather Verbs**

The detailed list of Nepali verb classes with their class members is given in the Appendix III of this thesis.

CHAPTER 4

ANALYSIS OF VERB FRAMES DATABASE

4.1 Introduction

This chapter is organized into five main sections, each containing corresponding sub-sections. Section 4.2 delves into the theoretical concept of verb frames as proposed in generative linguistics, examining their utility in natural language processing. In Section 4.3, we explore the implicational aspects associated with verb frames. Section 4.4 provides an in-depth breakdown of the various components comprising verb frames. Lastly, Section 4.5 is dedicated to the semantic categorization of Nepali verbs using distinctive verb frames.

4.2 What is a Verb Frame?

The verb stands as a fundamental grammatical category within a language, assuming a central role in shaping the meaning of a given sentence or discourse. By furnishing the relational and semantic structure within a sentence, it holds an indispensable position, rendering sentence construction unattainable in its absence. Consequently, in contemporary times, the linguistic understanding of verbs has emerged as a primary focal point within the realm of automatic language processing.

The concept of the verb frame finds its origins in the work of Chomsky (1965), who introduced the notion of a 'sub-categorization frame.' According to Chomsky (1965), sub-categorization refers to “the ability/necessity for lexical items (typically verbs) to specify the presence and types of syntactic arguments with which they can co-occur”. While both sub-categorization and valency concepts address the count and roles of arguments within a sentence, they initially differed in their scope. Sub-categorization, in its original definition, excluded the subject and focused solely on complements, objects, and oblique arguments. Modern theories have since expanded the sub-categorization frame to include the subject. Conversely, valency always encompassed the count of arguments, encompassing the subject from the outset. In this context, sub-categorization has evolved to align with valency, as contemporary phrase structure grammars now consider verbs as subcategorizing for both their subjects and objects.

In this study, the specific concept of verb frames can be traced back to the foundational work of Levin (1993). Levin's work served as a catalyst for subsequent verb-focused research projects, such as VerbNet, FrameNet, PropBank, WordNet, and others. These projects, in turn, played a pivotal role in advancing and refining the notion of verb frames, contributing significantly to its development and understanding.

In this study, the term "verb frame" refers to a structured representation of a verb, organized in tabular form, which encapsulates linguistic details concerning “the syntactic usage of a verb within a language” (Begum, 2017). This verb frame comprises linguistic information related to a verb, primarily categorized into two parts: the Verb Description and the Verb Frame.

The Verb Description encompasses various elements, including the verb's name, its specific sense, an English gloss or translation, its verb class, an illustrative example sentence, and the roles played by its arguments, known as theta roles, among others.

In contrast, the Verb Frame provides information regarding karaka relations, vibhakti (case markings or inflections), the necessity of arguments, lexical category, and ontological aspects associated with the verb. A comprehensive explanation of each field of information is provided in section 4.4 of this chapter.

4.3 Importance of Verb Frame

The verb holds a fundamental position within a sentence, serving as a cornerstone in the allocation of functions to various arguments. In the Paninian Grammatical tradition, verbs are regarded as the parent of all arguments, exerting comprehensive control over the conduct and interactions of these arguments. A verb frame assumes significance due to its ability to provide an extensive examination of a specific verb, encompassing aspects from its various senses to its argument structures. The importance of the verb frame becomes evident through the following aspects:

4.3.1 A Knowledge Base

A verb frame serves as a linguistic asset functioning as a knowledge repository, as it contains extensive details about a specific verb. This resource proves beneficial in

various knowledge-driven natural language processing (NLP) endeavors and applications, including tasks such as word sense disambiguation and machine translation.

4.3.2 Linguistic Analysis of Verb

The verb frame offers an extensive portrayal of a verb's characteristics, encompassing both lexico-syntactic and semantic dimensions. It encapsulates a wide array of linguistic information pertaining to a specific verb, including its sense, class, transitivity details, corpus-based examples, argument structure, theta roles, dependency relations, and the ontology of associated arguments. This wealth of linguistic information equips researchers with a profound understanding of a verb, enabling them to conduct comprehensive investigations into Nepali verbs for various research purposes.

4.3.3 Annotation of Verbs

Annotation of linguistic units at morphological, lexical, and syntactical levels is necessary to build various NLP applications. Morphological Analysis, Parts of Speech (POS) Tagging, and Parsing are three kinds of annotations for the above-mentioned linguistic units. Verb frames are useful for POS Tagging and Parsing since they provide information about the sentence's lexical category and dependency relations of arguments.

4.3.4 A Resource for Verb Sense Disambiguation

The main objective of developing verb frames in this study is to find possible strategies for handling ambiguous verbs. It aims to depict every linguistic information about each sense of an ambiguous verb. The verb frames in this study take account of the 200 most frequently used ambiguous Nepali verbs with separate frames for each sense of the verb. They represent detailed information about lexico-syntactic and semantic attributes of a verb which provides linguistic cues for their disambiguation task.

4.4 Constituents of Verb Frames

Verb frames in this study comprise two essential constituents: i) Verb Description and ii) Verb Frame. These components are organized within a data file referred to as 'verb entry.' Figure 4.1 below illustrates the structure of each constituent within the verb frame.

Figure 4.1 Verb Frame for the first sense of verb ‘au’

```

Verb::au
SID::au%VI%S1
Verb_Sense::agaman_hu
Eng_Gloss::Come
Verb_class::Verb of motion
Verbs_in_Same_Class::Synonyms>
Example::u          ghar          a-jo
          he.NOM   home.LOC   come-3.SG.PST
          'He came home'
Theta_Roles::AGENT GOAL VERB
Frame_ID::au%VI%S1%FID1
-----
arc-label  necessity  vibhakti  lextype  ontology
-----
k1          m          0          p/n      [+ani/+hum]
k2p         m          0          n        [+plc]
-----
*****

```

4.4.1 Description

The verb description encompasses nine distinct elements, including the verb's name, sense ID, verb sense, English gloss, verb class, verbs within the same class, an example sentence, theta roles, and frame ID. This segment of the verb frame serves as a repository of lexico-semantic details pertaining to the specific verb. The subsequent subsections offer comprehensive explanations of each individual component.

4.4.1.1 Verb Name

The verb name in Nepali is the specific identifier for a given verb. In Figure 4.1 'au' represents the verb name, and it is presented in the International Phonetic Alphabet (IPA) to accurately convey its pronunciation.

4.4.1.2 Sense ID

The Sense ID serves as a distinctive identifier for a specific sense of a verb. It comprises the verb name, verb type, and sense number, separated by the '%' symbol. In

Figure 4.1, the Sense ID for the first sense of the verb 'au' is represented as 'au%VI%S1,' where 'au' denotes the verb name, 'VT' signifies a 'transitive verb,' and 'S1' indicates 'Sense 1' of the verb. Our verb frame categorizes verbs into three types: Transitive (VT), Di-transitive (VDT), and Causative (VCAUS). Distinct verb senses are identified using the convention of S1, S2, S3, and so forth, corresponding to the number of senses a verb may possess.

4.4.1.3 Verb Sense

The verb sense signifies the synonymous meaning of a particular verb in Nepali. In Figure 4.1, the verb sense for 'au' is 'agaman_hu' ('come'), and it is denoted as 'Verb_Sense' within the verb frame.

4.4.1.4 English Gloss

The English gloss corresponds to the meaning of a verb in the English language. Within the verb frame, it is designated as 'Eng_Gloss.' In Figure 4.1, 'come' serves as the English gloss for the specific sense of the verb 'au.'

4.4.1.5 Verb Class

The verb class represents a semantic categorization of the verb based on Levin (1993). In Figure 4.1, 'come' is classified as 'Verb of Motion.' This inclusion of information in the study is valuable as it contributes to the disambiguation task of Nepali verbs. Verbs belonging to the same semantic class tend to exhibit similar behavior and share comparable argument structures. Moreover, it aids in the categorization of Nepali verbs into specific semantic classes, enhancing our understanding of their usage patterns and meaning distinctions.

4.4.1.6 Verb in Same Class

This information pertains to synonymous verb frames within the database. Specifically, it signifies that several verb frames with similar senses are considered synonymous, either because they share the same sense or belong to the same verb class. Verbs within the same class are identified by their unique frame ID. This data regarding

the same class is essential for assessing whether verbs within the same class exhibit similar behaviour.

Figures 4.2 and 4.3, provided below, serve as illustrations of instances where verbs belong to the same class. For instance, the second sense of the verb 'au' (au%VT%S2%FID2) and the second sense of the verb 'dʒan' (dʒan%VT%S1%FID1) not only share the same sense but also exhibit identical argument structures, semantic roles, dependency relations, and ontological features of the arguments. This type of information facilitates the grouping of verbs into specific semantic categories, enhancing our understanding of their shared characteristics and usage patterns.

Figure 4.2: Verb Frame for the verb second sense of verb ‘au’

```

SID::au%VT%S2
Verb_Sense::ʒan
Eng_Gloss::Know
Verb_class::Verb of perception
Verbs_in_Same_Class::Synonyms>ʒan%VT%S1%FID1
Example::us-lai dherai kura au-cha
          He-DAT many things.ACC know-3.SG.NPST
          'He knows many things'
Theta_Roles::EXPERIENCER TOPIC VERB
Frame_ID::au%VI%S2%FID2
-----
arc-label  necessity  vibhakti  lextype  ontology
-----
k4a        m         lai       n/p     [+ani/+hum]
k2         m         0        n       [+top]
-----
*****

```

Figure 4.3 Verb Frame for the second sense of verb ‘dʒan’

```

Verb::dʒan
SID::dʒan%VT%S1
Verb_Sense::thahΛ_pau
Eng_Gloss::Know
Verb_class::Verb of perception
Verbs_in_Same_Class::Synonyms>au%VT%S2%FID2
Example::tapaĩ        jo        kuro        dʒan-nu-hun-ca
          you.NOM    this  thing.ACC  know-INF-HH-2.SG.NPST
          'You know this thing'
Theta_Roles::EXPERIENCER TOPIC VERB
Frame_ID::dʒan%VT%S1%FID1
-----
arc-label  necessity  vibhakti    lextype    ontology
-----
k1          m          0          p/n        [+ani/+hum]
k2          m          0          n          [+top]
-----
*****

```

4.4.1.7 Example

This part of the verb frame comprises an example sentence for each sense of a verb, transcribed in IPA. The glossing for each example sentence follows the interlinear morpheme-by-morpheme glossing guidelines established by the Leipzig Glossing Rules (2015). For instance, the following example is provided for the verb 'au' in Figure 4.1:

```

u          ghΛr        a-jo
he.NOM     home.LOC    come-3.SG.PST
'He came home'

```

4.4.1.8 Semantic Roles

Semantic roles denote the specific functions carried out by each argument within a predicate, defined with respect to a limited but universal set of thematic functions (Crystal, 2003, p. 463). In our verb frame, we provide precise definitions for each of these semantic roles, which are as follows:

- a. Actor: Used in communication classes like verbs of social interaction and meeting when both arguments are symmetrical.

- b. Agent: Typically represents a human or animate subject and is primarily associated with volitional agents but can also be used for internally controlled subjects, such as forces.
- c. Asset: Employed in verb classes related to financial transactions, such as building and obtaining verbs, with 'currency' as a selectional restriction.
- d. Attribute: Refers to a quality of something being changed, as observed in verbs like 'price.'
- e. Beneficiary: Denotes the entity that benefits from a particular action, often used in the context of building, obtaining, and performance verbs.
- f. Cause: Primarily used in classes involving psychological states and verbs related to the body.
- g. Destination: Indicates the endpoint of motion or the direction toward which motion is directed. Applied in verb classes related to motion, sending, and carrying.
- h. Source: Represents the starting point of motion, commonly used in verbs of motion.
- i. Location: Refers to an underspecified destination, source, or place, often introduced by locative or path prepositional phrases.
- j. Experiencer: Applied to a participant who is aware of or experiencing something, commonly found in classes involving psychological states, verbs of perception, and verbs related to the body.
- k. Instrument: Used for objects or forces that come into contact with an object and cause a change in them.
- l. Material and Product: Employed in the context of building and growing classes to capture critical semantic components of arguments. Found in classes of verbs related to creation and transformation that allow for material/product alternation.
- m. Patient: Used for participants undergoing a process or those affected in some way. Commonly associated with verbs that explicitly or implicitly express state changes, often serving as the direct object.
- n. Predicate: Reserved for classes with a predicative complement.
- o. Recipient: Refers to the target of a transfer, often found in classes related to changes of possession, communication, and verbs involving the body.

- p. Stimulus: Employed by verbs of perception to describe events or objects that elicit a response from an experiencer.
- q. Theme: Represents participants in a location or changing location.
- r. Time: Refers to the class-specific role used in the 'begin' class to express time.
- s. Topic: Used in communication verbs to handle the theme or topic of conversation or message transfer.

These semantic roles provide a structured framework for understanding the functions of arguments within different verb predicates and contribute to the thorough analysis of verb behavior in Nepali.

4.4.1.9 Frame ID

The Frame ID serves the purpose of providing a distinctive name and unique identifier to each verb frame associated with each sense of a specific verb. In Figure 4.1, 'au%VI%S1%FID1' represents the Frame ID for the first sense of the verb 'au.' This Frame ID comprises two components: the Sense ID (e.g., 'au%VI%S1') and the Frame Number (e.g., 'FID1'), which are separated by a percentage sign. This identifier is essential for distinguishing and pinpointing a particular frame corresponding to a specific sense of a verb.

4.4.2 Verb Frame

In this study, the term 'verb frame' is employed in two distinct senses. In a broad sense, it encompasses the entirety of the data, which includes the 'verb entry' comprising both the 'description of the verb' and the 'verb frame.' In a specific sense, it refers to the tabular representation of a verb's argument structure, capturing information regarding dependency relations, necessity, vibhakti, lextypes, and ontology of arguments. To clarify, the verb frame, as defined by Begum (2017, p. 99), consists of six pieces of information about the argument structure of each sense of a verb. However, in our version of the verb frame, we have omitted two pieces of information, 'position' and 'relation,' while adding details about 'ontology.' Consequently, our verb frames encapsulate a total of five pieces of information, which are as follows:

4.4.2.1 Arc Label

The dependency relation, also known as the karaka relation, is represented in the first field of the verb frame, denoted as 'arc label.' This field serves to capture and signify the specific relationship between the verb and its associated arguments, providing crucial information about the syntactic and semantic structure of the sentence. The 'karaka' is a term used in Panini's Grammar *Astadhyayi* to denote 'the special relations of nouns, adjectives, and pronouns with the words expressing action' (Bhatta, 1998-99: 15). It is similar to thematic relation. However, they differ as 'one of the key features of the karaka relation is that they are verb centric' (Gorthi et al, 2014). It means karaka relation depends on the sentence's main verb. The karaka relations are expressed through case markers or post-nominal suffixes in Indian languages. There are mainly six types of karakas according to the karaka theory. They are - karta (Agentive), karma (Objective), karana (Instrumental), sampradana (Dative), apadana (ablative) and adhikarana (locative). In our verb frames, the annotation of each type and subtype of karaka is carried out in accordance with the "AnnCorra: TreeBanks for Indian Languages Guidelines for Annotating Hindi Tree Bank", as developed by IIIT Hyderabad (Bharati et al., 2012). Table 4.1 provides a comprehensive description of the various types and subtypes of karakas incorporated into our verb frames. This annotation methodology ensures consistency and adherence to established guidelines in the representation of linguistic information.

Table 4.1 karakas and other dependency relations

S. No.	<i>karaka</i>	Convention	Description
1	<i>karta</i>	k1	doer/agent/subject
		pk1, jk1, mk1	causer, causee, mediator-causer
2	<i>karma</i>	k2	object/patient
		k2p	Goal, Destination
		k2g	Secondary karma
3	<i>karana</i>	k3	Instrument
4	<i>sampradana</i>	k4	Recipient
	<i>anubhava karta</i>	k4a	Experiencer
5	<i>apadana</i>	k5	Source
	<i>prakruti apadana</i>	k5prk	source material
6	<i>vishayadhikarana</i>	k7	location elsewhere
	<i>kala adhikarana</i>	k7t	location in time

	<i>deshadhikarana</i>	k7p	location in space
7	<i>sambandha</i>	r6	genitive/possessive
8	<i>kriyavishesana</i>	adv	adverbs (only manner adverbs)
9	<i>Relation (prati)</i>	rd	Direction
10	<i>hetu</i>	rh	Reason

Table 4.1 illustrates the approach for representing karaka information within the verb frame. The sentence provided in Figure 4.1 exhibits two dependencies: k1 'u' and k2p 'ghar,' which correspond to the karaka roles associated with 'u' and 'ghar.' These dependencies are documented in the first field of the verb frame, denoted as the 'arc label,' offering a structured and organized representation of the syntactic and semantic relationships within the sentence.

4.4.2.2 Necessity of the argument

The second field in the verb frame, labeled as 'necessity,' conveys information regarding the requirement of arguments within a sentence. Arguments, in this context, are defined as "the participants that help a predicate express a certain proposition in a sentence" (Aarts, 2001, p. 91). In simpler terms, expressions with specific semantic roles are considered arguments. In the example sentence provided in Figure 4.1, 'u' ('he') and 'ghar' ('home') are two arguments of the predicate 'a-jo' ('came'), and they assume the semantic roles of 'agent' and 'location,' respectively.

Arguments can be classified as either 'mandatory' or 'desirable' based on the proposition conveyed by the predicate. The Nepali verb 'au-nu' ('to come') is a diadic predicate, meaning it requires both arguments in the sentence 'u ghar a-jo' ('he came home'). However, some arguments can be classified as 'desirable,' meaning they are needed for the verb's semantics but are comparatively less essential. In other words, "they can be omitted without significantly affecting communication and can generally be inferred from the context" (Begum, 2017, p. 99).

For instance, consider the sentences 'the door is opened' versus 'the door is opened by me.' In the latter sentence, 'me' is a desirable argument that can be omitted without causing a breakdown in communication. Therefore, the 'necessity' field in the verb frame includes conventions such as 'm' to denote mandatory arguments and 'd' to signify

desirable arguments, thereby capturing the essential information about the argument's necessity in a sentence.

For instance, consider the sentences 'the door is opened' versus 'the door is opened by me.' In the latter sentence, 'me' is a desirable argument that can be omitted without causing a breakdown in communication. Therefore, the 'necessity' field in the verb frame includes conventions such as 'm' to denote mandatory arguments and 'd' to signify desirable arguments, thereby capturing the essential information about the argument's necessity in a sentence.

4.4.2.3 vibhakti

The third field in the verb frame, labeled as 'vibhakti,' serves to convey information about the type of case marker or any other postpositional element associated with a particular argument. In the example sentences provided in Figure 4.4, there are three different vibhakti markers: '-le' (ergative), '-baṭṭa' (ablative), and '-lai' (dative), each attached to their respective arguments ('us,' 'd̪el,' 'ʌpʌradhi'). In Indian languages like Nepali, vibhakti information provides valuable insights into the morpho-syntactic functions of a given argument within a proposition. It helps elucidate the role and relationship of each argument in the sentence, contributing to a more comprehensive understanding of sentence structure and meaning

Figure 4.4: Verb Frame for the third sense of the verb ‘ukas’

```

SID::ukas%VT%S3
Verb Sense::mukṭa gar
Eng_Gloss::Bail out
Verb_class::Verb of removing
Verbs_in_Same_Class::Synonyms>uker%VT%S2%FID2
Examlpe::us-le ḡel-baṭṭa ʌpʌradhi-lai ukas-jo
           he-ERG jail-ABL culprit-ACC bail-3.SG.NPST
           'He bailed the culprit out from the jail'
Theta_Roles::AGENT SOURCE PATIENT VERB
Frame_ID::ukas%VT%S3%FID3
-----
arc-label  necessity      vibhakti      lextype      ontology
-----
k1          m             le/0          p/n          [+ani/+hum]
k5          d             baṭṭa/dekhi   n            [+plc]
k2          m             lai           p/n          [+ani/+hum]
-----
*****

```

4.4.2.4 Lexical Category

The information about the grammatical categories of an argument is represented in the 'lextype' field, which is the short form of 'lexical type' or 'lexical category.' In the example sentence, as depicted in Figure 4.4, all three arguments have lexical categories classified as nouns, and they are represented as 'n' in the verb frame. This lexical category information indicates that the arguments in the sentence are all nouns, providing valuable insights into the grammatical structure of the sentence. However, arguments of the verb 'ukas' can be pronouns also, that is why 'p' is written as an optional lexttype of the arguments.

4.4.2.5 Ontology

The term 'ontology' refers to the hierarchical collection of concepts and the relationship among those concepts. In other words, it is a knowledge base that provides information about the concepts existing in the world, their properties, and their relation to each other. King and Reinold (2008) assume that three kinds of concepts are represented in ontology. They are-

- a) Classes (general things) in the many domains of interest;
- b) The relationship that can exist among things; and,
- c) The properties (or attributes) those things may have.

Studying concepts based on ontology is necessary because it provides comprehensive, precise, consistent, and meaningful distinctions about the classes, properties, and relations of concepts. In this context, we have used the term 'ontology' in a particular sense that refers "to the collection of concepts arranged in a hierarchy of categories combined with the relationships between those concepts".

Indeed, ontology serves as a knowledge base that captures information about the semantic properties of concepts and their relationships with one another. It plays a crucial role in determining a verb's selectional restrictions, aiding in the disambiguation of words with multiple senses. This approach has been recognized as a practical method for word sense disambiguation by researchers such as Kang (2003; 2004) and Hung et al. (2009). Consequently, in this study, a field that provides information about the ontological attributes of arguments has been introduced.

The tag set devised for the 'Shared Task cum Workshop on OntoLex in Indian Languages,' organized by the Center for Applied Linguistics and Translation Studies at the University of Hyderabad in 2017, has been employed for the annotation of ontology within our verb frames. This standardized tagging system ensures consistency and compatibility in capturing ontological information, enhancing the study's effectiveness in analyzing verb semantics and disambiguating word senses.

Figure 4.5 demonstrates the hierarchical tag set used to annotate ontological features.

Figure 4.5 Tag Set for Ontology

1.	Concrete [+con]
	+Living, +Natural [+liv, +nat]
	+Animate [+ani]
	+Human [+hum]
	(+Infant [+inf], +Child [+chi], +Adolescence [+ado], +Adult [+adu],
	+Old Age [+old]
	(+Kinship [+kins {+adr, -adr}])
	-Human [-hum]
	+Animal [+anm], +Reptile [+rep], +Bird [+brd], +Micro Organism
	[+mic_org]
	-Animate [-ani]
	+Flora [+flo], (+Shrub [+shrb], +Plant [+plnt], +Climber [+clmb], +Tree
	[+tree])
	-Living [-liv]
	-Artifact, +Natural [-artf, +nat]
	+Solid [+sol]
	+Place [+plc], +Object [+obj]
	-Solid [-Sol]
	+Liquid [+liq], +Gas [+gas]
	+Artifact [+artf]
	+Object [+artfobj]
	+Place [+artfplc]
2.	-Concrete [-con] or +Abstract [+abs]
	+Cognition [+cog]
	+State [+sta]
	+Disease [+dis], +Biological State [+biosta], +Mental State [+mensta], +Social
	State [+sosta]
	+Skill [+ski]
	+Event [+evnt]
	+Natural Event [+natevnt], +Historical Event [+hisevnt], +Planned Event
	[+plnde vnt], +Social Event [+soevnt], +Fateful Event [+ftflevnt], +Fatal
	[+fatevnt]
	+Action [+act]
	+Physical Action [+phyact], +Communication [+com]
	+Quality [+qual]
	+Measurement [+msmt]
	+Time [+tme]
	+Period [+prd], +Season [+sea], +Historical Age [+hisage]
	+Direction [+dir]
	+Onomatopoeic Sound [+ona]
	+Terminologies [+ter]
	+Miscellaneous [+misc]

The ontology assigned to the arguments in the example sentence presented in Figure 4.4 is as follows: '+ani/+hum' (+animate/+human) for 'u' ('he'), '+plc' (+place) for 'jail,' and '+ani/+hum' for 'अपराधी' ('culprit'). This ontological information about the arguments proves invaluable in disambiguating a verb with multiple senses, such as 'ukas' in Nepali, which has three distinct senses: 'pull out,' 'bail out,' and 'provoke.' Figure 4.6 serves as an illustration of how an argument with different ontological features plays a pivotal role in determining the meaning of a specific verb, like 'ukas.' This highlights the significance of ontological attributes in disambiguating word senses and enhancing our understanding of verb semantics in Nepali.

Figure 4.6 Verb Frames for the verb ‘ukas’ having two different senses

Verb::ukas				
SID::ukas%VT%S1				
Verb_Sense::uthau				
Eng_Gloss::Pull out				
Verb_class::Verb of exerting Force				
Verbs_in_Same_Class::Synonyms>ukas%VT%S1%FID1,tan%VT%S1%FID1				
lAchar%VT%S1%FID1				
Example::us-le khaḍal-baṭa qhunggo ukas-jo				
he-ERG pit-ABL stone.ACC pull-3.SG.PST				
'He pulled out the stone from the pit'				
Theta_Roles::AGENT SOURCE THEME VERB				
Frame_ID::ukas%VT%S1%FID1				

arc-label	necessity	vibhakti	lextype	ontology

k1	m	le/0	p/n	[+ani/+hum]
k5	d	baṭa/dekhi	n	[+plc]
k2	m	0/lai	p/n	[+con]

SID::ukas%VT%S2				
Verb_Sense::utteḡit_par				
Eng_Gloss::Provoke				
Verb_class::Verb of psychological State				
Verbs_in_Same_Class::Synonyms>uker%VT%S2%FID2,uksi%VT%S6%FID6				
ucal%VT%S2%FID2				
Example::us-le tjaṣ keṭa-lai ukas-jo				
he-ERG that boy-ACC provoke-3.SG.PST				
'He provoked that boy'				
Theta_Roles::AGENT PATIENT VERB				
Frame_ID::ukas%VT%S2%FID2				

arc-label	necessity	vibhakti	lextype	ontology

k1	m	le/0	p/n	[+ani/+hum]
k2	m	lai	p/n	[+ani/+hum]

Figure 4.4 and 4.6 above, represented the verb frames for three senses of verb ‘ukas’ in Nepali: 'bailout,' 'pull out,' and 'provoke,' respectively which have three different argument structures and ontologies as shown in Table 4.2 below.

Table 4.2 Argument Structure and Ontology for different senses of ‘ukas’

Example	Argument Structure	Ontology
us-le dʒel-baʈʌ ʌpʌradhi-lai ukas-jo He-ERG jail-ABL culprit-DAT bail-3.SG.M.PST 'He bailed out the culprit from the jail'	{1<NP (AGENT)>, 2 <NP (SOURCE)>, 3<NP (PATIENT)>}	1 <NP> = [+ani/+hum] 2 <NP> = [+plc] 3 <NP> = [+ani/+hum]
us-le khaɖʌl-baʈʌ dʒuŋgo ukas-jo He-ERG pit-ABL stone-ACC pull-3.SG.PST 'He pulled out the stone from the pit'	{1<NP (AGENT)>, 2<NP (SOURCE)>, 3<NP (THEME)>}	1 <NP> = [+ani/+hum] 2 <NP> = [+plc] 3 <NP> = [+con]
us-le tʃʌs keʈʌ-lai ukas-jo He-ERG that boy-DAT provoke-3.SG.PST 'He provoked that boy'	{1<NP(AGENT)>, 2<NP(PATIENT)>}	1 <NP> = [+ani/+hum] 2 <NP> = [+ani/+hum]

As shown in Table 4.2 above, the example sentences for three different senses of verb ‘ukas’ has three different argument structures with different theta roles and ontology. This kind of information about an ambiguous verb can provide a significant cue for the disambiguation task.

Begum (2017) has incorporated two more pieces of information in her verb frames for Hindi. They are – Position and Relations. The position of a particular argument to the verb is captured in the 'posn' field which can be left or right.

Similarly, the relation of a verb to an argument is captured in the ‘reln’ field. However, the Paninian dependency grammar treats arguments as children of the verb, and the verb is considered a parent to them. Therefore, the relationship between the verb and its arguments is represented through a parent-child relation, indicated by 'p' and 'c' within her verb frame. This parent-child relationship helps establish the syntactic and semantic connections between the verb and its associated arguments, providing a structured and organized representation of the sentence's structure and meaning. For example, if we see in Figure 4.4 above, the example sentence given has three arguments ‘us’, ‘dʒel’ and ‘ʌpʌradhi’ respectively, and they can be considered as the children of verb ‘ukas-jo’. We have remove those two pieces of information from the verb frame since both 'position' and 'relation' remain the same in all verb frames. Because Indian languages like Nepali are SOV dominant phrase order languages, arguments' position to their predicate remains

left in almost all cases. Likewise, since the dependency grammar treats arguments as children of a verb, their relation also remains constant.

4.5 Statistics of Verb Frames

In this study, a total of 482 verb frames are created for 200 ambiguous Nepali verbs. The verb entries were selected, looking at their frequency of occurrences and the apparent ambiguity. The verbs having two to fifteen senses were taken for developing verb frames, and different verb frames were created for every sense of the verb since each sense shows differences in their semantic and syntactic structure. Table 4.3 demonstrates the number of verbs having a different number of senses:

Table 4.3 above shows that the verbs having two ambiguous senses are most common in Nepali. Out of 200 verbs, 166 verbs have two senses.

4.6 Classification of Nepali Verbs Based on Verb Frames

An attempt has been made here to classify Nepali verbs based on verb frames. In other words, the verbs having the same argument structure, especially the same dependency relation, have been grouped to see whether or not such classification pursues some understanding of the syntactic-semantic behavior of those verbs. The approach presented by Begum (2017) has been devised here for the classification.

Table 4.4 demonstrates the number of unique verb frames with their frequency of occurrence. The data in the table shows that we can classify 482 verb frames into 36 classes according to the unique pattern of dependency relations possessed by 482 senses of 200 Nepali verbs. It also exhibits the fact that the most frequent frames are 'k1+k2+verb' (166 occurrences) and 'k1+verb' (103 occurrences).

However, 'k1+k7t+k2+verb', 'k7t+k1+verb', 'rh+k2+verb', 'k5+k2+verb', 'k4+k3+verb', 'k1+k2+k3+verb', 'k1+k3+k2+verb', 'k1+k2+rd+verb', 'ras-k+k2+verb', 'k3+k7p+verb', 'k4a+k3+verb', 'pk1+mk1+k2+verb' are the least occurring frames with only one occurrence.

Table 4.3 Unique Verb Frames with their frequency of occurrence

S. No.	Frame No.	Dependency Relation	Frequency of Occurrence
1	Frame – 1	k1+k2	214
2	Frame – 2	k1+verb	105
3	Frame – 3	k1+k7p+verb	15
4	Frame – 4	k1+k7+verb	14
5	Frame – 5	k7+k1+verb	14
6	Frame – 6	k1+k5+verb	13
7	Frame – 7	k1+k7+k2+verb	12
8	Frame – 8	k7p+k1+verb	10
9	Frame – 9	k4a+k2+verb	9
10	Frame – 10	k1+rd+verb	8
11	Frame – 11	k1+k7p+k2+verb	7
12	Frame – 12	k1+k5+k2+verb	7
13	Frame – 13	k1+k3+verb	6
14	Frame – 14	k1+ras-k+verb	4
15	Frame – 15	k3+k2+verb	4
16	Frame – 16	k1+ras-k+k2+verb	3
17	Frame – 17	k1+k2+k7p+verb	3
18	Frame – 18	k5+k1+verb	3
19	Frame – 19	k1+k2p+verb	3
20	Frame – 20	k1+k4+k2+verb	3
21	Frame – 21	k1+k2+k7+verb	3
22	Frame – 22	k1+rsp+verb	3
23	Frame – 23	k1+k2+rh+verb	2
24	Frame – 24	k4+k2+verb	2
25	Frame – 25	k1+k7t+k2+verb	1
26	Frame – 26	k7t+k1+verb	1
27	Frame – 27	rh+k2+verb	1
28	Frame – 28	k5+k2+verb	1
29	Frame – 29	k4+k3+verb	1
30	Frame – 30	k1+k2+k3+verb	1
31	Frame – 31	k1+k3+k2+verb	1
32	Frame – 32	k1+k2+rd+verb	1
33	Frame – 33	ras-k+k2+verb	1
34	Frame – 34	k3+k7p+verb	1
35	Frame – 35	k4a+k3+verb	1
36	Frame – 36	pk1+mk1+k2+verb	1

It is also observed that the most frequent verb frame with dependency relation of ‘k1+k2+verb’ is transitive, whereas the second most frequent verb frame having ‘k1+verb’ relation is intransitive one.

Out of 482 verb frames 214 are having ‘k1+k2+verb’ construction and transitivity is their common feature. An instance of verb having this type of construction is given below:

us-le mera rahasja ukel-jo
he-ERG my secret.ACC reveal-3.SG.NPST
‘He revealed all my secrets’

The example above is for the verb ukel%VT%S2%FID2 and has ‘k1+k2+verb’ frame. It has following attributes:

arc-label	necessity	vibhakti	lextype	ontology
k1	m	le/0	p/n	[+ani/+hum]
k2	m	0	n	[+top]

It is also observed that several verbs share same semantic class and same karaka relation. Those verbs with their verb frames are given below:

A. Verb of Removing

i. ukhel%VT%S%FID1

arc-label	necessity	vibhakti	lextype	ontology
k1	m	le/0	p/n	[+ani/+hum]
k2	m	0	n	[+con]

ii. mur%VT%S2%FID2

arc-label	necessity	vibhakti	lextype	ontology
k1	m	le/0	n/p	[+ani/+hum]
k2	m	0	n	[+bdp]

B. Verb of Obtaining

i. harau%VT%S2%FID2

arc-label	necessity	vibhakti	lextype	ontology
k1	m	le/0	n	[+ani/+hum]
k2	m	lai	n	[+ani/+hum]

ii. sAmɑu%VT%S3%FID3

arc-label	necessity	vibhakti	lextype	ontology
k1	m	le/0	n	[+ani/+hum]
k2	m	0	n	[+ani/+hum]

The verbs above not only share the same semantic classes and dependency relations, but most of the arguments of those verbs share the same ontological features as well. Thus, we can assume that verbs with the same semantic class behave similarly, resulting in similar dependency structures.

The second class of verb-frame has ‘k1 + verb’ structure which is 103 in number. An example of such kind of verb is given below:

q̣hokɑ ughri-j̣o
 Door.NOM open-3.SG.PST
 'The door opened'

ughri%VI%S1%FID1

arc-label	necessity	vibhakti	lextype	ontology
k1	m	0	n	[+con]

The example given above is for the verb ughri%VI%S1, which is intransitive and has one place argument; q̣hokɑ ‘door’ is k1, which is in an agentive role but has zero *vibhakti*. It means the karta here has a nominative case. It shows that the verb having ‘k1+verb’ construction has zero *vibhakti*. The following verb frames having the same construction support the same fact:

i) **Verb:** uɾ%VI%S1%FID1

Example: cɑro uɾ-j̣o
 Bird-NOM fly-3.SG.PST
 'The bird flew away'

Verb Frame:

arc-label	necessity	vibhakti	lextype	ontology
k1	m	0	n	[+anm]

ii) **Verb:** udau%VI%S1%FID1

Example : gham uda-jo
 sun.NOM rise-3.SG.PST
 'The sun rose'

Verb Frame:

arc-label	necessity	vibhakti	lextype	ontology
k1	m	0	n	[+cels]

However, the verbs given below share same semantic classes and same argument structure under this category:

A. Verb of Appearance

i. ughri%VI%S1%FID1

arc-label	necessity	vibhakt	lextype	ontology
k1	m	0	n/p	[+con]

ii. udau%VI%S1%FID1

arc-label	necessity	vibhakti	lextype	ontology
k1	m	0	n	[+cels]

B. Verb of Psychological State

i. sanki%VI%S2%FID2

arc-label	necessity	vibhakti	lextype	ontology
k1	m	0	n	[+ani/+hum]

ii. gal%VI%S1%FID1

arc-label	necessity	vibhakti	lextype	ontology
k1	m	0	n	[+ani/+hum]

This way, we can conclude that the classification of verbs based on their frames is useful for understanding their syntactico-semantic behavior.

CHAPTER 5

VERB SENSE DISAMBIGUATION SYSTEM

5.1 Introduction

This chapter focuses on the implications of the verb frames database as a linguistic resource for verb sense disambiguation in Nepali. Section 5.2 introduces the theoretical concept of ambiguity in specific fields of linguistics and literature, while Section 5.3 explores the various types of ambiguity as described in previous literature. Section 5.4 delves into the theoretical concept of word sense disambiguation as a sub-task of Natural Language Processing (NLP), highlighting its approaches and implications. Finally, Section 5.5 outlines the system architecture of verb sense disambiguation using verb frames.

5.2 Nature and Scope of Ambiguity

Language is a human construct, an artifact that evolved in a social setting. It is a medium through which humans exchange their ideas and experiences. Under its formal attributes, language can be defined as a system of arbitrary verbal symbols structured systematically. One of the fundamental facts about a language is that it is loaded with ambiguous expressions, be it in a word, phrase, clause, sentence, or discourse level.

Ambiguity, in a literal sense, is an expression having more than one plausible sense of meaning when used in various contexts. It is “the state in which a word, statement, or any linguistic entity can be understood in more than one way” (Dixit, 2010, p. 51). Wasow et al. (2005) put that “meanings are regions in space, remaining agnostic as to its dimensionality, the metaphysical status of the points in it, and exactly how linguistic expressions get associated with regions”. They have used the term “denotation” to associate expressions and regions. They further state that “an expression is ambiguous if it has two or more distinct denotations - that is if it is associated with more than one region of the meaning space”. The word 'bank' in English is a common example of such ambiguity, which is interpreted as 'an edge of the river' or 'a financial institution' depending upon the context in which it is used.

Every concept is perceived differently in different areas of investigation, and it is very true to the concept of ambiguity as well. Researchers in literature and linguistics have different understandings of the concept of ambiguity. Literary texts are the storehouse of linguistic data, providing every possible attribute of language usage. So is the case with data on ambiguity as well. There is a long tradition of study of ambiguity in literature. Empson (1930) is a phenomenal work that has paved the way for research on ambiguity in literature in the West.

In Eastern literary philosophy, ambiguity has been frequently discussed since Acharya Bharatmuni's era under the *alankara* and *shabda shakti* theory. For example, *Alesha alankara*, *lakshana*, *vyanjanaa shabda shakti*, etc., resemble the concept of ambiguity in various ways.

However, in literary criticism, 'ambiguity' is considered an obligatory element of a literary text, whether in the East or the West. It is considered a literary device that helps create suspense, humor, satire, metaphor, or style of literary texts.

However, in linguistics, ambiguity is discussed and studied in Semantics, Psycholinguistics, Translation Studies, and Computational Linguistics. In Semantics, it is considered as 'meaningful properties of sense relations that hold between two or more expressions' (Allan, 1986, p. 140). While on the other hand, the concern of Psycholinguistics is more with the processes involved in the human brain while determining the meaning of an ambiguous expression. On the other hand, ambiguity is treated as one of the greatest hindrances in Translation Studies and Computational Linguistics. Their main concern is to find the resolution of such ambiguity.

5.3 Types of Ambiguity

Several classifications of ambiguity are found in various literatures of different disciplines. Since this study's main goal is to discuss and analyze ambiguity from a linguistic perspective, focusing more on the classification of ambiguity found in linguistics is legitimate. There are broadly four types of ambiguity found in any natural language. They are Lexical, Syntactic, Semantic, and Pragmatic ambiguity. The subsections below discuss the types of ambiguity in detail.

5.3.1 Lexical Ambiguity

The lexical ambiguity occurs when a lexical item or a word possesses more than one interpretation of meaning equally feasible in different contexts. English words like 'bank,' 'head,' 'chair,' 'book,' etc., are obvious instances of this ambiguity. Similarly, the words like *cāṅga*, *khera*, *ḍun*, *cola*, etc., in Nepali have two completely different meanings equally plausible in different contexts, as in the sentences below:

Sense 1: *cāṅga* (n.) - 'Kite'

makar sankranti-ma cāṅga ud-a-in-cha

makar sankranti-LOC kite.ACC fly-PASS-NPST

'Kite will be flown during makar sankranti'

Sense 2: *cāṅga* (n.) - 'Fit'

vajam gare-ra sarir-lai cāṅga rakh-na sak-in-cha

Exercise do-CONJ body-DAT fit keep-INF do-PASS-NPST

'One can keep the body fit by doing an exercise'

In the instances above, we see the word *cāṅga* has two different senses: 'kite' and 'fit.' The word can be taken as an example of lexical ambiguity in Nepali.

Similarly, the word *tira* is another example of lexical ambiguity in Nepali. It has two meanings depending upon the context: 'towards' and 'bank of the river.' The sentences given below exemplify the case:

Sense 1: *tira* - 'Towards'

ti keṭi iskol-bata soḍhai ghar-tira lag-i

that girl.NOM school-ABL direct home-towards go-3.SG.F.PST

'That girl went to her home directly after school'

Sense 2: *tirΛ* - 'Bank of the river

khola-ko tirΛ-ma madzhi bas-chΛ

River-GEN **bank**-LOC fisherman.NOM live-3.SG.M.NPST

'The fisherman lives at the bank of the river'

Likewise, the words like *kherΛ*, *dʒun*, and *cola* have two senses in two different contexts in Nepali. *kherΛ* is 'waste' as well as a 'moment,' *dʒun* is 'moon' and 'interrogative pronoun,' and *cola* is interpreted sometimes as 'blouse' and sometimes as 'life.' In this way, lexical ambiguity occurs when a lexical entity shows two completely different senses when used in two different contexts.

5.3.2 Syntactic Ambiguity

Syntactic ambiguity arises when a sentence is interpreted in more than one way, which is equally viable. Leech (1987), in this context, correctly puts that “ambiguity is a one-to-many relation between syntax and sense”. This type of ambiguity occurs when a phrase or sentence has multiple underlying structures. It arises because of” the relationship between the words and clauses of a sentence and the sentence structure underlying the word order they are in”. In another way, a sentence is considered syntactically ambiguous when a hearer or a reader perceives a sentence as having two different underlying structures and interpreted in two different senses. In English sentences like 'flying planes can be dangerous,' the word 'flying' can be interpreted as an adjective or a verb, resulting in two different understandings of its meaning. “The same thing is with the sentence ‘visiting relatives can be boring’ (Palmer, 1977, p. 108), which means that either 'relatives can be boring' or 'visits can be boring.' Likewise, the sentence given below in Nepali can be perceived as having two different structures, which results in two different senses as follows:

Sense 1:

maile bata-ma khel-da keṭa-haru-lai dekh-ẽ

I-ERG street-LOC play-CONJ boy-PL-ACC see-1.SG.PST

'I saw boys when I was playing on the street'

Sense 2:

mai-le bata-ma khel-da keṭa-haru-lai dekh-ẽ

I-ERG street-LOC play-MOD boy-PL-ACC see-1.SG.PST

'I saw the playing boys on the street.'

In the instances above, the word *khelda* is realized as two different grammatical elements in Nepali, i.e., a verb in the 'sense 1' sentence and a modifier in the 'sense 2' sentences, respectively. The realization of the same linguistic element, i.e., *khelda* in two different grammatical categories has resulted in an ambiguous interpretation of meaning. Hence, it is an instance that can be considered a case of structural ambiguity in Nepali.

5.3.3 Semantic Ambiguity

Human language is an output of an arbitrary relationship system between verbal signs and their referent, popularly known as the 'signifier' and 'signified' relation in Saussure's (1914) terminology. We often fail to describe the exact relationship between a language's signifier and signified elements. In this context, semantic ambiguity results from the relationship between an utterance and the object it entails. An expression with two or more plausible meanings in two different contexts is considered a case of semantic ambiguity. It incorporates the lexical and syntactic level ambiguities. It is due to 'the polyvalency (Polysemy and Homonymy) or matters related to referents.' For example, in a sentence like 'She cannot bear the children,' one has to determine whether the sentence means 'she cannot endure children' or 'she cannot go with children' depending upon the context.

5.3.4 Pragmatic Ambiguity

Pragmatic ambiguity, conversely, occurs at the contextual level and depends totally upon the hearer's perception of an expression. It is considered a non-linguistic ambiguity and often overlaps with semantic ambiguity. It is also known as situational ambiguity since the meaning of an expression greatly depends on the situation where a hearer stands. This kind of ambiguity is often used as a device of style in a literary text. An instance of such ambiguity in English is “John wants to marry a girl with green eyes” (Lyons, 1977, pp. 190-1). The sentence has two readings- “there is a particular girl with green eyes that John wants to marry” and “he has no specific girl in mind but merely requires that his future wife has green eyes”.

5.4 Word Sense Disambiguation and its Approaches

As discussed in section 5.2, a natural language is filled with vast occurrences of ambiguous expressions at lexical, syntactic, semantic, or pragmatic levels. Human brain is the storehouse of real-world knowledge, with the best equipped with mechanism to handle and resolve any ambiguity within a fraction of a second. However, the same task proved to be very difficult in the case of automatic language processing in a computer. With limited knowledge about the vast array of linguistic and non-linguistic behavior of the human community, computers often fail to determine the accurate sense of an ambiguous expression. Hence, Word Sense Disambiguation (WSD) is a computational task that tries to pick up an appropriate meaning of a word having multiple meanings. In other words, word sense disambiguation is the automated process of identifying the precise sense or meaning of a word when it has multiple possible interpretations. WSD is closely related to lexical semantics, whose main concern is to understand the relationship between 'word,' 'meaning,' and 'context.' In WSD, “meanings are called senses or concepts obtained from the sense inventory. The ambiguous word is referred to as the target word. The context in which the target word is used is called an instance” (Yesuf, 2015, p.3). It serves as a catalyst for various NLP tasks and applications, including machine translation, semantic interpretation, information retrieval, question answering, text mining, and parsing, rather than being a standalone goal.

Yesuf (2015) assumes that “in order to associate an occurrence of an ambiguous word with one of its senses, the WSD must have the following components”:

- i) “an inventory of the senses associated with each word to be disambiguated must be available”;
- ii) “a mechanism to associate word senses in context to individual senses must be developed”;
- iii) “an evaluation procedure must be adopted to measure how well this disambiguation mechanism performs”.

WSD considers ambiguity a natural classification problem, and it attempts to classify an occurrence of the word in a context into one or more of its sense classes with the help of knowledge bases such as a dictionary, thesaurus, WordNet, etc. Classification is done by using the context of the target word. ‘Determine the possible senses of each word and assigning the most appropriate sense to a word with high accuracy are two main tasks of WSD’ (Rajini, 2019: 7). There are primarily two approaches to Word Sense Disambiguation (WSD): the Knowledge-Based Approach and the Corpus-Based Approach.

i) **Knowledge-Based Approach** depends on the utilization of lexical resources such as machine-readable dictionaries, thesauri, WordNet, ontologies, and so on. These knowledge bases consist of pre-existing information about entities and concepts, usually created manually. This approach utilizes resources like machine-readable dictionaries, thesauri, and computational lexicons as part of its knowledge bases. Techniques employed in the Knowledge-Based Approach include semantic similarity assessment, selection restriction analysis, heuristic-based WSD, and overlap-based methods.

ii) **Corpus-based approach**, on the other hand, uses a large-scale corpus to extract contextual information about an ambiguous word which helps machines to learn disambiguation rules. ‘This approach applies statistical and machine learning methods for extracting the knowledge for disambiguation’ (Rajini, 2019: 8). This approach uses three techniques to extract information: Supervised, Semi-Supervised, and Unsupervised.

- a) **The supervised technique** relies upon manually annotated training data to identify the proper sense of an ambiguous word. “The instances in the

training data are annotated manually with their appropriate concepts from a sense inventory” (Yesuf, 2015, p. 24). Its concern is making disambiguation rules based on machine learning of certain “features associated with a particular word sense in the training data”. Decision List, Decision Trees, Naive Bayes, Genetic Algorithms, Neural Networks, Convolution Neural Networks, Exemplar/Memory Based Learning, Support Vector Machines, Ensemble Methods, Majority Voting, and AdaBoost are some of the techniques used for supervised methods (Rajini, 2019, pp. 15-17).

- b) **The semi-supervised technique** is a hybrid technique that makes some linguistic rules to carry out disambiguation tasks. Bootstrapping and Monosemous Relatives are two main types of semi-supervised techniques.
- c) **The unsupervised technique** is developed due to the limitations of supervised and semi-supervised techniques since they depend upon the availability of large amounts of annotated data which is time and effort-consuming. Hence, the unsupervised technique uses raw data in order to overcome difficulties faced by earlier techniques. This technique classifies words into clusters based on similar features of those words. “Unsupervised approaches map one of the labels to senses, rather than finding the actual sense of a word” (Rajini, 2019, p.21). Context Clustering Algorithms, Word Clustering Approaches, and Co-occurrence Graphs are some examples of unsupervised techniques.

5.5 Verb Sense Disambiguation System

Given that verbs have a limited number compared to other word categories, they tend to be more ambiguous in language. Each verb possesses a specific argument structure and exhibits distinct behavior during various alternation processes. Moreover, a verb's meaning can change depending on the alternations it undergoes, resulting in different morphological forms. Therefore, verbs are the most complex in terms of both their structure and meaning in a language.

In this context, Verb Sense Disambiguation (VSD) is a computational task that automatically selects the appropriate sense of a verb within a given context. Context

plays a pivotal role in determining the accurate sense of a word during sense disambiguation, and in the case of verbs, corresponding nouns and other elements provide this context. Additionally, the argument structure of a verb furnishes valuable contextual information. To elaborate, “the context and contextual information are defined in relation to the target word and the syntactic and semantic properties of the verb” (Murthy, 2003, p.83).

In another approach, corresponding nouns and other elements within a sentence need to be identified and annotated with their semantic features to discern the specific sense of the verb in the context.

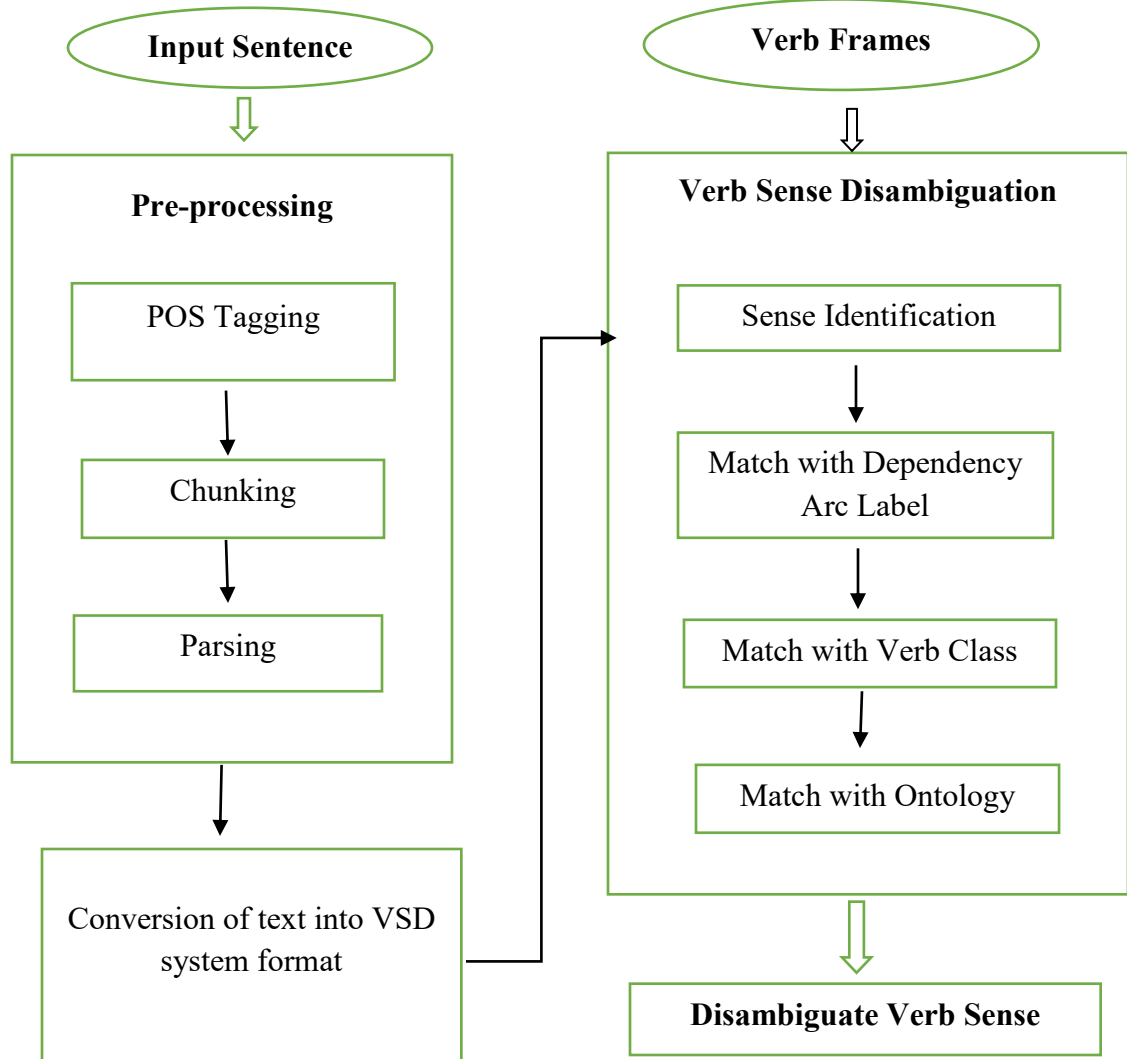
Chapter 4 of this thesis delves into the idea that our verb frame captures both lexico-syntactic and semantic information about a verb. Specifically, it offers insights into the semantic attributes of each argument, which is crucial for the sense identification task. The subsequent sections outline the design of a knowledge-based VSD system that we have constructed using verb frames as a foundational knowledge resource.

5.5.1 System Architecture

The architecture of the Nepali VSD system using verb frames has three main components: *pre-processing*, *conversion of text into VSD system format*, and *verb sense disambiguation*. Our system receives sentences having ambiguous verbs as input. In *pre-processing* stage, the corpus is annotated with three levels of linguistic knowledge: lexical (POS tagging), phrase (Chunking), and syntactic (Parsing).

The *conversion of text into VSD system format* takes parser output as an input and converts them into the VSD-compatible format. Finally, the *verb sense disambiguation* takes verb frames as input and identifies different senses of a verb then it undergoes the processes like *match with the karaka label*, *match with verb class*, and *match with ontology*. In this stage, the system retrieves *sense ID* to assign a proper meaning to a particular verb in the context. The flowchart in Figure 5.1 demonstrates a step-by-step architecture of the VSD system using verb frames.

Figure 5.1: System Architecture



5.5.2 Input Sentence

The input sentence is the term used in NLP to refer to the textual data or sentences used for analysis, processing, and other NLP tasks. The input sentences vary in form, length, and content, which can be extracted from different domains of language use like literature, tourism, entertainment, health, sports, science and technology, and so on. Input sentences in our system are extracted from a monolingual Nepali corpus of 200K sentences. A total of 1000 sentences having ambiguous sense relations have been taken as input here. Only simple sentences with simple verb forms are considered here to avoid the complex nature of subordinate clauses and compound verb forms.

5.5.3 Pre-Processing

The processing of the text refers to the automation, creation, and manipulation of electronic text. It involves different techniques and algorithms to understand, interpret and extract meaningful information from the text. The first component of our VSD system is pre-processing of the text. It takes sentences as input and makes them suitable for further processing. This stage is, in fact, the data preparation stage that occurs before the processing of text for other system components. Our system performs three tasks in the pre-processing phase: POS tagging, Chunking, and Parsing.

The pre-processing of the text for our system has been carried out using a Nepali parser developed by Pradhan et al. (2020). The parser is developed using the Conceptual Graph Approach and Paninian Grammatical Framework. The parser takes sentences as input and provides POS, Chunk, and Parsing level information about a sentence. It also provides information about sentence types such as affirmative, declarative, interrogative, non-interrogative, etc., and tense information about a verb, like past and non-past tense, which are irrelevant for our system. The subsections below elaborate on each task in detail. Figure 5.2 presents a sketch of the parser output, including all the information provided by it.

Figure 5.2: Parser Output

```
*****
ऊ
PRP
घर
NN
आयो
VF
|
Subject is Affirmative Simple Past Tense
Subject is Third person
1. Sentence is Non Interrogative.
1. Sentence is affirmative.
5. Subject is third person.
7. Sentence is in Simple Past Tense.
  Noun chunks [0] = ऊ
  Noun chunks [1] = घर
  आयो is the verb phrase
Part of Speech of ऊ is PRP
Relation of ऊ with verb is Karta (K1)
Part of Speech of आयो is VF
Part of Speech of घर is NN
Next Relation of घर with verb is Location (K2p)
Part of Speech of आयो is VF
*****
```

5.5.3.1 Parts of Speech Tagging

Part-of-speech tagging (POS tagging) is a fundamental NLP task involving the assignment of specific parts of speech to individual words within a given text. According to Georgiev et al. (2012), it can be described as the process of "assigning each word in a given piece of text a contextually suitable grammatical category." This task serves as a foundational and indispensable tool in natural language processing, constituting the most fundamental level of syntactic analysis.

The parts of speech or grammatical categories in a language are divided into certain types like noun, pronoun, verb, adjective, adverb, conjunction, preposition or postposition, and so on. The number and types of parts of speech differ from language to language.

The parts of speech tagging for our system is done using the Bureau of Indian Standard (BIS) tag set (2010) which is a hierarchical tag set having “a total of 38 annotation level tags and are common to all the Indian languages” (BIS Tag Set, 2010:1). Table 5.1 presents each tag with the convention used by BIS.

Table 5.1: BIS Tag Set for Indian Languages (2010)

S. No.	Category			Label	Convention
	Top Level	Subtype (Level 1)	Subtype Level 2		
1	Noun			N	N
1.1		Common		NN	N NN
1.2		Proper		NNP	N NNP
1.3		Verbal		NNV	N NNV
1.4		NLoc		NST	N NST
2	Pronoun			PR	PR
		Personal		PRP	PR PRP
		Reflexive		PRF	PR PRF
		Relative		PRL	PR PRL
		Reciprocal		PRC	PR PRC
		Wh-word		PRQ	PR PRQ
		Indefinite		PRI	PR PRI
3	Demonstrative			DM	DM
		Deictic		DMD	DM DMD
		Relative		DMR	DM DMR
		Wh-word		DMQ	DM DMQ
		Indefinite		DMI	DM DMI

4	Verb			V	V
4.1		Main		VM	V VM
			Finite	VF	V VM VF
			Non-finite	VNF	V VM VNF
			Infinitive	VINF	V VM VINF
			Gerund	VNG	V VM VNG
4.2		Auxiliary		VAUX	V VAUX
5	Adjective			JJ	
6	Adverb			RB	
7	Postposition			PSP	
8	Conjunction			CC	CC
8.1		Coordinator		CCD	CC CCD
8.2		Subordinator		CCS	CC CCS
8.3		Quotative		UT	CC CCS UT
9	Particles			RP	RP
9.1		Default		RPD	RP RPD
9.2		Classifier		CL	RP CL
9.3		Interjection		INJ	RP INJ
9.4		Intensifier		INTF	RP INTF
9.5		Negation		NEG	RP NEG
10	Quantifiers			QT	QT
10.1		General		QTF	QT QTF
10.2		Cardinals		QTC	QT QTC
10.3		Ordinals		QTO	QT QTO
11	Residuals			RD	RD
11.1		Foreign word		RDF	RD RDF
11.2		Symbol		SYM	RD SYM
11.3		Punctuation		PUNC	RD PUNC
11.4		Unknown		UNK	RD UNK
11.5		Echowords		ECH	RD ECH

5.5.3.2 Chunking

Chunking is the process of grouping words within a sentence that are syntactically related into chunks based on their grammatical structure and relationships. This task represents the second level of syntactic analysis following part-of-speech tagging. Abney (1991) provides a definition of a 'chunk' as "a single content word surrounded by a constellation of function words." In simpler terms, a chunk can be understood as "a minimal (non-recursive) phrase or partial structure comprised of closely related and inseparable words or entities, with no disruption to the dependencies within the chunk"

(Bharati et al., p. 2006). There are certain types of chunks based on the head of a phrase or chunk, like noun chunk, verb chunk, adjective chunk, adverb chunk, etc. The parser has followed the tag set prepared by Bharati et al. (2006) for the task of chunking. Table 5.2 below represents the chunk tag set prepared by Bharati et al. (2006) to consistently annotate Indian languages.

Table 5.2: Chunk Tag Set for Indian Languages

S. No.	Chunk Type	Tag Name
1	Noun Chunk	NP
2.1	Finite Verb Chunk	VGF
2.2	Non-finite Verb Chunk	VGNF
2.3	Infinitival Verb Chunk	VGINF
2.4	Verb Chunk (Gerund)	VGNN
3	Adjectival Chunk	JJP
4	Adverb Chunk	RBP
5	Chunk for Negatives	NEGP
6	Conjuncts	CCP
7	Chunk Fragments	FRAGP
8	Miscellaneous	BLK

5.5.3.3 Parsing

Analyzing the grammatical structure of a sentence or a piece of text is known as parsing. It helps to identify each constituent part of a sentence and its grammatical roles. It also provides insight into how those constituents are related to each other and how they function together to make a sentence meaningful. There are mainly two types of parsing:

- i. Dependency Parsing; and,
- ii. Constituency Parsing.

In our system, we have used the dependency approach for parsing, and in dependency parsing, the relationship between words is perceived as dependencies or directed links. Dependencies are represented through a hierarchical structure called a dependency tree or a syntactic parse tree. The dependency parsing assigns each word in a sentence with a specific role or grammatical function such as subject, object, modifier, etc. The roles are represented through arc labels where the head of the arc is considered the governing word, and the dependent is the word governed. The two most common

grammatical frameworks for dependency parsing are Paninian Grammar and Universal Dependency Grammar. Among these two, our verb frames represent the argument structure of a verb based on the Paninian Grammatical Framework that treats a verb as the head of the arc and other arguments as their dependencies.

5.5.4 Conversion of Text into VSD System Format

In Figure 5.2, we observed that the parser developed by Pradhan et al. (2020) has a specific output format which is incompatible with our system. It contains irrelevant information such as sentence type, TAM information, etc. To address this, we have transformed the output into another format and cleaned irrelevant information for our purpose. Additionally, due to the ongoing development of the parser, sentence annotation at each level needs to be fully refined. To rectify this, we have manually made corrections during this stage. These corrections include editing information about POS, Chunk, and Dependency Relation. Notably, each verb has been annotated with its root form in IPA and its corresponding semantic class. Figure 5.3 represents the converted text format as per our system requirement.

Figure 5.3: Conversion of Text Format

```
Sent ID: 1
Sentence 1 : ऊ घर आयो |
POS Tags 1 : { ऊ , PRP } { घर , NN } { आयो , VF } { | , PUNC }
Chunks: ( ऊ )( घर )
DREL: ( { ऊ , PRP } )_k1 ( { घर , NN } )_k2p { आयो , VF ( au : root >
verb of motion ) } { | , PUNC }

Sent ID: 2
Sentence 2 : उसलाई धेरै कुरा आउँछ |
POS Tags 2 : { उस , PRP } { लाई , PSP } { धेरै , JJ } { कुरा , NN } { आउँछ , VF
} { | , PUNC }
Chunks: ( उसलाई )( धेरै कुरा )
DREL: ( { उस PRP } { लाई , PSP } )_k4a ( { धेरै , JJ } { कुरा , NN } )_k2
{ आउँछ , VF ( au : root > verb of perception ) } { | , PUNC }
```

5.5.5 Verb Frames Database

We have developed a comprehensive database comprising verb frames for 200 ambiguous Nepali verbs. Within our database, there is a total of 482 frames available, each associated with one of these 200 ambiguous verbs. The verb frames included in our study provide two essential pieces of linguistic information about each verb: a Description and a Frame.

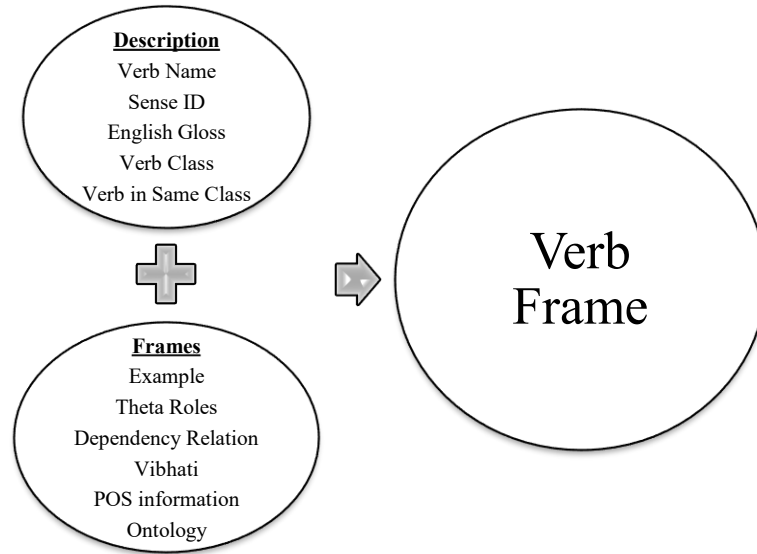
The Description of each verb encompasses key details such as the verb's name, sense ID, verb sense, English gloss, verb class, and other verbs within the same class, offering valuable context and classification.

On the other hand, the Frame section of our database consists of example sentences for each sense of a verb. These examples are designed to illustrate the verb's usage in various contexts and provide a detailed breakdown of its argument structure. This includes a specific focus on theta roles, dependency relations, vibhakti (a grammatical case system in Sanskrit and related languages), lexical type, and the ontology of arguments.

For more extensive information and a deeper understanding of our verb frame database, you can refer to Chapter 4 of this thesis, where we offer a detailed description and analysis of our database's content and structure.

The verb frames in our study are created to be used as resource for verb disambiguation tasks. The information furnished by our verb frames, such as the semantic class of the verb, dependency relation and ontology, is useful for identifying an appropriate sense of a particular verb. Figure 5.2 represents the overall schema of components of our verb frames.

Figure 5.4 Schema of Nepali Verb Frame



Our VSD system takes verb frames as a source of linguistic knowledge for each sense of a particular ambiguous verb and supplies disambiguation cues to the system at different levels.

5.5.6 Verb Sense Disambiguation

The verb sense disambiguation task in our system goes through four stages of data processing: Sense Identification, Match with Karaka label, Match with Verb Class, and Match with Ontology. The following subsections describe them in detail.

5.5.6.1 Sense Identification

The process of determining the most suitable sense of a word within a particular context is referred to as sense identification. In the context of Verb Sense Disambiguation (VSD), sense identification employs various approaches, including supervised, unsupervised, and knowledge-based methods, among others, to achieve its goal. Among these, we have used a knowledge-based method since we have extracted senses of verbs from the database of verb frames. In our verb frame database, each sense of a specific verb is associated with a unique sense ID, which comprises the verb's name, its type, and a sense number. Therefore, the process of identifying the sense of a verb corresponds to the identification of its corresponding sense ID, as these sense IDs are designed to uniquely represent and distinguish between different senses of the verb.

5.5.6.2 Match with Karaka Label

After the identifying the senses of the verb, our system undergoes the next stage called 'match with karaka label.' In this stage, information about the karaka or dependency relation of the arguments in a sentence is compared with the information about other senses of the same verb. While observing the system output, it is found that in many cases, the karaka labels for different verb senses are matched. If they are matched, they need to be refined through two additional filters: verb class and ontology. This is because when two senses of a verb have the same karaka relation, it indicates that they still need to be disambiguated using other cues. Figure 5.5 (a) and Figure 5.5 (b) below demonstrate the output for the verb 'ukel' having same karaka relation.

Figure 5.5 (a): System output for first sense of the verb 'ukel'

```
Sent ID: 10
Sentence 10 : उसले खाएका कुरा जम्मै उकेल्यो ।
POS Tags 10 : { उस , PRP } { ले , PSP } { खाएका , VF } { कुरा , NN } {
जम्मै , JJ } { उकेल्यो , VF } { । , PUNC }
Chunks: (उसले) (खाएका कुरा जम्मै)
DREL: (( { उस , PRP } { ले , PSP } )_k1 (( { खाएका , VF } { कुरा , NN } { जम्मै
, JJ } )_k2 { उकेल्यो , VF ( ukel : root > verb involving the body ) }
{ । , PUNC }
Matching sides only on krel basis: ukel%VT%S1, ukel%VT%S2
Matching sides on ontology basis: ukel%VT%S1
Matching sides on verb class basis: ukel%VT%S1
```

Figure 5.5 (b): System output for second sense of the verb ‘ukel’

```
Sent ID: 11
Sentence 11 : उसले मेरा रहस्य उकेल्यो ।
POS Tags 11 : { उस , PRP } { ले , PSP } { मेरा , PRP } { रहस्य , NN } {
उकेल्यो , VF } { । , PUNC }
Chunks: ( उसले ) ( मेरा रहस्य )
DREL: ({ उस , PRP } { ले , PSP })_k1 ({ मेरा , PRP } { रहस्य , NN })_k2 {
उकेल्यो , VF ( ukel : root > verb of communication ) } { । , PUNC }
Matching sides only on krel basis: ukel%VT%S1, ukel%VT%S2
Matching sides on ontology basis: ukel%VT%S2
Matching sides on verb class basis: ukel%VT%S2
```

As we can see in the provided figures, the karaka relations of both sentences, each having different sense for the verb ‘ukel’, are same i.e. k1+k2+verb. With this information, our system proceeded to the next two filters: verb class and ontology, which are distinct for both sentences. Consequently, the information about the verb class and ontology has provided the disambiguation cue for the verb.

5.5.6.3 Match with Ontology

Our verb frame provides additional semantic information about a verb, namely ontology. The ontology of a verb categorizes its various senses based on the roles they play in sentences. It also represents the semantic relationships they establish with other words in a sentence in a structured way. This type of structured representation can aid in disambiguating the different senses of a verb by offering a clear framework for understanding how the verb operates in different contexts. In the stage of match with ontology, our system compares the ontological information carried by the arguments of a particular verb. The sentences having different senses of a verb sometimes carry same ontology and sometimes different. In the case of same ontology, it again goes for another filter called verb class. Figure 5.6 demonstrates the system output for the verb ‘markau’ whose arguments are matched based on ontology.

Figure 5.6: System output for the two senses of the verb ‘markau’

```
Sent ID: 341
Sentence 341: उसले कम्मर मर्कियो ।
POS Tags 341 : { उस , PRP } { ले , PSP } { कम्मर , NN } { मर्कियो , VF }
{ | , PUNC }
Chunks: ( उसले ) ( कम्मर )
DREL: ({ उस , PRP } { ले , PSP })_k1 ({ कम्मर , NN })_k2 { मर्कियो , VF
( markau : root > Verb of motion ) } { | , PUNC }
Matching sides only on krel basis: markau%VT%S1, markau%VT%S2
Matching sides on ontology basis: markau%VT%S1, markau%VT%S2
Matching sides on verb class basis: markau%VT%S1

Sent ID: 342
Sentence 342: उसले हात मर्कियो ।
POS Tags 342 : { उस , PRP } { ले , PSP } { हात , NN } { मर्कियो , VF }
{ | , PUNC }
Chunks: ( उसले ) ( हात )
DREL: ({ उस , PRP } { ले , PSP })_k1 ({ हात , NN })_k2 { मर्कियो , VF
( markau : root > Verb of psychological state ) } { | , PUNC }
Matching sides only on krel basis: markau%VT%S1, markau%VT%S2
Matching sides on ontology basis: markau%VT%S1, markau%VT%S2
Matching sides on verb class basis: markau%VT%S2
```

As we observe in Figure 5.6, the two senses of the verb ‘markau’ have same argument structure based on karaka relation and ontology. The verb class is the only feature which provides disambiguation cue in this case.

5.5.6.4 Match with Verb Class

The verb class refers to the categorization of verbs based on their shared semantic properties. Verbs within the same semantic class often exhibit similar syntactic and semantic characteristics, and they behave almost in same way. In our verb frame, it represents the “semantic class of the verb based on Levin (1993)”, and it offers significant disambiguation cue for an ambiguous verb.

In this stage, our system compares the semantic classes of the arguments which a verb having more than one sense possessed. This is the third and final filter of our system which provides a significant clue for disambiguation task. Figure 5.7 demonstrates the system output for the two senses of the verb ‘ukhel’, which hold the same verb class, differing only on the basis of ontological feature.

Figure 5.7 System output for the two senses of the verb ‘ukhel’

```
Sent ID: 14
Sentence 14 : उसले घाँस उखेल्यो ।
POS Tags 14 : { उस , PRP } { ले , PSP } { घाँस , NN } { उखेल्यो , VF }
{ । , PUNC }
Chunks: ( उसले ) ( घाँस )
DREL: ({ उस , PRP { ले , PSP }})_k1 ({ घाँस , NN }}_k2 { उखेल्यो , VF
{ ukhel : root > verb of removing } { । , PUNC }
Matching sids only on krel basis: ukhel%VT%S2, ukhel%VT%S1
Matching sids on ontology basis: ukhel%VT%S1
Matching sids on verb class basis: ukhel%VT%S2, ukhel%VT%S1

Sent ID: 15
Sentence 15 : सरकारले अवैध घरहरू उखेल्यो ।
POS Tags 15 : { सरकार , NN } { ले , PSP } { अवैध , JJ } { घर , NN } {
हरू , PSP } { उखेल्यो , VF } { । , PUNC }
Chunks: ( सरकारले ) ( अवैध घरहरू )
DREL: ({ सरकार , NN } { ले , PSP }})_k1 ({ अवैध , JJ } { घर , NN } { हर्
PSP }}_k2 { उखेल्यो , VF ( ukhel : root > verb of removing ) } { । ,
PUNC }
Matching sids only on krel basis: ukhel%VT%S2, ukhel%VT%S1
Matching sids on ontology basis: ukhel%VT%S2
Matching sids on verb class basis: ukhel%VT%S2, ukhel%VT%S1
```

In the data represented in Figure 5.7, we can observe that the features like karaka relation and verb class are matched in both senses of the verb ‘ukhel’. Therefore, it is necessary for it to undergo through the filter of ontology for its disambiguation.

CHAPTER 6

EVALUATION AND ERROR ANALYSIS

6.1 Introduction

This chapter is dedicated to the evaluation and analysis of errors in the verb sense disambiguation (VSD) system that we have developed in this research. Section 6.2 provides an overview of the evaluation methodology, encompassing the techniques and the corpus employed to assess the system's performance. Section 6.3 goes on to delineate the errors that were observed in the system's output. Furthermore, this section conducts a thorough analysis of the types of errors that occurred and delves into the underlying reasons for these errors.

6.2 Evaluation

As evaluating the performance is a crucial aspect of this study, we have conducted a series of experiments to assess the effectiveness of the proposed approach for verb sense disambiguation. These experiments serve as a fundamental part of our research, allowing us to gauge how well the approach performs in disambiguating verb senses and providing valuable insights into its overall effectiveness. We have evaluated the proposed knowledge-based VSD system using the rule-based method. As discussed in the previous chapters, our system is developed using a knowledge-based approach, i.e., our system utilizes information provided by a verb frames database to assign senses to verbs. Our database contains 482 verb frames for 200 ambiguous Nepali verbs, (See Appendix I).

We have evaluated our system using a test dataset of 1,000 sentences containing ambiguous verbs. The sentences were sourced from ILCI, TDIL, Samakalinsahitya.com, and <https://github.com/sharad461/nepali-translator>, as demonstrated in Table 2.2.

The system's performance was evaluated using two measures: Precision and Recall. Precision refers to the number of verbs taken for testing, while Recall is the number of disambiguated verb senses or correctly retrieved verb frames. The performance evaluation criteria for our system can be formulated as follows:

$$\text{Precision} = \frac{\text{No. of Relevant retrieved verb frames}}{\text{No. of all retrieved verb frames}}$$

$$\text{Recall} = \frac{\text{No. of relevant retrieved verb frames}}{\text{No. of all relevant verb frames}}$$

Table 6.1 presents the precise test results, indicating that our system successfully retrieved 442 verb frames out of the 482 verb frames available in our database. With the guidance provided by these verb frames, our system was able to disambiguate 844 out of the 1000 test sentences. However, 156 sentences remained undistinguished due to various errors, as elaborated upon in Section 6.3 of this chapter.

Table 6.1: Test Result

Relevant Retrieved Verb Frames	Disambiguated Sentences	Non-Disambiguated Sentences
442	844	156
Total No. of Verb Frames = 482	Total Number of Sentences = 1000	

Hence, the precision and recall for our system are 91.7% and 84.4%, respectively as formulated below:

i) $\text{Precision} = 442/482 \times 100 = 91.7\%$

ii) $\text{Recall} = 844/1000 \times 100 = 84.4\%$

6.3 Error Analysis

Error Analysis involves identifying, categorizing, and understanding the types of errors that occur and their underlying causes or patterns of occurrence. It provides valuable insights into the strength and weakness of the system that would guide improvement and refinement. In our study, error analysis involves examining the cases where our VSD system failed to assign senses to verbs correctly and determining the reasons behind these failures. While evaluating the system, we found five types of errors that occurred in our system as follows:

- i. Preprocessing Error
- ii. Out of Verb Frame Database
- iii. Out of Ontological Lexicon
- iv. Out of Verb Class Lexicon
- v. Errors due to multiple matches in each level of information

Sub-sections below analyze each type of error with examples.

6.3.1 Pre-processing Error

The pre-processing error has occurred in our system during the conversion of text. In this stage, we have cleaned and edited parser output manually. This was necessary because the parser we employed in our research is still in its developmental stage which needs manual validation at each annotation step, such as POS tagging, Chunking and Parsing.

There are three types of preprocessing error found in our system, occurred due to three different reasons. They are exemplified in Figure 6.1 below. The first type of preprocessing error has occurred due to the incomplete information about the verb class like in the sentence ID 46, there is a missed phrase ‘Verb of’ in the information given. It is written as ‘oṭh: root > dressing’ instead of ‘oṭh: root > Verb of dressing’ due to which it has caused no matching information on verb class basis.

The second type of preprocessing error as given in Figure 6.1 is due to the misspelling of the verb root. For example in sentence ID 62, the verb root is written as ‘kaṭ’ instead of ‘kaṭh’.

Likewise, the third type of preprocessing error is occurred due to the wrong parenthesis information as in sentence ID 99, the bracket symbol ‘(’ is missing after ‘{मा , PSP }’.

These types of errors can be fixed with revised validation of the input sentences prior to the processing of the text for programming.

Figure 6.1: Preprocessing Error

Sent ID: 46
Sentence 46 : उसले टोपी ओढ्यो |
POS Tags 46 : { उस , PRP } { ले , PSP } { टोपी , NN } { ओढ्यो , VF } { | , PUNC }
Chunks: (उसले) (टोपी)
DREL: ({ उस , PRP } { ले , PSP })_k1 ({ टोपी , NN })_k2 { ओढ्यो , VF (oth : root > dressing) }
{ | , PUNC }
Matching sids only on krel basis: oth%VT%S1, oth%VT%S2
Matching sids on ontology basis: oth%VT%S1, oth%VT%S2
Matching sids on verb class basis:

Sent ID: 62
Sentence 62 : समाजले हरिलाई पानीमा काढ्यो |
POS Tags 62 : { समाज , NN } { ले , PSP } { हरि , NNP } { लाई , PSP } { पानी , NN } { मा , PSP }
{ काढ्यो , VF } { | , PUNC }
Chunks: (समाजले) (हरिलाई) (पानीमा)
DREL: ({ समाज , NN } { ले , PSP })_k1 ({ हरि , NNP } { लाई , PSP })_k2 ({ पानी , NN } { मा ,
PSP })_k7 { काढ्यो , VF (kar : root > Verb of removing) } { | , PUNC }
Matching sids only on krel basis:
Matching sids on ontology basis:
Matching sids on verb class basis:

Sent ID: 99
Sentence 99 : उसका मनमा कुरा खेल्यो |
POS Tags 99 : { उसका , PRP } { मन , NN } { मा , PSP } { कुरा , NN } { खेल्यो , VF } { | ,
PUNC }
Chunks: (उसका मनमा) (कुरा)
DREL: ({ उसका , NN } { मन , NN } { मा , PSP })_k7 ({ कुरा , NN })_k2 { खेल्यो , VF (khel :
root > Verb of psychological state) } { | , PUNC }
Matching sids only on krel basis:
Matching sids on ontology basis:
Matching sids on verb class basis:

6.3.2 Out of Verb Frame Database

Our VSD system depends on a database of verb frames to obtain information about how a verb behaves in a particular context or sense. We have a database of 482 verb frames corresponding to the 200 ambiguous Nepali verbs, with each frame associated with a particular sense of these verbs. As seen in the system architecture in Figure 5.1, our system goes through four stages of verb sense disambiguation: sense identification, match with karaka label, match with verb class, and match with ontology. These four stages rely entirely on the knowledge provided by the verb frame database. With the help of this database, the system first identifies the different senses of a verb, and then it matches the karaka, verb class, and ontological information of arguments that a particular sense of a verb takes with the information provided by the verb frame.

Throughout the evaluation process, it's important to acknowledge that errors may arise due to the absence of a verb frame for a particular sense of a verb. This highlights the significance of having a comprehensive and well-structured verb frame database to support accurate sense disambiguation. In cases where a specific sense lacks adequate representation in the database, the system may struggle to make the correct sense identification, resulting in errors. This underscores the importance of continually expanding and refining the database to enhance the overall performance of the verb sense disambiguation system. Figure 6.2 exemplifies such errors as follows:

Figure 6.2: Error due to lack of Verb Frame Database

Sent ID: 491

Sentence 491 : मलाई जाडो भयो।

POS Tags 491 : { म, PRP } { लाई, PSP } { जाडो, NN } { भयो, VF } { |, PUNC }

Chunks: (मलाई)(जाडो)

DREL: ({ म, PRP } { लाई, PSP })_k4a ({ जाडो, NN })_k2 { भयो, VF (hu : root > Verb of psychological state) } { |, PUNC }

Matching sids only on krel basis:

Matching sids on ontology basis:

Matching sids on verb class basis:

Sent ID: 487

Sentence 487 : उसले मारुनी गायो।

POS Tags 487 : { उस, PRP } { ले, PSP } { मारुनी, NN } { गायो, VF } { |, PUNC }

Chunks: (उसले)(मारुनी)

DREL: ({ उस, PRP } { ले, PSP })_k1 ({ मारुनी, NN })_k2 { गायो, VF (ga : root > Verb of Performance) } { |, PUNC }

Matching sids only on krel basis:

Matching sids on ontology basis:

Matching sids on verb class basis:

In sentences IDs 491 and 487, the verb roots are 'hu' and 'ga', which are not present in our verb frames database. In other words, our database does not contain any frames for these verbs. Therefore, our system is unable to retrieve 'match' information for these verbs, which has resulted in errors.

6.3.3 Out of Verb Class Lexicon

The semantic class of a verb provides an important clue for determining the correct sense of a verb in a particular context. Verb class information, based on Levin (1993), constitutes one of the main components of our verb frame. The VSD system we have developed incorporates this information in the disambiguation process through the stage of ‘match with verb class’. Out of the 51 classes of verbs provided by Levin (1993), we have utilized 38 classes in our verb frames. Given the relatively small size of our verb frame database, which encompasses only 482 verb frames, it's important to recognize that our lexicon of verb classes is also limited in number. Consequently, our Verb Sense Disambiguation (VSD) system may occasionally encounter errors stemming from the absence of verb class information. This limitation highlights the need for ongoing efforts to expand and enrich both the verb frame database and the lexicon of verb classes, as a more comprehensive resource can significantly enhance the accuracy and reliability of the VSD system. Figure 6.3 exemplifies such type of errors.

Figure 6.3: Error due to lack of Verb Class Information

```
Sent ID: 58
Sentence 58 : रामले भारतको सिमाना काट्यो |
POS Tags 58 : { राम , NNP } { ले , PSP } { भारतको , NNP } { सिमाना , NN } { काट्यो , VF } { | , PUNC }
Chunks: ( रामले ) ( भारतको सिमाना )
DREL: ({ राम , NNP } { ले , PSP })_k1 ({ भारतको , NNP } { सिमाना , NN })_k2 { काट्यो , VF (kaṭ : root) } { | , PUNC }
Matching sids only on krel basis: kaṭ%VT%S1,kaṭ%VT%S2,kaṭ%VT%S3,kaṭ%VT%S4,kaṭ%VT%S6
Matching sids on ontology basis:
Matching sids on verb class basis:

Sent ID: 67
Sentence 67 : कुखुराले चल्ला काट्यो |
POS Tags 67 : { कुखुरा , NN } { ले , PSP } { चल्ला , NN } { काट्यो , VF } { | , PUNC }
Chunks: ( कुखुराले ) ( चल्ला )
DREL: ({ कुखुरा , NN } { ले , PSP })_k1 ({ चल्ला , NN })_k2 { काट्यो , VF (kaṭh : root) } { | , PUNC }
Matching sids only on krel basis:kaṭh%VT%S2,kaṭh%VT%S3,kaṭh%VT%S4,kaṭh%VT%S5,kaṭh%VT%S6,kaṭh%VT%S7
Matching sids on ontology basis:
Matching sids on verb class basis:
```

In Figure 6.3 above, the sentence IDs 58 and 67 contain verb roots 'kaṭ' ('cross') and 'kaṭh' ('hatch'), respectively. These verbs lack verb class information in the input sentences because they are not present in our verb class lexicon entries.

6.3.4 Out of Ontological Lexicon

The ontological features of arguments provide another crucial clue for disambiguating verbs with the assistance of our verb frames. The ontology tag set, prepared by CALTS (2017), has been devised to depict ontological features of arguments within our verb frames, and we have created a small ontological dictionary for our VSD system. The 'match with ontology' stage in our system uses ontological information as a disambiguation cue. However, because our ontological dictionary contains a very limited number of words, it sometimes fails to cover all the words found in test sentences. As a result, it may encounter errors due to the lack of an ontological lexicon for some words.

Figure 6.4: Error due to lack of Ontological Information

```
Sent ID: 71
Sentence 71 : तिमीले धान कुट्यौ।
POS Tags 71 : { तिमी, PRP } { ले, PSP } { धान, NN } { कुट्यौ, VF } { |, PUNC }
Chunks: ( तिमीले ) ( धान )
DREL: (( { तिमी, PRL } { ले, PSP } ))_k1 (( { धान, NN } ))_k2 { कुट्यौ, VF ( kuṭ : root > Verb of
contactby impact ) } { |, PUNC }
Matching sids only on krel basis: kuṭ%VT%S1%FID1, kuṭ%VT%S2
Matching sids on ontology basis:
Matching sids on verb class basis: kuṭ%VT%S1%FID1, kuṭ%VT%S2

Sent ID: 78
Sentence 78 : हरिले आइफल खायो।
POS Tags 78 : { हरि, NNP } { ले, PSP } { आइफल, NN } { खायो, VF } { |, PUNC }
Chunks: ( हरिले ) ( आइफल )
DREL: (( { हरि, NNP } { ले, PSP } ))_k1 (( { आइफल, NN } ))_k2 { खायो, VF ( ( kha : root > Verb of
ingesting ) } { |, PUNC }
Matching sids only on krel basis: kha%VT%S2, kha%VT%S1, kha%VT%S10, kha%VT%S11,
kha%VT%S3, kha%VT%S5, kha%VT%S6, kha%VT%S7, kha%VT%S8, kha%VT%S9
Matching sids on ontology basis:
Matching sids on verb class basis: kha%VT%S2, kha%VT%S1, kha%VT%S3
```

Figure 6.4 above shows the two instances of errors that persisted due to the lack of ontological information for the words धान [dhan ‘paddy’] and आइफल [aiphall ‘apple’] as in sentence IDs 71 and 78, respectively.

6.3.5 Error due to multiple matches in each level of information

This type of error occurs when there is multiple sense IDs that match with each other for a particular verb given in the test input. The presence of multiple matches in the karaka relation, ontology, and verb class information indicates that our system is unable to disambiguate the verb based on the information provided by our verb frames database at these three levels. Figure 6.5 exemplifies such errors as follows:

Figure 6.5: Error due to multiple matches in each level of information

Sent ID: 420
Sentence 420 : मलाई मात लाग्यो |
POS Tags 420 : { म, PRP } { लाई, PSP } { मात, NN } { लाग्यो, VF } { |, PUNC }
Chunks: (मलाई) (मात)
DREL: (({ म, PRP } { लाई, PSP })_k4a ({ मात, NN })_k2 { लाग्यो, VF (lag : root > Verb of psychological state) } { |, PUNC }
Matching sids only on krel basis: lag%VT%S2, lag%VT%S10, lag%VT%S11, lag%VT%S5, lag%VT%S8
Matching sids on ontology basis: lag%VT%S2, lag%VT%S10
Matching sids on verb class basis: lag%VT%S2, lag%VT%S10

Sent ID: 56
Sentence 56 : भाइले मेरो कुरा काट्यो |
POS Tags 56 : { भाइ, NN } { ले, PSP } { मेरो, PRP } { कुरा, NN } { काट्यो, VF } { |, PUNC }
Chunks: (भाइले) (मेरो कुरा)
DREL: (({ भाइ, NN } { ले, PSP })_k1 ({ मेरो, PRP } { कुरा, NN })_k2 { काट्यो, VF (kat : root > Verb of communication) } { |, PUNC }
Matching sids only on krel basis: kat%VT%S1, kat%VT%S2, kat%VT%S3, kat%VT%S4, kat%VT%S6
Matching sids on ontology basis: kat%VT%S3, kat%VT%S4
Matching sids on verb class basis: kat%VT%S3, kat%VT%S4

In the figure above, the sentence IDs 420 and 56 have verbs 'lag' and 'kat', respectively, which have multiple matches at each level of disambiguation information, including karaka, ontology, and verb class. Errors of this nature require specific treatment to identify additional disambiguation cues.

CHAPTER 7

SUMMARY AND CONCLUSION

7.1 Summary

There is a dire need to build different kinds of linguistic resources in order to enhance language technologies for low-resourced languages like Nepali. The fact that natural language processing without linguistic knowledge is next to impossible encourages researchers to pursue linguistics computationally. We see tremendous efforts being carried out on personal and institutional levels to create linguistic knowledge bases for different languages worldwide. In the same scenario, the Nepali language is at the primary stage of natural language processing, which has various thrust areas where we can make efforts and do more research. In this context, this study is an effort to build a knowledge base of ambiguous Nepali verbs for enhancing NLP tasks and applications. As we see in Chapter 1 of this thesis, ambiguity in any natural language has proved to be a great barrier to getting desirable accuracy in every NLP application, such as Machine Translation, Semantic Interpretation, Information Retrieval and Extraction, etc. It is, in fact, an ‘AI-complete problem’ which seeks strategies to handle it in more specific ways. Verb Sense Disambiguation is an NLP task that resolves verbal ambiguities through different methods and approaches. The knowledge-based approach finds resolution strategies in language resources such as dictionaries, thesaurus, WordNet, etc. This study proposes disambiguating Nepali verbs using verb frames, a knowledge base enriched with lexical, syntactic, and semantic information about a particular verb.

Chapter 2 of this thesis has provided a comprehensive overview of the methodological aspects of the study, encompassing two main sections: Collection of Data and Theoretical Framework. In the first section, the study described the extensive process of gathering data to build verb frames for Nepali. A total of 482 verb frames were meticulously developed, covering each sense of 200 ambiguous verbs. The accuracy and reliability of these verb senses were confirmed through the meticulous examination of a corpus comprising 200,000 sentences and cross-referenced with four reputable Nepali dictionaries in print form. The chapter elucidated the sources and nature of the data collection, highlighting the diverse domains and rich linguistic resources employed for

this purpose. The second section of this chapter is about the theoretical framework delved into the foundational underpinnings of the research, divided into two sub-sections. The first sub-section, introduced the concept of a Verb Frame as a comprehensive knowledge repository that encapsulates lexico-syntactic and semantic knowledge about individual verbs. Drawing inspiration from the work of Begum (2017), the study adapted and modified the structure of verb frames to serve the specific purpose of sense disambiguation in Nepali verbs. The modifications included the incorporation of additional information such as 'verb class' and 'ontology' to enhance the accuracy of sense disambiguation. Notably, certain elements like 'position' and 'relation' were omitted from the verb frame, as they remained constant in Nepali sentences due to language-specific characteristics. The chapter provided a detailed breakdown of the components of the verb frame, distinguishing between the Description of the verb and the Verb Frame itself. Key components included information about the verb's name, sense ID, verb sense, English gloss, verb class, verbs in the same class, example sentences, theta roles, and frame ID. Additionally, it outlined the tabular representation of the argument structure of a verb, detailing elements like arc-label, necessity, vibhakti, lextype, and ontology. The second sub-section delved into the Knowledge-Based Approach for Verb Sense Disambiguation (VSD). The study highlighted the categorization of VSD approaches, including Supervised Machine Learning, Unsupervised, Semi-Supervised, and Knowledge-Based methods. In this research, a Knowledge-Based Approach was adopted, leveraging external linguistic resources such as dictionaries, thesauri, and WordNet. This approach relied on the comprehensive verb frames constructed earlier, which contained a wealth of linguistic knowledge encompassing verb senses and their associated attributes. These attributes, including karaka relations, verb class, and ontology, were identified as valuable assets for enhancing the accuracy of verb sense disambiguation.

Chapter 3 of this thesis presents a preliminary introduction to Nepali verb morphology. Morphologically, Nepali is an agglutinating language that exhibits SOV phrase order in its syntax. The Nepali language has a nominal and verbal system rich in subject-verb or subject-object-verb agreement patterns. Particularly, the Nepali verbs inflect for gender, number, person, honorific status, tense, aspect, and mood, and for their alternation processes like passivization, negativization, and causativization. The transitive

verb in Nepali takes up to three objects in a sentence. Thus, the transitive verbs in Nepali are mono-transitive, di-transitive, or tri-transitive. The process of causativization increases the valence of a verb, resulting in the formation of tri-transitive construction in Nepali. The formal structure of finite verbs in Nepali is ‘root+copula’ as they end in [chu/chaũ/chas/chau/cha/chan] or [ho/thiyo] copular suffixes. The non-finite verbs in Nepali have [dai/da/i/era] endings. Based on morphological structure, verbs in Nepali can be divided into simple and compound verbs. Simple verbs refer to those that are made out of only one root, whereas compound verbs are the result of a combination of more than one root and copula. The auxiliaries in Nepali are suffixed to the verb root in order to make a verb stem. The suffixes like [-cha,-jo], and [ho] are auxiliaries in Nepali that exhibit various morphological forms according to tense, aspect, mood, and other inflections. Nepali has two grammatical genders based on inflectional markers that subject and verb take in a sentence. They are feminine and non-feminine. The linguistic entity that denotes other than human does not show grammatical gender distinction in Nepali. The human masculine entities and non-human entities, including non-living things, fall under the non-feminine gender in this language. First, second, and third persons with no inclusive and exclusive distinctions in second person pronouns are found in Nepali. For the first-person singular and plural forms and second and third-person plural forms, the verb does not show any variations in cases of gender and honorific status. However, the Nepali verb shows variations for number and person for each unit. The auxiliary verb carries tense, aspect, and mood information in Nepali. The two-fold distinctions for tense are found in Nepali: past and non-past. The aspects in Nepali are simple, continuous, perfect, habitual, and unknown.

Among these aspects, habitual and unknown aspects are for past tense only. There are five types of grammatical moods in Nepali. They are: assertive, imperative, interrogative, indicative and presumptive. In order to make a sentence with passive voice, the *-ijo* morpheme is suffixed to a verb stem. There are five methods of causative formation in Nepali: by *-a* suffix, by *-a/al* suffix, by changing [ʌ] into [a], by changing [u] into [o], and by *-a* insertion. In Nepali, *-na* is a negative marker. It is either prefixed or suffixed to a verb stem to make a negative sentence.

Chapter 4 of the thesis discusses the procedures and methodology of building verb frames for Nepali. A total of 482 verb frames have been created for 200 ambiguous Nepali verbs. The verbs having two to fifteen senses have been taken for verb frames, and a corpus analysis has made the extraction and selection of verbs of 2 lac sentences. The verbs are selected based on their frequency of occurrence and their obvious ambiguity. The senses of each verb have been verified using IndoWordNet and some of the standard dictionaries of Nepali. Verb frames are created following the theoretical framework provided by Begum (2017). They comprise two parts of information: Description and Verb frame. The description refers to the lexico-semantic features of a particular verb with special reference to its name, sense, sense ID, gloss, and semantic class and verbs in the same class. Verb frame, on the other hand, refers to the specific table representing syntactic information such as example sentences, theta roles, dependency relations, and the necessity of the arguments, vibhaktis, lexical type, and ontology. The example sentences for each sense of a particular verb have been extracted from the corpus and transcribed in IPA using interlinear morpheme-by-morpheme glossing provided by Leipzig Glossing Rules (2015). The verbs in the example sentences are converted into simple past tense form to avoid the complexity of verbal inflections a verb can hold due to TAM and other inflections. Theta roles of the arguments are depicted using the tagging module provided by VerbNet (2006). At the same time, dependency relations are represented with the help of the Paninian Grammatical Framework. The necessity of the arguments is captured either as desirable or mandatory while the type of vibhaktis are given based on AnnCora Guidelines annotating Hindi Treebank (2014). Detailed information about the ontologies of arguments is also represented in verb frames, which are tagged based on the tagset provided by CALTS, University of Hyderabad (2017). The fourth chapter of this thesis also tries to capture unique verb frames for classifying Nepali verbs. The classification proposes 35 types of verbs having unique frames.

In chapter 5 of this thesis, we have delved into the intricate world of ambiguity in language and its various manifestations. From the fundamental concept of ambiguity to its different types, we have explored how language can be a rich source of multiple meanings. This exploration has underscored the need for effective tools and techniques to

disambiguate these ambiguous expressions, especially in the realm of natural language processing (NLP). Ambiguity, as we have seen, can manifest at various levels, including lexical, syntactic, semantic, and pragmatic. Each type of ambiguity presents unique challenges, and resolving them is crucial for enhancing the accuracy of NLP applications. Word Sense Disambiguation (WSD) emerges as a critical task in this context, as it aims to automatically determine the correct sense of a word in a given context. We have discussed the two main approaches to WSD: the knowledge-based approach and the corpus-based approach. The knowledge-based approach relies on predefined lexical resources and expert knowledge, while the corpus-based approach utilizes large-scale textual data and statistical methods. Both approaches have their strengths and limitations, and the choice between them depends on the availability of resources and specific application requirements. To illustrate the practical application of WSD, the chapter also presents the system architecture of Verb Sense Disambiguation using verb frames.

Verb Sense Disambiguation (VSD) is a computational task that aims to automatically select the appropriate sense of a verb within a given context. Contextual information, including corresponding nouns, and the argument structure of the verb, is crucial for accurate sense disambiguation. The Nepali VSD system using verb frames consists of three main components: pre-processing, conversion of text into VSD system format, and verb sense disambiguation. Each stage plays a vital role in preparing the data and applying linguistic knowledge to disambiguate verb senses. Input sentences for the system are extracted from a monolingual Nepali corpus and are carefully selected to avoid complex structures. Simple sentences with straightforward verb forms are chosen to facilitate the disambiguation process. Pre-processing involves three tasks: Part-of-Speech (POS) tagging, chunking, and parsing. These tasks help in analyzing the grammatical structure of sentences and preparing the data for subsequent stages. The output from the parser is transformed into a format compatible with the VSD system. Irrelevant information is removed, and corrections are made to ensure the data aligns with the system's requirements. A database of verb frames is created, containing information about 200 ambiguous Nepali verbs and their various senses. These frames provide valuable semantic and syntactic information for disambiguation. The core task of the system involves four stages: Sense Identification, Match with Karaka Label, Match with

Verb Class, and Match with Ontology. These stages employ different cues, such as sense IDs, karaka labels, verb classes, and ontological information, to disambiguate verb senses effectively. The final step involves retrieving sense IDs and matching them with corresponding verb senses, resulting in the selection of the appropriate sense for a given verb in context. Overall, this chapter has outlined the comprehensive process of disambiguating verb senses using verb frames in the Nepali language. It highlights the importance of linguistic knowledge and context in tackling the ambiguity inherent in verbs and demonstrates the systematic approach employed in the VSD system.

In chapter 6, we have delved into the evaluation and analysis of the verb sense disambiguation system developed in the course of our research. Our system, which relies on a knowledge-based approach utilizing a verb frames database, was subjected to rigorous evaluation using a test dataset of 1,000 sentences containing ambiguous verbs. The evaluation was carried out using precision and recall as performance metrics, yielding precision of 91.7% and recall of 84.4%. These figures demonstrate the system's effectiveness in disambiguating verb senses in a Nepali language context. However, our analysis also uncovered various types of errors, each with its own distinct characteristics. These errors include preprocessing errors, instances where the system is unable to find verb frames in its database, absence of verb class information, lack of ontological lexicon coverage, and errors arising from multiple matches in the disambiguation process. Each type of error was examined in detail, providing insights into areas where the system can be refined and improved. Overall, this chapter has provided a comprehensive overview of the performance and limitations of our verb sense disambiguation system, setting the stage for future enhancements and refinements to address these identified errors and further improve the system's accuracy and reliability.

7.2 Conclusion

In conclusion, this research highlights the significant challenges and opportunities in the field of natural language processing for low-resourced languages like Nepali. The study underscores the critical need to develop linguistic resources to enhance language technologies for such languages.

The research has shown that computational linguistics can play a pivotal role in addressing linguistic complexities and ambiguities. It has demonstrated the importance of linguistic knowledge in achieving accurate results in various natural language processing applications.

The creation of a comprehensive knowledge base of ambiguous Nepali verbs, consisting of 482 verb frames for 200 ambiguous verbs, is a significant contribution to the field. These verb frames, enriched with lexical, syntactic, and semantic information, serve as valuable assets for enhancing natural language processing tasks and applications.

Ambiguity, a prevalent challenge in natural language, has been identified as a major barrier to achieving desirable accuracy in NLP applications. The study recognizes ambiguity as an 'AI-complete problem' and focuses on strategies to handle it effectively.

The adoption of a knowledge-based approach for Verb Sense Disambiguation using verb frames has been demonstrated as a promising method. The incorporation of linguistic resources such as dictionaries, thesauri, and ontological information has proven to be effective in enhancing the accuracy of sense disambiguation.

Throughout the research, a thorough understanding of Nepali verb morphology and its intricacies has been presented. This knowledge is fundamental in developing accurate verb frames and, by extension, improving NLP applications for the Nepali language.

The evaluation of the VSD system has provided valuable insights into its strengths and weaknesses. The identified errors, ranging from preprocessing issues to database limitations and multiple matches, serve as guidance for future refinements and improvements.

In essence, this research represents a significant step towards advancing natural language processing capabilities for Nepali and similar low-resourced languages. It highlights the importance of linguistic resources, context, and systematic approaches in addressing linguistic complexities and achieving accurate results in NLP applications.

The journey continues, with opportunities for further research and enhancement in the field of computational linguistics for underrepresented languages.

References

- Aarts, Baas. (2001). *English Syntax and Argumentation* (2nd Edition). New York: PALGRAVE.
- Acharya, Jairaj. (1991). *A Descriptive Grammar of Nepali and An Analyzed Corpus*. Washington, D.C: Georgetown University Press.
- Adhikary, Rameshwar Prasad. (1980). ch- As a Non-Past Marker in Nepali. In *Nepalese Linguistics*, Vol.1. Kathmandu, Nepal: Linguistic Society of Nepal.
- Adhikari, H. (2052 BS). *Nepali Karaka Vyakarana*. Kathmandu, Nepal: Nepal Rajajiya Prajna Pratishthan.
- Bal, Bal Krishna. (2009). Towards Building Advanced Natural Language Applications - An Overview of the Existing Primary Resources and Applications in Nepali. Retrieved from <https://aclanthology.org/W09-3424.pdf>
- Baker, Collin F, Charles J. Fillmore and John Lowe. (1998). The Berkeley FrameNet Project. In *COLIN 98, Proceedings of the conference*. Retrieved from <https://aclanthology.org/P98-1013.pdf>.
- Basnyat, S. (1986). *Componential Analysis*. Unpublished PhD dissertation. Kathmandu, Nepal: Central Department of Linguistics, Tribhuvan University.
- Basnet, Santa B. and Shailesh B. Pandey. (2009). Morphological Analysis of Verbs in Nepali. In *Nepalese Linguistics*, Vol.24. Kathmandu, Nepal: Linguistic Society of Nepal.
- Bhatta, V.P. (1988-89). Theory of Karaka. Bulletin of the Deccan College Post Graduate and Research Institute, Vol.47/48. Retrieved from <https://www.jstor.org/stable/42930207>.
- Bharati, Akshar, Dipti Misra Sharma, Samar Husain, Lakshmi Bai, Rafiya Begam, Rajeev Sangal. (2012). AnnCorra: TreeBanks for Indian Languages Guidelines for Annotating Hindi TreeBank. Hyderabad: Language Technologies Research Center, IIIT, Hyderabad. Retrieved from <https://www.academia.edu/53949232>.
- Bhattacharyya, Pushpak. (2010). IndoWordNet. Retrieved from http://www.lrec-conf.org/proceedings/lrec2010/pdf/939_Paper.pdf.
- Begum, Rafiya. (2017). *Developing A Pilot Hindi Treebank Based on Computational Paninian Grammar*. An unpublished PhD Dissertation. Hyderabad: International Institute of Information Technology.
- Brandy E, King and Kathy Reinold. (2008). *Finding the concept, not just the word: A librarians guide to ontologies and semantics*. Oxford, England: Chandos Publishing.
- Chandra Sharma (Eds.). (2013). *Pragya Nepali-to-English Dictionary*. Kathmadu, Nepal:

- Nepal Academy.
- Chomsky, N. (2015). *Syntactic Structures*. Martino Publishing
- Dhungana, Udaya Raj, Subarna Shakya, Kabita Baral and Bharat Sharma. (2014). Word Sense Disambiguation using WSD specific WordNet of Polysemy Words. In *International Journal on Natural Language Computing*, Vol. 3, No.4. Retrieved from <http://airccse.org/journal/ijnlc/papers/3414ijnlc05.pdf>.
- Dhungana, Udaya Raj and Subarna Shakya. (2014). Word Sense Disambiguation using Clue Words. In *Journal of the Institute of Engineering*, Vol. 10, No. 1, pp. 192–198. Retrieved from <http://www.nepjol.info/index.php/JIE/issue/view/716>.
- Dhungana, Udaya Raj and Subarna Shakya. (2014). Word Sense Disambiguation in Nepali Language. In *Proceedings of Fourth International Conference on Digital Information and Communication Technology and its Applications*. Retrieved from <http://ieeexplore.ieee.org/document/6821655/>
- Gorthi, Sai Kiran, Ashish Palakurthi, Radhika Mamidi, Dipti Misra Sharma. (2014). Identification of Karaka relations in an English sentence. Retrieved from <http://cdn.iiit.ac.in/cdn/lttrc.iiit.ac.in/icon/2014/proceedings/File69-p147.pdf>.
- Ghosh, Sanjukta. (2015). Making Verb Frames for Bangla Vector Verbs. Retrieved from <http://cdn.iiit.ac.in/cdn/lttrc.iiit.ac.in/icon2015>.
- Hung, Jason C. and Che-Yu Yang. (2006). Word sense disambiguation using word ontology and concept distribution. Retrieved from <https://doi.org/10.1080/02533839.2009.9671494>.
- Ibrahim, Wayees Jejjud. (2005). Ambiguity. In *College of Basic Education Researchers Journal*, Vol. 2, No. 4. Mosul University College Of Basic Education.
- Jha, Girishnath. (2010). The TDIL program and the Indian Language Corpora Initiative (ILCI). Retrieved from http://www.lrecconf.org/proceedings/lrec2010/pdf/874_Paper.pdf.
- Kang, Sin Jang. (August 2004). Ontology Construction and Its Application to Disambiguate Word Senses. DOI: 10.3745/KIPSTB.2004.11B.4.491. Retrieved from <https://www.researchgate.net/publication/264107411>.
- Kharel, Dashrath. (2008). *Darjeeling ra Sikkim ma boline Nepali bhasako Varnanatmak adhyayan*. An unpublished PhD Dissertation. Darjeeling: Department of Nepali, University of North Bengal.
- LDCIL. 2014. <https://www.ldcil.org/resourcesParallelTextCorp.aspx>.
- LDCIL, 2014. <https://www.ldcil.org/resourcesTextCorp.aspx>.

- LDCIL, 2014. <https://www.ldcil.org/resourcesSpeechCorp.aspx>.
- Levin, Beth. (1993). *English Verb Classes and Alternations*. Chicago: University of Chicago Press.
- Leech, G. (1987). *Semantics*. Harmondsworth: Penguin Books.
- Lohani, Ram Raj. (1999). Permissive Constructions in Nepali: Evidence from Complex Predicates. In *Nepalese Linguistics*, Vol.16. Kathmandu, Nepal: Linguistic Society of Nepal.
- Lohani, Ram Raj. (2001). Valency Increase in Nepali Predicates: The Processes of Transitivity and Causativization. In *Nepalese Linguistics*, Vol.18. Kathmandu, Nepal: Linguistic Society of Nepal.
- Lohani, Shridhar Prasad and Rameshwar Prasad Adhikari (Eds.). (2010). *Ekta Concise Nepali-to-English Dictionary*. Kathmandu, Nepal: Ekta Books.
- Lyons, J. (1977). *Semantics*. Cambridge: Cambridge University Press.
- Manger, Krishna Maya. (August, 2014). Numeral Classifiers in Nepali. In *Language in India*, Vol.20:8. Retrieved from <http://www.languageinindia.com/aug2020/krishnanepalinumeralclassifiersfinal1.pdf>
- Manger, Krishna Maya. (June, 2023). Development of Verb Frames for Nepali. In *Language in India*, Vol.23:6. Retrieved from <http://www.languageinindia.com/june2023/krishnaverbframesNepalifinal.pdf>
- Maurya, A.S. and Bahadur, P. (2022). A Detailed Analysis of Word Sense Disambiguation Algorithms and Approaches for Indian Languages. Retrieved from https://link.springer.com/chapter/10.1007/978-981-16-3346-1_56.
- Miller, George A. (1995). WordNet: A Lexical Database for English. Retrieved from <https://dl.acm.org/doi/10.1145/219717.219748>.
- Palmer Martha, Dan Gildea, Paul Kingsbury. (2005). The Proposition Bank: A Corpus Annotated with Semantic Roles. In *Computational Linguistics Journal*, Vol. 31:1. Retrieved from <https://aclanthology.org/J05-1004.pdf>.
- Pant, G.P. (2000). Semantico-Syntactic Classification of Nepali Verbs. M.A Dissertation. Kathmandu, Nepal: Central Department of Linguistics, Tribhuvan University.
- Pant, Ganga Ram. (2001). Logical Structure of Nepali Verbs. In *Nepalese Linguistics*, Vol.18. Kathmandu, Nepal: Linguistic Society of Nepal.
- Parajuli, Krishna Prasad (Ed.). (2010). *Nepali Brihat Sabdakosh* (7th Edition). Kathmandu, Nepal: Nepal Academy.
- Pokharel, Madhav P. (1991). Compound Verbs in Nepali. In *Contribution to Nepalese*

- Studies*, Vol.18:2. Kathmandu, Nepal: CNAS, Tribhuvan University.
- Pokharel, Madhav P. (November, 1992). Moods in Nepali. In *Nepalese Linguistics*, Vol.9. Kathmandu, Nepal: Linguistic Society of Nepal.
- Pokharel Madhav P. (1997). *Nepali VakyaVyakaran*. Kathmandu, Nepal: Ekta Publishing House.
- Pokharel, M.P. (2054 BS). *Nepali VakyaVyakarana*. Kathmandu, Nepal: Nepal Rajkiya Prajna Pratisthan.
- Pokharel, M.P. (1999). Middle Verbs in Nepali. Paper presented in *5th Himalayan Language Symposium*, Kathmandu.
- Poudyal, Netra Prasad. (2009). The syntactic alignments across three-argument verbs in Nepali. In *Nepalese Linguistics*, Vol.24. Kathmandu, Nepal: Linguistic Society of Nepal.
- Poudel, Madhav Prasad. (2015). *Nepali Kriyahaaruko kosha*. Kathmandu, Nepal: Vidhyarthi Prakashan Private Limited.
- Prakasam, V. and Anvita Abbi. (1993). *Semantic Theories and Language Teaching*. New Delhi: Allied Publishers.
- Prasain, Balaram. (November, 2008). Computational analysis of Nepali basic verbs (written forms). In *Nepalese Linguistics*, Vol.23. Kathmandu, Nepal: Linguistic Society of Nepal.
- Prasain, Balaram. (2011). *Computational Analysis of Nepali Morphology: A Model for Natural Language Processing*. An unpublished PhD Dissertation. Kathmandu, Nepal: Central Department of Linguistics, Tribhuvan University.
- Roy, Arindam, Sunita Sarkar and Bipul Syam Purkayastha. (June 2014). Knowledge-Based Approaches to Nepali Word Sense Disambiguation. In *International Journal on Natural Language Computing*, Vol. 3, No.3. Retrieved from <http://airccse.org/journal/ijnlc/papers/3314ijnlc05.pdf>.
- Sheth, M., Popat, S. and Vyas, T. (2018). Word Sense Disambiguation for Indian Languages. Retrieved from https://link.springer.com/chapter/10.1007/978-981-10-4741-1_50.
- Schuler, Karin Kipper. (2005). VerbNet: A broad-coverage, comprehensive verb lexicon. Retrieved from <https://repository.upenn.edu/dissertations/AAI3179808>.
- Schulte Im Walde, Sabine. (2009). The induction of verb frames and verb classes from corpora. Retrieved from <http://verbs.colorado.edu/~mpalmer/Ling7800/SabineReview.pdf>.
- Shrestha, Niraj Patrick, A.V. Hall and Sanat K. Bista. (2008). Resources for Nepali Word

- Sense Disambiguation. In *Proceedings of International Conference on Natural Language Processing and Knowledge Engineering*, 19-22 October, 2008. Beiyou Building of Science and Technology, Beijing, China.
- Singh, Satyendra. (2015). Role of Karaka Relations in Hindi Word Sense Disambiguation. In *Journal of Information Technology*, Vol. 8:3. DOI: 10.4018/JITR2015070102. Retrieved from <https://www.igi-global.com/article>.
- Singh, S., Rauniyar, R., Manohar, M. (2021). Nepali Word-Sense Disambiguation Using Variants of Simplified Lesk Measure. Retrieved from https://doi.org/10.1007/978-981-16-1681-5_4.
- Vendler, Z. (1967). *Linguistics in Philosophy*. New York: Cornell University Press.
- Wasow, Thomas, Amy Perfors, and David Beaver. (2005). The Puzzle of Ambiguity. Retrieved from <https://web.stanford.edu/~wasow/Lapointe.pdf>.
- Wallace, W.D. (1979). The Syntax and Semantics of Causatives in Nepali. In Kachru B. (Ed.). *South Asian Language Analysis-I*. Urbana, Champaign: Department of Linguistics, University of Illinois.
- Yesuf, Segid Hassen. (2015). Amharic Word Sense Disambiguation using WordNet. Retrieved from <https://www.researchgate.net/publication/305995585>.
- Yoonus, M. Mohamed and Samar Sinha. (2011). A Hybrid Based POS Tagger for Indian Languages. In *Language in India*, Vol.11:9, September 2011. Retrieved from <http://www.languageinindia.com/sep2011/yoonusposttaggersfinalp.pdf>

APPENDIX – I

List of Ambiguous Verbs taken for building Verb Frames

S.N	Verb	No. of Sense	S.N.	Verb	No. of Sense	S. N.	Verb	No. of Sense	S.N.	Verb	No. of Sense
1	<i>au</i>	2	51	<i>chuɿ</i>	3	101	<i>pirol</i>	2	151	<i>mic</i>	2
2	<i>āk</i>	2	52	<i>dʒamɬ</i>	2	102	<i>pug</i>	2	152	<i>mil</i>	5
3	<i>ukas</i>	3	53	<i>dʒam</i>	2	103	<i>puch</i>	2	153	<i>mal</i>	2
4	<i>uker</i>	2	54	<i>dʒal</i>	2	104	<i>purjau</i>	2	154	<i>muɿ</i>	2
5	<i>ukel</i>	2	55	<i>dʒan</i>	2	105	<i>pokh</i>	2	155	<i>much</i>	2
6	<i>uksi</i>	2	56	<i>dʒag</i>	2	106	<i>phɬ</i>	2	156	<i>munʈi</i>	2
7	<i>ukhel</i>	2	57	<i>dʒogau</i>	2	107	<i>phāɿ</i>	2	157	<i>meɿ</i>	2
8	<i>ughri</i>	2	58	<i>dʒoɿ</i>	3	108	<i>phuk</i>	2	158	<i>rac</i>	2
9	<i>ughar</i>	2	59	<i>dʒhar</i>	2	109	<i>phuɿ</i>	2	159	<i>rasau</i>	2
10	<i>ucal</i>	2	60	<i>dʒhik</i>	2	110	<i>phul</i>	2	160	<i>raɬɬ</i>	3
11	<i>uɰh</i>	6	61	<i>dʒhos</i>	2	111	<i>pher</i>	2	161	<i>reɿ</i>	2
12	<i>uɿ</i>	2	62	<i>ɬāg</i>	2	112	<i>phjāk</i>	2	162	<i>ranki</i>	3
13	<i>utar</i>	2	63	<i>ɬās</i>	2	113	<i>bag</i>	2	163	<i>rakh</i>	2
14	<i>utri</i>	2	64	<i>ɬip</i>	2	114	<i>baɬɰ</i>	2	164	<i>rāki</i>	3
15	<i>udau</i>	2	65	<i>ɬek</i>	2	115	<i>baɰh</i>	4	165	<i>rapi</i>	2
16	<i>uphri</i>	3	66	<i>ɬok</i>	2	116	<i>ban</i>	2	166	<i>rūg</i>	2
17	<i>ubɰzi</i>	2	67	<i>ɬag</i>	2	117	<i>bas</i>	2	167	<i>ruc</i>	2
18	<i>ubhi</i>	2	68	<i>ɬunɰ</i>	2	118	<i>baɬau</i>	2	168	<i>rop</i>	2
19	<i>umri</i>	2	69	<i>ɬhel</i>	2	119	<i>basau</i>	2	169	<i>ɬagau</i>	4
20	<i>oɰh</i>	2	70	<i>ɬhok</i>	3	120	<i>bādh</i>	3	170	<i>ɬachar</i>	2
21	<i>kɬmau</i>	3	71	<i>qɬh</i>	2	121	<i>bigrɬ</i>	2	171	<i>ɬaɿ</i>	2
22	<i>kɬrau</i>	2	72	<i>qɬɬ</i>	2	122	<i>bit</i>	3	172	<i>ɬɰthar</i>	2
23	<i>kɬs</i>	2	73	<i>qɬhak</i>	2	123	<i>bisau</i>	2	173	<i>ɬanɰi</i>	2
24	<i>kaɿ</i>	6	74	<i>ɬap</i>	2	124	<i>bhan</i>	2	174	<i>ɬammi</i>	2

25	<i>kaṭh</i>	7	75	<i>taṛ</i>	2	125	<i>bhaṛ</i>	2	175	<i>lag</i>	15
26	<i>kiṭ</i>	2	76	<i>taṛka</i>	2	126	<i>bhag</i>	2	176	<i>li</i>	2
27	<i>kuṭ</i>	2	77	<i>tan</i>	2	127	<i>bhidḡ</i>	2	177	<i>ljau</i>	2
28	<i>ker</i>	2	78	<i>tok</i>	4	128	<i>bhaṭkur</i>	2	178	<i>saḱ</i>	3
29	<i>khap</i>	2	79	<i>thak</i>	2	129	<i>bhaṭka</i>	2	179	<i>saṅki</i>	2
30	<i>khas</i>	2	80	<i>tham</i>	2	130	<i>bhasi</i>	2	180	<i>sapri</i>	2
31	<i>kha</i>	2	81	<i>thun</i>	2	131	<i>bhāṭ</i>	2	181	<i>saṃau</i>	3
32	<i>khijau</i>	2	82	<i>dabau</i>	2	132	<i>bhaṭka</i>	2	182	<i>sābhal</i>	2
33	<i>khul</i>	6	83	<i>dag</i>	3	133	<i>bhaṅki</i>	2	183	<i>saḱki</i>	2
34	<i>khel</i>	4	84	<i>di</i>	2	134	<i>bhir</i>	2	184	<i>saṛ</i>	3
35	<i>khodḡ</i>	2	85	<i>dekh</i>	2	135	<i>bhuṭ</i>	2	185	<i>sarḱi</i>	2
36	<i>khol</i>	4	86	<i>nag</i>	2	136	<i>bhul</i>	2	186	<i>suk</i>	2
37	<i>gaḷ</i>	2	87	<i>nap</i>	2	137	<i>bheṭ</i>	2	187	<i>sulki</i>	2
38	<i>gaṛ</i>	2	88	<i>nikal</i>	2	138	<i>bhog</i>	2	188	<i>haḷli</i>	2
39	<i>ghaṭ</i>	2	89	<i>niski</i>	2	139	<i>bhjau</i>	2	189	<i>haṛau</i>	2
40	<i>ghum</i>	2	90	<i>nihuri</i>	2	140	<i>maḱci</i>	2	190	<i>haṭ</i>	2
41	<i>caṭh</i>	2	91	<i>paḱra</i>	2	141	<i>maṇau</i>	2	191	<i>hāḱ</i>	3
42	<i>caṃka</i>	2	92	<i>paḱau</i>	2	142	<i>maṭh</i>	2	192	<i>hal</i>	4
43	<i>caṛka</i>	2	93	<i>paḡla</i>	2	143	<i>maṛ</i>	4	193	<i>haṭ</i>	2
44	<i>caḷ</i>	6	94	<i>paḱau</i>	2	144	<i>maṛka</i>	2	194	<i>her</i>	6
45	<i>cjat</i>	2	95	<i>paṛ</i>	4	145	<i>maḱki</i>	2	195	<i>han</i>	4
46	<i>cjap</i>	2	96	<i>paḱ</i>	2	146	<i>maḡh</i>	2	196	<i>hu</i>	2
47	<i>chāṭau</i>	2	97	<i>paḱau</i>	2	147	<i>man</i>	2	197	<i>haḱar</i>	2
48	<i>char</i>	2	98	<i>pau</i>	2	148	<i>mat</i>	2	198	<i>har</i>	2
49	<i>chaṭ</i>	2	99	<i>pak</i>	2	149	<i>mar</i>	2	199	<i>haṃka</i>	2
50	<i>chan</i>	2	100	<i>pal</i>	2	150	<i>mas</i>	2	200	<i>hūṭaḷ</i>	2

APPENDIX - II

Semantic Classification of Verbs based on Levin (1993)

1 Verbs of Putting

1.1 Put Verbs

Class Members: milaunu, dʒuraunu ‘arrange’; dʒubanu, gaɾnu, bʰidʒaunu ‘immerse’; dʒaɾnu ‘install’; rakʰnu, ʌɾnu ‘lodge’; chʌɾnu, ukɭanu, baɾnu ‘mount’; rakʰnu, milaunu, thʌkjaunu ‘place/put’; ɬaɲnu ‘sling’; lukaunu ‘stash’ etc.

1.2 Verbs of Putting in a Spatial Configuration

Class Members: dʒhundʒinu, ʌɬkʌnu ‘dangle’; dʒhulaunu, ɬaɲnu ‘hang’; bisaunu, bichjaunu, ʌdʌunu ‘lay’; ʌdʌs ʌɲnu, dʰessinu ‘lean’; basnu, stʰit hunu ‘sit’; sustaunu, ʌɬnu, dʰʌɭkʌnu ‘rest’; ubhinu, ʌglo hunu, ɬʌdʒæunu ‘stand’; ʌdʒæunu ‘suspend’, etc.

1.3 Funnel Verbs

Class Members: baɬʒarnu ‘bang’; chopnu, dʒubki ʌunu, dʒʰarnu ‘dip’; thuparnu, khādn̩u, phjākn̩u ‘dump’; khʌɲjaunu ‘ladle’; hʌlnu ‘pound’; push ‘ɬhelnu’; tannu, khicnu ‘rake’; piɬnu, sʌmma pʌnu ‘ram’; ubhaunu, udʒæunu, udʒhelnu, uɬhaunu ‘scoop’; khurkʌnu, kornu, kotʌnu ‘scrape’; hʌllaunu, ghʌcghʌcæunu, dʒʰʌɬkʌnu, hūɖʌlnu ‘shake’; nimoɬnu, nicornu, ɬosnu, hulnu ‘squeeze’; baɖʌlnu ‘sweep’; dobrjaunu, khumejaunu ‘tuck’; ɬʌlnu ‘wad’; puchnu ‘wipe’, etc.

1.4 Verbs of Putting with a Specified Direction

Class Members: khʌsalnu, cuhaunu ‘drop’; uɬhaunu ‘hoist’; ucalnu, boknu ‘lift’; ohralnu, utarnu, khʌsalnu, dʒʰarnu ‘lower’; udʒjaunu, ʌgljaunu ‘raise’, etc.

1.5 Pour Verbs

Class Members: nithranu, nikhranu, turkaunu, cuhaunu, cuhunu ‘dribble’; tʌrkʌnu, ɬʌpkʌnu, tʌpkʌnu ‘drip’; poknu, khʌɲjaunu, ghoptjaunu ‘pour’; piɬnu, ɬʰoknu ‘slosh’; vanta gaɾnu ‘spew’; chirkjaunu ‘spurt’, etc.

1.6 Coil Verbs

Class Members: baɬarinu, baɬinu ‘coil’; ghumrinu, ghūgurinu ‘curl’; gāɬho pʌnu ‘loop’; ghumnu ‘spin’; ghumnu, phʌnkinu ‘twirl’; riɲnu ‘whirl’ etc.

1.7 Spray/Load Verbs

Class Members: kocnu, bharnu ‘cram’; dʒhjamminu ‘crowd’; thicnu ‘dab’, potnu ‘daub’, dʰaknu, ɬaɲgnu ‘drape’; simsim pani pʌnu ‘drizzle’; dʒhundʒinu, dʒholinu, ʌrkʌnu ‘hang’; thopnu, culinu ‘heap’; hʌlnu ‘inject’; rokinu, ʌɖkinu ‘jam’; ladnu, bokaunu, bharnu ‘load’; ropnu ‘plant’; lep ʌunu ‘plaster’; ghocnu, khopnu ‘prick’; ghoɬnu, rʌgaɬnu, maɬnu, dʌlnu, ghʌsnu ‘rub’; chʌnu ‘scatter’; basnu, thaminu, ʌɬinu ‘settle’; siunu ‘sew’; dʌrkʌnu, barsanu ‘shower’; ʌɬpatinu ‘smear’; chʌnu ‘sow’; chjapnu ‘spatter’; chʌrkʌnu, sirkaunu ‘spray’; pasarnu, phīdʒarnu, phʌilaunu ‘spread’;

chamkanu ‘sprinkle’; thak parnu ‘stack’; tāsnu, tasnu ‘stick’; gūthnu, gāsnu, unnu ‘string’; thicnu, koccinu, latarnu ‘stuff’; sosnu ‘swab’; dhunu, pakhalnu ‘wash’; bernu, lapetnu, gutmutjaunu ‘wrap’, etc.

1.8 Fill Verbs

Class Members: sadzaunu ‘adorn’; bādhnu ‘bind’; cheknu, roknu, budzjaunu ‘block’; sosnu, latpatinu ‘blot’; sarkinu, nissasinu ‘choke’; lukaunu ‘cloak’; thunnu, thuninu ‘clog’; mailinu ‘contaminate’; chopnu, dhaknu, oṭhnu, khaṇnu ‘cover’, bādhnu, roknu ‘dam’, sīgarnu, adzaunu ‘decorate’; nibhaunu, bhidzhaunu ‘douse’; bhidznu ‘drench’; ghernu ‘encircle’; posnu ‘enrich’; phāṣnu, laṭharinu, dželinu, gāsnu ‘entangle’; bharnu, purnu ‘fill’; urlanu ‘flood’; bhārinu, rāṇgaunu ‘imbue’; dzaḍnu ‘inlay’, misnu ‘interlard’; khāḍnu ‘pad’; baḍznu ‘ring’; kūḍnu, khopnu ‘stipple’; gaḍnu ‘stud’; ghernu ‘surround’; bernu ‘swaddle’; lapetnu ‘swathe’; chaṇnu ‘tile’; chāṭnu ‘trim’; chopnu ‘veil’, etc.

1.9 Butter Verbs

Class Members: lobhjaunu ‘bait’; chepre parnu, lolopoto garnu, chiplo launu ‘butter’; birko launu ‘cap’; lukaunu ‘cloak’, laṭṭinu, matnu ‘drug’; cheknu ‘fence’; budzo launu ‘gag’; cipjaunu, cillo parnu ‘grease’; kaṇnu ‘groove’; kaṣnu ‘harness’; tel launu ‘oil’; ṭalnu ‘patch’; ṭalkaunu ‘polish’; sīcnu, chiṭnu ‘water’; potnu, chuna launu ‘whitewash’, etc.

1.10 Pocket Verbs

Class Members: phela parnu ‘bag’; kaḍ garnu ‘cage’; thunnu ‘coop’; pakrinu ‘corral’; caraunu ‘pasture’; ot lagnu ‘shelter’; thaṅkaunu ‘shelve’; kādh halnu ‘shoulder’; dzhirma unnu ‘skewer’; djalma parnu ‘snare’; thuknu ‘spit’; pasoma parnu ‘trap’, etc.

2 Verbs of Removing

2.1 Remove Verbs

Class Members: dzhiknu, nikalnu ‘abstract’; ṭipnu ‘cull’; meṭaunu ‘delete’; choṭnu, haṭaunu, phjāknu ‘discharge’; pharkaunu ‘disgorge’; ukhelnu, phukalnu ‘dislodge’; kharidz garnu ‘dismiss’; chuṭnu, mukṭa garnu ‘disengage’; tannu ‘draw’; nikasnu, kaṭhnu ‘eject’; haṭaunu ‘eliminate’; ukhelnu ‘eradicate’; nikalnu ‘evict’; kaṭnu ‘excise’; nikalnu, uchitjaunu, milkaunu, galhatjaunu ‘expel’; dzhiknu, phukalnu, thutnu, ukhelnu ‘extract’; ṭhelnu, ghaceṭnu ‘extrude’; chimalnu ‘lop’; meṭanu ‘omit’; bādhnu, vibhadzan garnu, barnu, cheknu ‘partition’; cijaunu ‘pry’; kaṭnu ‘reap’; haṭaunu, paṅchaunu ‘remove’; chuttjaunu, alag garnu ‘separate’; lagharu ‘shoo’; ghāṭaunu ‘substract’; ukhelnu ‘uproot’; phirta linu, paṅchaunu, khicnu ‘withdraw’; tannu, baralnu, khosnu ‘wrench’, etc.

2.2 Banish Verbs

Class Members: nikalnu, pabandi launu ‘banish’; khali garnu, phernu, haṭnu ‘evacuate’; nikalnu, galhatjaunu, uchitjaunu, milkaunu ‘expel’; sumpānu ‘extradite’; samzhanu ‘recall’; haṭaunu ‘remove’, etc.

2.3 Clear Verbs

Class Members: phāḍḍḍu, ugharnu, kholnu ‘clear’; sapha gāṛnu, madḥhnu, lipnu ‘clean’; baḥaunu, nitharnu, nikharnu, tārkinu, tāḥarīnu ‘drain’; rittinu, ritjaunu, sakinu, siddhinu, khali gāṛnu ‘empty’, etc.

2.4 Wipe Verbs

Manner Subclass

Class Members: hannu, thicnu ‘dab’; tapkjaunu ‘distill’; kernu, meṭnu ‘erase’; kaṭnu, haṭaunu ‘expunge’; paḥhalnu ‘flush’; ritjaunu ‘leach’; caṭnu ‘lick’; ṭipnu ‘pluck’; ṭaḷkaunu ‘polish’; chimāḷnu, chāṭnu, kaṭnu ‘prune’; kharnu ‘purge’; dhunu, chuṭhnu, chopāḷnu ‘rinse’; ghoṭnu, ghasnu, raḡeṭnu, raḡaṭnu, maṭnu ‘rub’; maṣkaunu, madḥhnu ‘scour’; khurkaṇu, tachnu, kornu ‘scrape’; madḥhnu, māḷnu ‘scrub’; khourānu ‘shave’; tar kaṭhnu ‘skim’; cilljaunu ‘smooth’; bhidjaunu, sosnu, ḡhaḍjaunu ‘soak’; nicornu, nimoṭhnu ‘squeeze’; channu, calnu ‘strain’; naṅgjaunu ‘strip’; chusnu, tannu ‘suck’; sosnu ‘swab’; baṭarnu ‘sweep’; milaunu ‘trim’; dhunu, paḥhalnu ‘wash’; paḥhirinu, launu ‘wear’, goṭnu ‘weed’; ḡḡharnu, uḡaunu ‘whisk’; niphāṇnu, kelaunu ‘winnow’; puchnu ‘wipe’, etc.

Instrument Subclass

Class Members: burus launu ‘brush’; kornu, baṭnu ‘comb’; reṭnu, kūḍnu, reṭnu ‘file’; channu, calnu ‘filter’; istri launu ‘iron’; puchnu ‘mop’; soḥarnu ‘rake’; kaṭnu ‘shear’; khaṇjaunu, uṭhaunu ‘shovel’; meṭaunu ‘sponge’; sosnu, puchnu ‘towel’; cusnu ‘vacuum’, etc.

2.5 Verbs of Possessional Deprivation: Steal Verbs

Class Members: harnu, bhāgaunu ‘abduct’; magnu ‘cadge’; paḥraṇu ‘capture’; khosnu, ḡḡapht gāṛnu ‘confiscate’; paḥraṇu ‘cop’; mukta gāṛnu ‘emancipate’; masnu, marnu ‘embezzle’; mānsaunu, phuknu, ḡḡharnu ‘exorcise’; dhutnu, ṭhāgnu ‘extort’; ḡḡhiknu, phukaḷnu, thutnu, ḡḡharnu, kaḍhnu ‘extract’; cornu, bheṭṭaunu, phutkaunu ‘filch’; korra ṭhoknu ‘flog’; luchnu, khosnu ‘grab’; thunnu ‘impound’; harnu, lukaunu, aṇaḥaṇ gāṛnu ‘kidnap’; mukta gāṛnu, choṭnu ‘liberate’; ucalnu, boknu ‘lift’; samatnu ‘nab’; caṇkaunu ‘pilfer’; cimoṭnu ‘pinch’; luṭnu ‘pirate’; bhavaḥaṇ gāṛnu ‘plagiarize’; cornu, uṭaunu ‘purloin’; pharkaunu ‘recover’; nikhaṇnu, cukaunu, phirta gāṛnu ‘redeem’; raksha gāṛnu, uddar gāṛnu, baḥaunu ‘rescue’; ḡḡapht gāṛnu ‘seize’; taskari gāṛnu ‘smuggle’; khosnu, thutnu, ḡḡhamṭīnu ‘snatch’; linu ‘take’, etc.

2.6 Verbs of Possessional Deprivation: Cheat Verbs

Class Members: choṭnu, maph gāṛnu ‘absolve’; chuṭkara paunu ‘acquit’; napaunu, niras parnu ‘balk’; Dāljaunu, luṭnu, khosnu ‘bereave’; ṭhāgnu, chāḷnu ‘bilk’; cornu ‘burgle’; Dhaṭnu, chāḷnu, dhoka dinu, ghat gāṛnu, chākaunu ‘cheat’; ṭipnu ‘cull’; niko parnu, ḡḡṭati parnu ‘cure’; dhutnu, paḥ parnu, chāḥkaṇaṭ gāṛnu ‘defraud’; nango parnu, utarnu ‘denude’; ritjaunu, siddhjaunu ‘deplete’, etc.

2.7 Pit Verbs

Class Members: bhuknu ‘bark’; choḍjaunu ‘hull’; kuṭnu, dāḷnu, naṅgjaunu ‘husk’; louse, milk, tachnu, upkaunu, choḍaunu, koṭjaunu, bungjaunu, bokjra ukkaunu ‘peel’; paḥheṭa kaṭnu, hat bādhnu, ḡḡkaṭjaunu ‘pinion’, khaṭāḥma haḷnu ‘pit’; kosa lagnu ‘pod’; kaṇal/kes kaṭnu, koṭarnu, ṭuppo kaṭnu ‘poll’, etc.

2.8 Debone Verbs

Class Members: bahiskar garnu ‘debark’; uDus marnu/dhapaunu ‘debug’; dʒaŋgal phāṛnu ‘deforest’; paḡalnu ‘defrost’, etc.

2.9 Mine Verbs

Class Members: mine, quarry

3 Verbs of Sending and Carrying

3.1 Send Verbs

Class Members: dʒanaunu, osarnu, purjaunu, laiḍʒanu ‘convey’; paṭhaunu, baṭhaunu, presit garnu, calan garnu ‘forward’; saḡhaunu, dinu ‘hand’; par garnu, parit hunu, calnu, bitnu, ṭarnu ‘pass’; pharkaunu ‘return’; paṭhaunu ‘send’; sarnu ‘shift’; ciplanu, raṭkanu ‘slip’; taskari garnu ‘smuggle’; sarnu, sarnu ‘transfer’; Dhuvani garnu, laiḍʒanu, osarnu ‘transport’, etc.

3.2 Slide Verbs

Class Members: uphranu, upharnu ‘bounce’; baḡnu, tairinu, taḡharinu, utranu ‘float’; calnu, sarnu ‘move’; guṽDulkinu, guDnu ‘roll’; ciplanu ‘slide’, etc.

3.3 Bring and Take

Class Members: ljaunu ‘bring’; linu, samatnu, pakranu, paunu, prapta garnu, laiḍʒanu, khosnu ‘take’, etc.

3.4 Carry Verbs

Class Members: vaḡan garnu, bhirnu, laiḍʒanu, boknu ‘carry’; tannu, ghissjaunu, ghicjaunu, ghisarnu, laḡharnu ‘drag’; ucalnu, uṭhaunu ‘heave/heft/hoist’; lat hannu, bhakunDjaunu ‘kick’; tannu ‘pull’; ṭhelnu, dhakelnu, ḡhaḡeṭnu, ṭhosnu, hulnu, hutjaunu ‘push’, etc.

3.5 Drive Verbs

Class Members: hāḡnu, kudaunu ‘drive’; uDnu, uDaunu ‘fly’; khijaunu ‘row’; guDaunu ‘wheel’, etc.

4 Verbs of Exerting Force: Push/Pull Verbs

Class Members: tannu ‘draw’; ḡhaḡcjaunu, upharnu ‘jerk’; thicnu ‘press’; tannu ‘pull’; ṭhelnu, dhakelnu, ḡhaḡeṭnu, ḡhokrjaunu ‘push’, thutnu ‘yank’, etc.

5 Verbs of Change of Possession

5.1 Give Verbs

Class Members: khuvaunu, palnu, hercaḡ garnu ‘feed’; dinu ‘give’; kutma dinu ‘lease’; paṭiṽco dinu, sapṭaṭ dinu ‘lend’; rin kaṭhnu ‘loan’; tirnu, cukaunu ‘pay’; pharkaunu, phirta dinu ‘refund’; budḡhaunu, sumpānu ‘render’; bhaṭama dinu ‘rent’; becnu ‘sell’; paṣkanu ‘serve’, etc.

5.2 Contribute Verbs

Class Members: jogdan dinu ‘contribute’; behornu, kharca garnu, ‘disburse’; bāDnu, vitaran garnu ‘distribute’; canda/dan dinu ‘donate’; arpan garnu, caDaunu ‘proffer’; bharnu ‘reimburse’; tjag/utsarga dinu ‘relinquish/sacrifice’; chama/maph garnu ‘remit’; pharkaunu ‘return’; budzhaunu, prastut garnu ‘submit’; atmasamarpan garnu ‘surrender’, etc.

5.3 Verbs of Future Having

Class Members: baDnu, briddhi hunu, aghi sarnu, agrasar hunu ‘advance’; bāDnu, bhag launu ‘allocate’; toknu ‘allot’; sumpānu ‘assign’; puraskar dinu ‘award’; mannu ‘concede’; barthaunu ‘extend’; svikar garnu ‘grant’; dzari garnu, dinu, ‘issue’; choDnu ‘leave’; carthaunu, tākrijaunu ‘offer’; vacan dinu, kasam khanu ‘promise’; matadan dinu ‘vote’; utpadan garnu ‘yield’, etc.

5.4 Verbs of Fulfilling

Class Members: patjaunu, bisvas garnu ‘trust/entrust’; upasthit hunu, hadzir hunu, prastut garnu, upahar dinu ‘present’; upalabdha garaunu ‘provide’; seva garnu ‘serve’; purjaunu, apurti garnu, dzuṭaunu ‘supply’, etc

5.5 Equip Verbs

Class Members: bodzh hunu ‘burden’; chatipurti garnu ‘compensate’; vjaja garnu ‘invest’; inam dinu ‘reward’; bhari bokaunu, bodzh ladnu ‘saddle’, etc.

5.6 Verbs of Obtaining

Get Verbs

Class Members: kaṭnu, kinnu ‘book’; kinnu ‘buy’; bolaunu ‘call’; pakranu ‘catch’; channu, rodznu ‘choose’; kamaunu ‘earn’; ljaunu, tannu ‘fetch’; paunu, phela parnu ‘find’; napha hunu, dzitnu ‘gain’; bhela garnu ‘gather’; paunu, bhetṭaunu ‘get’; bhaṭama linu ‘hire’; rakhnu ‘keep’; kutma linu ‘lease’; agja/ades/hukum dinu ‘order’, phulnu ‘flower’; phul ṭipnu ‘pluck flower’; pugnu ‘reach’; bhaṭama linu ‘rent’; arakshan garnu ‘reserve’; dzogaunu ‘save’; surakchit garaunu ‘secure’; kaṭnu, marnu, capkanu ‘slaughter animal’; cornu ‘steal’; dzitnu ‘win’, etc.

Obtain Verbs

Class Members: svikar garnu, svikarnu ‘accept’; ardzan garnu, paunu ‘acquire’; paico/sapaṭi linu ‘borrow’; magnu ‘cadge’; bhela garnu, sankalan garnu ‘collect’; samatnu, pakranu ‘grab’; prapta garnu, paunu ‘obtain’; kinnu ‘purchase’; channu, rodznu ‘select’; khosnu, thutnu ‘snatch’, etc.

5.7 Verbs of Exchange

Class Members: saṭnu, saṭpher garnu ‘barter’; badlinu, phernu ‘change’; saṭnu ‘exchange/substitute/swap’, etc.

5.8 Berry Verbs

Class Members: macha marnu ‘fish’, cara samaunu/marnu ‘fowl’, etc.

6 Learn Verbs

Class Members: ardʒan ɡarnu ‘acquire’; rɒʈnu, ɡhoknu, kanʈhɒ ɡarnu ‘cram’; siknu ‘learn’; ɡhoknu, kanʈhasthɒ ɡarnu ‘memorize’; pɒʈnu ‘read’; ɒdhjɒʒan ɡarnu ‘study’, etc.

7 Hold and Keep Verbs

7.1 Hold Verbs

Class Members: cɒjapnu ‘clasp’; ɒʈhjaunu ‘clutch’; muʈhjaunu ‘grasp’; pɒkrinu ‘grip’; samatnu, thamnu, rakhnu, roknu ‘hold’, etc.

7.2 Keep Verbs

Class Members: sangetnu, zamma ɡarnu ‘hoard’; rakhnu, sɒɡalnu, thanko launu, thunnu ‘keep’; bɒki choʈnu ‘leave’; thankjaunu, sɒcnu, zalgera ɡarnu ‘store’, etc.

8 Verbs of Concealment

Class Members: roknu, bandhɒ ɡarnu ‘block’; thunnu ‘cloister’; dɒbaunu ‘conceal’; parda launu, cheknu ‘curtain/screen’; luknu, lukaunu ‘hide’; chutjaunu, ɒlɒɡjaunu ‘isolate/quarantine sequester’; panchinu, panchjaunu ‘seclude’; asɒʒɒ dinu, ot lagnu ‘shelter’, etc.

9 Verbs of Throwing

9.1 Throw Verbs

Class Members: phjɒknu, dhalnu ‘cast’; guleli/ghujetro hannu ‘catapult’; chuck, rankanu, phankanu, salkanu ‘fire’; pjattɒ hannu ‘flick’; dʒhɒʈkarnu ‘fling’; hannu ‘hit ball’; hutjaunu ‘hurl’; ʈhoknu ‘knock’; dhakelnu ‘shove’; cɒʈkaunu, thɒppɒʈ hannu ‘slap’; zhɒʈaro hannu ‘sling’; toʈnu, kicnu ‘smash’; phjɒknu ‘throw’, etc.

9.2 Pelt Verbs

Class Members: hannu, kuʈnu, piʈu, cuʈnu, ʈɒʈʈaunu ‘buffet’; dɒrkjaunu ‘shower’; dʒhungjaunu ‘stone’, etc.

10 Verbs of Contact by Impact

10.1 Hit Verbs

Class Members: bɒʈʒarnu ‘bang’; toʈnu ‘batter’; hirkaunu ‘beat’; thɒccinu, thɒcarnu ‘bump’; dash, drum, hammer, hannu, ʈhoknu, hirkaunu ‘hit’; lat hannu ‘kick,’ etc.

10.2 Swat Verbs

Class Members: cɒbaunu, ʈoknu, dɒsnu, cilnu ‘bite’; citharnu, cɒjatnu, nɒɒrjaunu ‘claw’; ʈhuɒnu ‘peck’; ghussi hannu, muʈkile hannu ‘punch’; kornu ‘scratch’; ɡhocnu ‘slug’; ɡhopnu, ceʈnu ‘stab’; kicnu ‘swat’, etc.

10.3 Spank Verbs

Class Members: piʈnu, ʈhoknu ‘birch’; ʈauko phornu ‘brain’; ɡolmoljaunu, thuparnu ‘clobber’; ɒʈnu ‘conk’; kaʈnu ‘knife’; khijaunu ‘paddle’; cɒpetnu ‘spank’, etc.

10.4 Non-Agentive Verbs of Contact by Impact

Class Members: *baḍḍarnu* ‘bang’; *thacarnu* ‘bump’; *ṭakraunu*, *munṭjaunu* ‘crash’; *hannu*, *hirkaunu* ‘hit’, etc.

11 Poke Verbs

Class Members: *khannu*, *khosrinu*, *khotalnu* ‘dig’; *ghocnu* ‘jab/prick’; *ceṭnu* ‘pierce’; *ghocnu*, *ṭhosnu*, *ghaceṭnu* ‘poke’; *ṭāsinu* ‘stick’, etc.

12 Verbs of Contact: Touch Verbs

Class Members: *sumsumjaunu*, *chamnu* ‘caress’; *khosrinu*, *kornu* ‘graze’; *mvaĩ khanu* ‘kiss’; *caṭnu* ‘lick’; *kottjaunu* ‘nudge’; *ṭhūṇnu* ‘peck’; *cimoṭnu* ‘pinch’; *ghocnu* ‘prod’; *cilnu*, *ṭhūṇnu* ‘sting’; *kutkujaunu*, *kaukuti launu* ‘tickle’; *chunu* ‘touch’, etc.

13 Verbs of Cutting

13.1 Cut Verbs

Class Members: *kaṭnu*, *phornu* ‘chip’; *kotrānu* ‘clip’; *kaṭnu* ‘cut’; *reṭnu*, *sernu*, *chimalnu* ‘hack’; *cheṭnu* ‘hew’; *ara kaṭnu* ‘saw’; *khurkānu*, *tachnu* ‘scrape’; *kornu* ‘scratch’, etc.

13.2 Carve Verbs

Class Members: *khosrinu*, *pachnu* ‘bruise’; *kūḍnu* ‘carve’; *ṭukrjaunu* ‘chop’; *kiṭnu*, *micnu*, *pelnu*, *dālnu* ‘crush’; *kuccjaunu*, *thepcaunu* ‘dent’; *khopnu* ‘drill’; *reṭnu* ‘file’; *ghasnu* ‘grate’; *pīdhnu*, *pisnu*, *pinnu* ‘grind’; *mucnu* ‘mash’; *guṭnu*, *kuṭkjaunu* ‘mince’; *notch*; *cheṭnu* ‘perforate’; *chimalnu* ‘prune’; *cjatnu* ‘shred’; *ṭukrjaunu* ‘slice’; etc.

14 Verbs of Combining and Attaching

14.1 Mix Verbs

Class Members: *misaunu* ‘blend’; *milnu*, *misinu* ‘combine/mingle/mix’; *dzoṭinu* ‘concatenate/connect/ join/ link’; *gabhinu* ‘merge’; etc.

14.2 Amalgamate Verbs

Class Members: with: *gabhnu* ‘affiliate’; *dzoṭinu*, *misrit hunu* ‘amalgamate’; *saṅglagna hunu* ‘associate’; *dzuṭnu* ‘coalesce’; *saṅjog banu* ‘coincide’; *tulana garu* ‘compare’; *almalinu* ‘confuse’; *dḷamma garu* ‘consolidate’; *dāḍḍnu* ‘contrast’; *berinu*, *baṭnu* ‘entwine’; *phāsnu*, *dḷelinu* ‘entangle’; etc.

14.3 Shake Verbs

Class Members: *poko parnu* ‘bundle’; *dḷhumminu*, *dḷhuttinu* ‘cluster’; *bhela garu* ‘gather’; *ḍallinu* ‘lump’; *package*; *hallaunu* ‘shake’; *phiṭnu* ‘shuffle’; *calaunu* ‘stir’; *bhela garu*, *saṅkalān garu* ‘collect’; *gāsnu* ‘attach’; *bādhnu* ‘bind’; *kaṣnu* ‘fasten’; *siuranu* ‘graft’; *moor*; *siunu* ‘sew’; etc.

14.4 Tape Verbs

Class Members: *agljaunu* ‘bolt’; *ṭāk launu* ‘button’; *kaṣnu* ‘clamp’; *cjaṇnu* ‘clasp’; *nel halnu* ‘fetter’; *ankus launu* ‘hook’; *gāṭho parnu* ‘knot’; *tala launu* ‘lock’; *loop*; *ṭāsnu* ‘paste’; etc.

14.5 Cling Verbs

Class Members: gāsinu, tāsīnu, uldghinu ‘adhere’; lipsanu ‘cleave’; āgalo halnu ‘cling’; etc.

15 Verbs of Separating and Disassembling

15.1 Separate Verbs

Class Members: chuttjaunu, bhed gānu ‘differentiate/distinguish’; kaṭnu, vicched gānu ‘disconnect’; bhag gānu, bāṭnu ‘divide’; parpacuke gānu, chuttjīnu ‘divorce’; ālagjaunu ‘segregate’; chuttjaunu ‘separate’; etc.

15.2 Split Verbs

Class Members: phuknu ‘blow’; toṭnu, bhācnu, tuṭnu, cūṭnu ‘break’; kaṭnu ‘cut’; kholnu, phukaḷnu ‘pry’; tannu ‘pull’; thēnu, ghacēṭnu ‘push’; ciṭṭanu ‘slip’; cirinu ‘split’; cjaṭnu ‘tear’; etc.

15.3 Disassemble Verbs

Class Members: chuttjaunu ‘detach’; phukaḷnu ‘disassemble’; kaṭnu ‘disconnect’; barnu, cheknu ‘partition’; etc.

15.4 Differ Verbs

Class Members: ālagginu ‘differ’; chuttjīnu ‘diverge’; etc.

16 Verbs of Coloring

Class Members: raṅgjaunu ‘color/dye/paint’; tālkinu ‘glaze’; dag lagnu ‘stain’; etc.

17 Image Creation Verbs

17.1 Verbs of Image Impression

Class Members: khodnu, kūdnu ‘engrave’; chapnu ‘imprint’; cheṭnu ‘incise’; kūdnu ‘inscribe’; chinha/chap launu ‘mark’; raṅgaunu ‘paint’; milaunu ‘set’; sahi gānu, haṣṭakhaṛ gānu ‘sign’; chap launu ‘stamp’; khopnu ‘tattoo’; etc.

17.2 Scribble Verbs

Class Members: kūdnu ‘carve’; sarnu ‘copy’; kornu ‘doodle/draw’; māsi launu ‘ink’; raṅgaunu ‘paint’; chapnu ‘print’; kornu ‘scratch’; tṭāṅkaṅ gānu ‘type’; lekhnu ‘write’; etc.

17.3 Illustrate Verbs

Class Members: sambodhaṅ gānu ‘address’; saḍjaunu ‘adorn/decorate/embellish’; ānumodaṅ gānu ‘endorse’; udjalo pānu ‘illuminate’; vjakhya gānu ‘illustrate’; sahi gānu ‘initial’; namaḷaṅ gānu ‘label’; etc.

17.4 Transcribe Verbs

Class Members: sarnu ‘copy’; photo kicnu ‘photograph’; lekhnu, rekard gānu ‘record’; lipjaṅkaṅ gānu ‘transcribe’; etc.

18 Verbs of Creation and Transformation

18.1 Build Verbs

Class Members: milaunu ‘arrange’; dʒoɽnu ‘assemble’; banaunu ‘build/make’; kūdnu ‘carve’; dhalnu ‘cast’; gholnu ‘churn’; sankalān/ saṅgrahaḥ gānu ‘compile’; pakāunu ‘cook’; bunnu ‘crochet/knit’; kaṭnu ‘cut’; vistar gānu ‘develop’; paṭṭjaunu ‘fold’; pisnu, pīdhnu ‘grind’; baṭnu, palaunu ‘grow’; silaunu ‘stitch’; tunnu, baṭnu ‘weave’; tachnu ‘whittle’; etc.

18.2 Grow Verbs

Class Members: vikas gānu ‘develop/evolve’; baṭnu, palaunu ‘grow’; korālnu, othra kaṭhnu ‘hatch’; chippinu, paknu ‘mature’; etc.

18.2 Verbs of Preparing

Class Members: khana pakāunu ‘cook meal’; anda tarnu ‘fry egg’; seknu, pilsaunu, polnu ‘grill’; ago balnu ‘light fire’; dhunu ‘wash’; etc.

18.3 Create Verbs

Class Members: banaunu, nirman gānu ‘coin/construct/form’; racnu, sirdʒana gānu ‘create’; racnu ‘compose’; utpatti gānu ‘derive’; dhāca banaunu ‘design’; khaṇnu ‘dig’; gaṭhnu, kaṭhnu ‘fabricate’; aviskar gānu ‘invent’; utpadan gānu ‘manufacture’; aḷodʒan gānu ‘organize’; ubdʒaunu ‘produce’; etc.

18.4 Knead Verbs

Class Members: kuṭnu ‘beat’; nihurinu ‘bend’; baṭarnu ‘coil’; dʒamma/sankalān gānu ‘collect’; khādnū ‘compress’; paṭṭjaunu ‘fold’; muchnu ‘knead’; pāgalnu ‘melt’; hallaunu ‘shake’; kicnu ‘squash’; nicornu ‘squeeze’; etc.

18.5 Turn Verbs

Class Members: saṭnu, phernu ‘alter/change’; badlinu ‘convert/transform’; rupantar gānu ‘metamorphose’; ghumnu, phirnu, ultjaunu, pharkaunu ‘Turn’; etc.

18.6 Performance Verbs

Class Members: prarthana gānu ‘chant prayer’; nacaṇa sikaunu ‘choreograph dance’; khelnu ‘play’; chitra kornu ‘draw picture’; git badʒaunu ‘play music’; khelnu ‘game’; abritti gānu ‘recite poem’; git gaunu ‘sing song’; photo kicnu ‘take picture’; suselnu ‘whistle tune’; kitab lekhnu ‘write book’; etc.

19 Engender Verbs

Class Members: paunu, utpanna gānu ‘beget/generate’; parnu, magnu, garaunu ‘cause’; sirdʒana gānu, banaunu ‘create’; dʒammaḥ dinu ‘engender’; rup/akar dinu ‘shape’; etc.

20 Calve Verbs

Class Members: bjaunu ‘calve/cub/fawn/foal/kitten/lamb/litter/pup/whelp’; dʒanmaunu ‘spawn’; etc.

21 Verbs with Predicative Complements

21.1 Appoint Verbs

Class Members: mannu, sakarnu ‘acknowledge’; kokh linu, god linu ‘adopt’; nijukta garnu ‘appoint’; thaharaunu ‘consider’; gunnu, thannu ‘deem’; toknu, kiṭnu, manonajan garnu ‘designate’; cunnu, cajan garnu ‘elect’; mannu, pudḡnu, adar garnu ‘esteem’; kalpana garnu, socnu ‘imagine’; cinhit garnu, cinu launu ‘mark’; manonit garnu ‘nominate’; dar toknu ‘rate’; gannu ‘reckon’; chahanu ‘want’; etc.

21.2 Characterize Verbs

Class Members: svikar garnu ‘accept’; sambodhan garnu ‘address’; prasansa garnu, kadar garnu ‘appreciate’; pramanit garnu ‘certify’; cunnu ‘choose’; ullekh garnu ‘cite’; bāṭnu, vargikaran garnu ‘classify’; puṣṭi garnu ‘confirm’; gannu ‘count’; paribhasit garnu ‘define’; vjakhja garnu ‘describe’; khuttjaunu ‘diagnose’; nijukta garnu ‘employ’; pasnu, bhitriu ‘enter’; samana garnu ‘envisage’; isthapit garnu ‘establish’; ghosana garnu ‘herald’; bharāma linu ‘hire’; samman garnu ‘honor’; cinnu ‘identify’; kalpana garnu ‘imagine’; samaves garnu ‘incorporate’; citaunu ‘intend’; caṭhaunu, ṭakraunu ‘offer’; virodh garnu ‘oppose’; prasansa garnu ‘praise’; etc.

21.3 Dub Verbs

Class Members: abhisek garnu ‘anoint’; baptisma linu ‘baptize’; bolaunu ‘call’; pratistha garnu ‘consecrate’; taz/mukuṭ pahirjaunu ‘crown’; agja dinu ‘decree’; man dinu ‘dub’; nam dinu ‘label’; banaunu ‘make’; nam dinu ‘name’; upanam dinu ‘nickname’; uccaran garnu ‘pronounce’; sasan garnu ‘rule’; chap launu ‘stamp’; etc.

3.4.21.4. Declare Verbs

Class Members: anuman garnu ‘assume/judge’; svikar garnu ‘avow’; bisvas garnu ‘believe’; sakarnu ‘confess’; ghosana garnu ‘declare’; bheṭnu, phela parnu, patto launu ‘find’; thannu ‘presume’; pratigja garnu ‘profess’; pramanit garnu ‘prove’; ṭkal launu ‘suppose’; socnu ‘think’; etc.

21.5 Conjecture Verbs

Class Members: svikar garnu ‘admit’; mannu, agja dinu ‘allow’; kiṭnu ‘assert’; anuman launu ‘conjecture’; asvikar garnu, radda garnu, namannu ‘deny’; aviskar garnu ‘discover’; anubhāṇ/bodh garnu ‘feel’; ṭhamjaunu ‘figure’; mandḡur garnu ‘grant’; āknu, ṭkal kaṭnu ‘guess’; mannu ‘hold’; zannu, thaha paunu ‘know’; rakhnu, palnu ‘maintain’; arthjaunu ‘mean’; dekhnu, avalokan garnu ‘observe’; cinnu, ṭhamjaunu ‘recognize’; gannu ‘repute’; dekhaunu ‘show’; sankā garnu ‘suspect’; etc.

21.6 Masquerade Verbs

Class Members: kam garnu ‘act’; vjavahar garnu ‘behave’; bhes badalnu ‘camouflage’; gannu ‘count’; svang parnu ‘masquerade’; rup linu, bannu ‘pose’; etc.

21.7 Orphan Verbs

Class Members: langaṭaunu ‘cripple’; sahid hunu ‘martyr’; tuhuro hunu ‘orphan’; vidhava hunu ‘widow’; etc.

21.8 Captain Verbs

Class Members: hāikam calaunu ‘boss’; sataunu ‘bully’; chapkaunu ‘butcher’; kaptan hunu ‘captain’; raksha garnu ‘guard’; goṭhala garnu ‘shepherd’; gavahi dinu ‘witness’; etc.

22 Verbs of Perception

22.1 See Verbs

Class Members: patto launu ‘detect’; dekhnu, budzhnu ‘discern’; bodh/anubhāv garnu ‘feel’; sunnu ‘hear’; dekhnu ‘notice’; dekhnu, hernu ‘see’; džannu, thaha paunu ‘sense’; smell; cakhnu, svad linu ‘taste’; etc.

22.2 Sight Verbs

Class Members: dekhnu ‘descry’; patto launu ‘discover’; džācnu ‘examine/inspect/investigate’; hernu ‘eye’; ṭipnu ‘note’; avālokān garnu ‘observe’; ṭhamjaunu, bujhnu ‘perceive’; cinnu ‘recognize’; mannu ‘regard’; svad dinu ‘savor’; niarnu, džācnu ‘scrutinize’; dekhnu ‘sight’; džasusi garnu, pahiljaunu ‘spy’; paṭhnu, adhjaṇ garnu ‘study’; sarvekshan garnu ‘survey’; hernu ‘view’; hernu ‘watch’; gavahi banu ‘witness’; etc.

22.3 Peer Verbs

Class Members: džācnu ‘check on’; tolaunu ‘gape’; nijalnu ‘gaze’; džhalakkā hernu ‘glance’; ā kha taru ‘glare’; lolaunu ‘goggle’; karke ā nkha launu ‘leer’; sunnu ‘listen to’; hernu ‘look’; ā nkha launu ‘ogle’; cijaunu ‘peek/peep/snoop’; nijalnu ‘peer’; sungnu ‘sniff’; ṭuluṭulu hernu ‘stare’; etc.

22.4 Stimulus Subject Perception Verbs

Class Members: anubhāv/bodh garnu ‘feel’; hernu ‘look’; sunnu ‘smell’; sunnu ‘sound’; cakhnu ‘taste’; etc.

23 Psych-Verbs Verbs of Psychological State

23.1 Amuse Verbs

Class Members: laḍḍaunu ‘abash’; prabhav parnu ‘affect’; dukha dinu ‘afflict’; apaman garnu ‘affront’; capnu ‘aggravate’; birodh garnu ‘agitate’; dukha dinu ‘agonize’; cheṭavni dinu ‘alarm’; chakkā parnu ‘amaze’; ramaunu ‘amuse’; risaunu ‘anger’; džiskjaunu ‘annoy’; tarṣaunu ‘appall’; phuljaunu ‘appease’; džagaunu ‘arouse’; santa garnu ‘assuage’; cakit parnu ‘astonish’; bismit parnu ‘astound’; haraunu ‘baffle’; bahakānu, phusjaunu ‘beguile’; džilla parnu ‘bewilder’; mukdha/makkha parnu ‘bewitch’; džhaskānu, tarṣānu ‘boggle’; alchi lagnu, ‘bore’; džhizho mannu ‘bother’; hāiran parnu ‘bug’; samja/santa parnu ‘calm’; lobhjaunu, mohit parnu ‘captivat/charm’; prerna/hausala dinu ‘cheer’; santavana dinu ‘comfort/console’; lagau hunu ‘concern’; akamakkā parnu ‘confound’; almalinu ‘confuse’; citta budzhaunu ‘content’; patjar parnu ‘convince’; laṭṭha parnu ‘daze’; tirmiraunu ‘dazzle’; dikka mannu ‘deject’; niras parnu ‘disappoint’; etc.

23.2 Admire Verbs

Class Members: PoSITIVE VERBS: adar garnu ‘admire’; upasana garnu ‘adore’; prasansa garnu ‘appreciate’; kadar garnu ‘cherish’; ramaunu ‘enjoy’; mannu ‘esteem’; vibhor hunu ‘exalt’; lahasinu ‘fancy’; kripa garnu ‘favor’; pudza garnu ‘idolize’; man paraunu ‘like’; maja/prem garnu ‘love’; samdzhanu, jad garnu ‘miss’; inam dinu ‘prize’; adar/samman garnu ‘respect’; samarthan garnu, saghaunu ‘support’; sahanu ‘tolerate’; sa ~cnu ‘treasure’; patjaunu, bisvas garnu ‘trust’; mahatva dinu ‘value’; pudznu ‘worship’; etc.

NEGATIVE VERBS: ghinaunu, ghrina garnu ‘abhor/detest’; vilap garnu ‘deplore’; hela garnu ‘despise/disdain’; man naparaunu ‘dislike’; napatjaunu, bisvas nagarnu ‘distrust’; tarsanu, dāraunu ‘dread’; dāha garnu ‘envy’; sarapnu ‘execrate’; dāraunu ‘fear’; ghrina garnu ‘hate’; vilap garnu ‘lament’; sok garnu ‘mourn’; dāja garnu ‘pity’; pachuto/pascatap garnu ‘regret’; etc.

23.3 Marvel Verbs

Class Members: dzhidzho launu ‘bother’; khjal rakhnu ‘care’; hercah garnu, rekhdekh garnu, hos garnu ‘mind’; rankanu ‘rage’; khusi hunu ‘rejoice’; bhattjaunu ‘rhapsodize’; surtaunu ‘worry’; tamsanu ‘enthuse’; ra ~kinu ‘fume’; khusi parnu ‘gladden’; pagal banaunu ‘madden’; cakit hunu ‘marvel’; dukhi hunu ‘sadden’; bimar hunu ‘sicken’; murcha parnu ‘swoon’; pulakit hunu ‘thrill’; chakka parnu ‘wonder’; runu ‘cry’; dāraunu ‘fear’; bodh garnu ‘feel’; sok garnu ‘grieve/mourn’; runu ‘weep’; dukhnu ‘hurt’; bhognu ‘suffer’; ananda mannu ‘glory’; modz garnu ‘luxuriate’; ramaunu ‘revel’; svikar garnu ‘approve’; sacet hunu ‘beware’; thaknu ‘tire’; risaunu ‘anger’; bhokaunu ‘hunger’; dzhukkjaunu ‘puzzle’; thuk aunu, ral kachnu ‘salivate’; surtaunu ‘sorrow’; pratikrija dekhaunu ‘react’; etc.

23.4 Appeal Verbs

Class Members: abhari/kritagja hunu ‘grate’; bazhnu ‘jar’; prarthana garnu ‘appeal’; etc.

24 Verbs of Desire

24.1 Want Verbs

Class Members: lobh garnu ‘covet’; magnu ‘crave’; iccha garnu ‘desire’; rahar garnu ‘fancy’; kha co/jaruri/avasjak hunu ‘need’; चाहानु ‘want’; etc.

24.2 Long Verbs

Class Members: lobhinu ‘dangle’; cukukaunu ‘hanker’; vasana zagnu ‘lust’; tirkhaunu ‘thirst’; rahar garnu ‘yearn’; etc.

25 Judgment Verbs

Class Members: PosiTIVE VERBS: sanjog parnu ‘acclaim’; prasansa garnu ‘applaud/commend/compliment/praise’; asis/asirvad dinu ‘bless’; utsav manaunu ‘celebrate’; chatipurti garnu ‘compensate’; badhai dinu ‘congratulate’; bahana/nihu ~ banaunu ‘excuse’; abhinandan garnu ‘felicitate’; chama/maph garnu ‘forgive/pardon’; abhivadan garnu ‘greet’; samman garnu ‘honor’; sahraunu, sarahana garnu ‘laud’;

NEGATIVE VERBS: gali gāru ‘abuse’; kura kaṭnu ‘backbite’; ninda gāru ‘calumniate’; haṇkaunu ‘castigate’; d̥ʒhaṛkanu ‘censure’; kharnu ‘chasten’; daṇḍa dinu ‘chastise’; uḍʒur gāru ‘chide’; ninda gāru ‘condemn’; khisi gāru ‘criticize’; hoċjaunu ‘decry’; baḍnam gāru ‘defame’; hiċjaunu ‘denigrate’; baḍkhvai~ gāru ‘denounce’; khilli uḍaunu ‘deride’; dos dinu ‘fault’; ḍa~ṭnu, ḍʒarimana tiraunu ‘fine’; abhiċog/bat launu ‘impeach’; araṃan gāru ‘insult’; malign; giċjaunu ‘mock’; sasti dinu ‘persecute’; mudda halnu ‘prosecute’; sazaḷa dinu ‘punish’; d̥ʒhaṇarnu ‘rebuke’; bhaṭsaṛna gāru ‘reproach’; gali gāru ‘scold’; hela/upekcha/tiraṣkar gāru ‘scorn’; saṛam/lad̥ʒ parnu ‘shame’; hepnu ‘snub’; sikar baṇaunu ‘victimize’; hoċjaunu ‘vilify’; etc.

Class Members: vislesan garnu ‘analyze’; nirupan garnu ‘assess/review’; lekhađzokha garnu/ dʒɑ̃ch garnu ‘audit’; muljaŋkan garnu ‘evaluate’; dʒɑ̃cnu ‘scrutinize’; ʌdhjaŋjan garnu ‘study’; etc.

27.1 Hunt Verbs

27.2 Search Verbs

27.3 Stalk Verbs

27.4 Investigate Verbs

27.5 Rummage Verbs

27.6 Ferret Verbs

120

28 Verbs of Social Interaction

28.1 Correspond Verbs

Class Members: mannu, sahamat hunu ‘agree’; tarka garnu ‘argue’; upahas garnu ‘banter’; mol garnu ‘bargain’; mukh lagnu ‘bicker’; ba dzhnu ‘brawl’; bhiṛnu ‘clash’; sath dinu ‘collaborate’; ṭakkraunu ‘collide’; daja garnu ‘commiserate’; khabar dinu ‘communicate’; ispardha garnu ‘compete’; sahamat hunu ‘concur’; kurakani garnu ‘confabulate’; dzhagaṛa garnu ‘conflict’; sahajog garnu ‘cooperate’; patrachar garnu ‘correspond’; bhinna hunu ‘differ’; asahamat hunu ‘disagree’; virodh garnu ‘dispute’; poila zanu ‘elope’; thattā garnu ‘joke’; djudhnu ‘joust’; dzoṛa milaunu ‘mate’; misinu ‘mingle/mix’; āgalo halnu ‘neck’; milaunu ‘negotiate’; dzoṛa milaunu ‘pair’; sangharsa garnu ‘struggle’; etc.

28.2 Marry Verbs

Class Members: prem garnu ‘court’; lutuputu hunu, pjaro garnu ‘cuddle’; parpacuke garnu ‘divorce’; āgalo halnu ‘embrace/hug’; mvaī khanu ‘kiss’; vivah garnu ‘marry’; guṭumuṭu hunu ‘nuzzle’; etc.

28.3 Meet Verbs

Class Members: djudhnu ‘battle’; mukka khelnu ‘box’; sallaha garnu ‘consult’; tarka garnu ‘debate’; larāi garnu ‘fight’; bhetnu ‘meet’; khelnu ‘play’; etc.

29 Verbs of Communication

29.1 Verbs of Transfer of a Message

Class Members: sodhnu ‘ask’; ullekh garnu ‘cite’; darsaunu ‘demonstrate’; lekhaunu ‘dictate’; vjakhja garnu ‘explain’; vistar garnu ‘explicate’; bhannu, varnan garnu ‘narrate’; upades dinu ‘preach’; uddhrit garnu ‘quote’; paṛhnu ‘read’; abritti garnu ‘recite’; dekhaunu ‘show’; paṛhaunu ‘teach’; bhannu ‘tell’; lekhnu ‘write’; etc.

29.2 Tell Verb

Class Members: bhannu, dḷanaunu, bataunu ‘tell’;

29.3 Verbs of Manner of Speaking

Class Members: barbaraunu ‘babble’; bhuknu ‘bark’; karaunu, hakarnu, dzhāpnu, cjaṭinu ‘bawl’; kurlanu, karṭkanu ‘bellow/roar’; bolaunu, pukarnu, dāknu ‘call’; phalaknu ‘chant’; phatphataunu ‘chatter/grumble’; basnu ‘cluck’; gungunaunu ‘croon’; runu ‘cry’; legro tannu ‘drawl’; bhunbhunaunu ‘drone’; groan; dzharkanu, huṅkar garnu ‘growl’; sitkar garnu ‘hiss’; cicjaunu ‘holler’; dhikkarnu ‘hoot’; tote bolnu ‘lisp’; kannu ‘moan’; murmurinu ‘murmur’; gangan garnu ‘mutter’; ghurnu ‘purr’; rankanu ‘rage’; cicjaunu, citkar garnu ‘scream’; karaunu ‘shout’; gaunu ‘sing’; dzhamtanu ‘snap’; bhakbhakaunu ‘stammer’; sauti/kanekhusi garnu ‘whisper’; suselnu ‘whistle’; etc.

29.4 Verbs of Instrument of Communication

Class Members: mel garnu ‘e-mail’; phon garnu ‘phone’; etc.

29.5 Talk Verbs

Class Members: bolnu ‘speak’; kura garnu, bat marnu ‘talk’; etc.

29.6 Chitchat Verbs

Class Members: badzhnu ‘argue’; gaph garnu ‘chat’; chatter; bakhak garnu ‘chitchat’; vartalap garnu ‘converse’; gaph cutnu ‘gossip’; etc.

29.7 Say Verbs

Class Members: ghosana garnu ‘announce/declare’; uccaran garnu ‘articulate’; pvakka bolnu ‘blurt’; dabi garnu ‘claim’; sakarnu ‘confess’; ullekh garnu ‘mention’; tipnu ‘note’; prastav rakhnu ‘propose’; bakhhan garnu ‘recount’; dohorjaunu ‘reiterate/repeat’; varnan garnu ‘relate’; hernu ‘remark’; prakat garnu ‘reveal’; bhannu ‘say/state’; suzhav dinu ‘suggest’; etc.

29.8 Complain Verbs

Class Members: sarkanu ‘boast’; udzur/gunaso garnu ‘complain’; thuskinu ‘grouch’; phatphaatunu ‘grouse’; apatti dzanaunu, virodh garnu ‘object’; etc.

29.9 Advise Verbs

Class Members: cetauni dinu ‘admonish’; sujhau dinu ‘advise’; sacet garaunu ‘alert’; savdhan garaunu ‘caution’; paramarsa dinu ‘counsel’; cetaunu ‘warn’; etc.

30 Verbs of Sounds Made by Animals

Class Members: bhuknu ‘bark’; dukranu ‘bellow’; cirbiraunu ‘chirp’; bhunbhunaunu ‘buzz’; basnu ‘cackle’; ghurnu ‘coo’; huṅkar garnu ‘growl’; mjau garnu ‘mew’; ba garnu ‘moo’; hinhinaunu ‘neigh’; cjačja garnu ‘peep’; etc.

31 Verbs of Ingesting

31.1 Eat Verbs

Class Members: khanu ‘drink/eat’ only

31.2 Chew Verbs

Class Members: capaunu/cabaunu ‘chew/munch’; catnu ‘lick’; toknu ‘nibble’; tipnu ‘pick’; thuṅnu ‘peck’; ghuṭkjaunu ‘sip’; suruppa khanu ‘slurp’; cusnu ‘suck’; etc.

31.3 Gobble Verbs

Class Members: nilnu ‘gobble/swallow’; bukjaunu ‘gulp’; etc.

31.4 Devour Verbs

Class Members: upabhog garnu ‘consume’; ghicnu, hasurnu ‘devour’; dhoknu ‘imbibe/swill’; nilnu ‘ingest’; etc.

31.5 Dine Verbs

Class Members: khaza khanu ‘breakfast’; bhozan garnu, khana khanu ‘dine/lunch’; bhoz khanu ‘feast’; carnu ‘graze’; vanbhoz khanu ‘picnic’; cuski linu ‘sup’; etc.

31.6 Gorge Verbs

Class Members: khuvaunu ‘feed’; phastanu ‘flourish’; peṭ bharnu ‘gorge’; ziunu, bacnu ‘live/survive’; saprinu, phapnu ‘prosper’; maulaunu ‘thrive’; etc.

31.7 Verbs of Feeding

Class Members: dudh khuvaunu ‘breastfeed’; khuvaunu ‘feed’; etc.

32 Verbs Involving the Body

32.1 Verbs of Bodily Processes

Hiccup Verbs

Class Members: dʌkarnu ‘belch/burp’; rattinu ‘blush’; hʌɽbʌɽaunu ‘flush’; baɽuli lagnu ‘hiccup’; kʌpnu ‘pant’; hachiũ ɡʌrnu ‘sneeze’; sũɡnu ‘sniffle’; ghurnu ‘snore’; nilnu ‘swallow’; hai kaɽhnu ‘yawn’; etc.

Breathe Verbs

Class Members: ɾʌɡʌt aunu ‘bleed’; sas phernu ‘breathe’; khoknu ‘cough’; runu ‘cry’; ɾʌl cuhaunu ‘dribble’; vanta ɡʌrnu ‘puke’; thuknu ‘spit’; ɾʌsina aunu ‘sweat’; ulɽi ɡʌrnu ‘vomit’; etc.

Exhale Verbs

Class Members: sas phernu ‘exhale/inhale’; ɾʌsina aunu ‘perspire’; etc.

32.2 Verbs of Nonverbal Expression

Class Members: muskaunu ‘beam’; kurnu ‘cackle’; khitka chaɽnu ‘chortle’; khisikka hasnu ‘chuckle’; khoknu ‘cough’; runu ‘cry’; ʌkha tʌrnu ‘frown/glare’; dhʌkaunu, ʌtalinu ‘gasp’; hʌsnu ‘giggle/laugh’; dʒhʌrkʌnu, huŋkar ɡʌrnu ‘growl’; citkar ɡʌrnu ‘howl’; gidʒjaunu ‘jeer’; suksukaunu ‘moan’; oɽh lebraunu ‘pout’; ucchʌvas linu ‘sigh’; mʌskinu ‘simper’; muskuraunu ‘smile’; ghurnu ‘snore’; suselnu ‘whistle’; hai kaɽhnu ‘yawn’; etc.

32.3 Verbs of Gestures/Signs Involving Body Parts

Wink Verbs

Class Members: ʌnkha dʒhimkjaunu ‘wink eye/blink eye’; tali bʌdʒaunu ‘clap hands’; ɽʌuko nihurjaunu ‘nod head’; ʌũla ɽhʌɽjaunu ‘point finger’; kum hʌllaunu ‘shrug shoulders’; dʒɛdʒhe ʌnkhole hernu ‘squint eyes’; hat hʌllaunu ‘wave hand’; etc.

Crane Verbs

Class Members: chillinu ‘arch back’; ʌɡalo halnu ‘neck’; dʌt dekhaunu ‘bare teeth’; nak phulaunu ‘blow nose’; muɽɽhi kʌsnu ‘clench fists’; ʌula doɽɽjaunu ‘crook finger’; hat bʌdhnu ‘cross arms/fold arms’; ɾʌkheɽa ɕʌʌaunu ‘flap wings’; dʌt tʌlkaunu ‘flash teeth’; cuɽki bʌdʒaunu ‘flick finger’; dʌt kiɽnu ‘gnash teeth/grind teeth’; ʌnkha kholnu ‘open eyes’; hat mʌlnu ‘rub hands’; ɽʌuko hʌllaunu ‘shake head’; etc.

Curtsey Verbs

Class Members: dʒhuknu, nihurinu ‘bow’; ghũdʌ ɽeknu ‘kneel’; salam ɡʌrnu ‘salaam/salute’ etc.

32.4 Snooze Verbs

Class Members: ũghnu, dʒhulnu, dʒhʌkaunu ‘doze’; lolaunu ‘drowse’; nidaunu, sutnu ‘nap/sleep/snooze’; etc.

32.5 Flinch Verbs

Class Members: hissʌ/hisrikka ɾʌrnu ‘balk’; ʌɽɽʌnu ‘cower’; dʌbnu ‘cringe’; hʌckʌnu ‘flinch’; dʒhʌskʌnu ‘recoil’; khumcinu ‘shrink’; kʌnnu ‘wince’; etc.

32.6 Verbs of Body-Internal States of Existence

Class Members: kãpnu, kamnu ‘convulse/quake/shiver’; lãtrãnu ‘cower’; hãllãnu ‘quiver/shake’; cãtpãtãunu ‘writhe’; etc.

32.7 Suffocate Verbs

Class Members: sãrkãnu ‘choke’; dũbnu ‘drown’; nissasinu ‘suffocate’; etc.

32.8 Verbs of Bodily State and Damage to the Body

Pain Verbs

Class Members: dukhnu, piṭã hunu ‘ache/pain’; bother; coṭ lagnu, biḍhnu ‘hurt’; cilaunu ‘itch’; etc.

Tingle Verbs

Class Members: gungunaunu ‘hum’; ghopnu ‘prickle’; riṅṭã lagnu ‘head reel’; ṭhuṅnu ‘sting’; cãskãnu, ṭãnkãnu ‘throb’; ḍḡhãmdḡhãmaunu, polnu, cãhrjãunu, pãrpãraunu, phãrpãhãraunu, rãmrãmaunu, ḍḡhãndḡhãnaunu ‘tingle’; etc.

Hurt Verbs

Class Members: oṭh ṭoknu ‘bite lip’; thãccinu, thãcãrnu ‘bump’; polnu, zãlnu ‘burn’; toṛnu, phuṭãunu ‘break’; pãlhar gãrn ‘bruise’; kãṭnu ‘cut’; dukhaunu, coṭ pãrnu ‘hurt/injure’; phuṭnu, bhãccinu ‘rupture’; polnu, ḍãmnu ‘scald’; etc.

Verbs of Change of Bodily State

Class Members: kalonilo hunu ‘blanch’; murchã pãrnu, ãcet hunu, behos hunu ‘faint/swoon’; rogaunu ‘sicken’; etc.

33 Verbs of Grooming and Bodily Care

33.1 Verbs of Caring for the Whole Body

Dress Verbs

Class Members: nuhaunu ‘bathe/shower’; phernu ‘change’; luga phukãlnu ‘disrobe’; pãlhirinu, launu ‘dress’; vjãjam gãrnu ‘exercise’; ṭimkinu ‘preen’; khãurãnu ‘shave’; udhãrnu ‘strip’; dhunu, pãkhalnu ‘wash’; etc.

Groom Verbs

Class Members: tãjar pãrnu, sīgãrnu, sãḍḡãunu ‘groom’; etc.

33.2 Verbs of Caring for a Specific Body Part

Floss Verbs

Class Members: brus gãrnu ‘brush teeth’; darhi khãurãnu ‘shave beard’; hat dhunu ‘wash hands’; etc.

Braid Verbs

Class Members: kes bãṭnu ‘braid hair’; kes kornu ‘comb hair’; kes kãṭnu ‘cut hair’; kes dhunu/pãkhalnu ‘rinse hair’; kes milãunu ‘set hair’; mukh puchnu ‘towel face’; kes chãṭnu ‘trim hair’; etc.

33.3 Verbs of Dressing

Simple Verbs of Dressing

Class Members: launu, pãlhirinu ‘don/wear’; phukãlnu, utãrnu ‘doff’; etc.

Verbs of Dressing Well

Class Members: sīgãrinu ‘doll’; pãlhirinu ‘dress’; ṭhãṭinu ‘spruce’; etc.

Verbs of Being Dressed

Class Members: pāhiraunu/launu ‘attire/garb’; etc.

34 Verbs of Killing

34.1 Murder Verbs

Class Members: hatja gānu ‘assassinate’; chapkaunu ‘butcher’; etc.

34.2 Poison Verbs

Class Members: dūbaunu, copānu, cobānu ‘drown’; dzhundjaunu ‘hang’; kaṭnu ‘knife’; bis dinu ‘poison’; goli hannu ‘shoot’; nissasjaunu ‘smother/suffocate’; ghopnu, cheṭnu ‘stab’; āṭhjaunu ‘strangle’; etc.

35 Verbs of Emission

35.1 Verbs of Light Emission

Class Members: prakasit hunu ‘beam’; dzhimkinu ‘blink’; balnu ‘blaze’; djalnu ‘flame’; balnu ‘flare’; ṭalkānu ‘flash/glare’; pilpilaunu ‘flicker’; ṭalpāl gānu ‘glimmer’; camkinu ‘glint/glisten’; ṭalkinu ‘glitter/glow’; dzagmagānu ‘scintillate’; ṭalkānu, camkānu, dzhalkānu ‘shine/sparkle’; camcamānu ‘twinkle’; etc.

35.2 Verbs of Sound Emission

Class Members: barbaraunu ‘babble’; kurlānu ‘bellow’; nad phuknu ‘blare’; visphoṭ hunu ‘blast’; carkinu ‘blat’; umlinu ‘bubble’; bhankinu ‘buzz’; phatphataunu ‘chatter’; ṭaṇṭānu ‘clang’; tali badjaunu ‘clap’; paṭkinu ‘crack’; kaṭkinu ‘crackle’; ṭakkraunu ‘crash’; cuṭkinu ‘creak’; karaunu ‘cry’; bhunbhunaunu ‘hum’; cicjaunu ‘jangle’; dzhandzhānu ‘jingle’; ghurnu ‘purr’; dhalakjaunu ‘knock’; kānu ‘moan’; murmurinu ‘murmur’; paṭpaṭaunu ‘patter’; ghankānu ‘plunk’; gurgurānu ‘rattle’; etc.

35.3 Verbs of Smell Emission

Class Members: sūgnu ‘smell’; gānhaunu, gandha aunu ‘stink’; etc.

35.4 Verbs of Substance Emission

Class Members: dākānu ‘belch’; ragat aunu ‘bleed’; umlinu ‘bubble’; cuhunu ‘leak/dribble’; tapkānu ‘drip’; cuhunu ‘exude’; phīdž aunu ‘foam’; bagnu ‘gush’; rasaunu ‘ooze’; khaṇjaunu ‘pour’; phulnu ‘puff’; khasalnu, dzharnu ‘shed’; okānu, phalaknu ‘spout’; umrinu ‘sprout’; bhulbhulaunu ‘spurt’; chjapnu ‘squirt’; baphinu ‘steam’; pasina kaṭhnu ‘sweat’; etc.

36 Destroy Verbs

Class Members: dhvasta parnu ‘annihilate’; sotta parnu ‘decimate’; vidhvans parnu, bhātkaunu, nas parnu ‘demolish/destroy/devastate/wreck’; masnu ‘exterminate’; unmul parnu ‘extirpate’; meṭnu ‘obliterate’; taḥasnaḥas parnu ‘ravage’; meṭaunu, kornu ‘raze’; svaha parnu ‘ruin’; bigānu, phuknu, hūṭelnu, uṭaunu, sidhjaunu, khiinu, suknu, bitnu, sakinu, aphalnu, khera djanu/phalnu ‘waste’; etc.

‘remain’; basnu ‘reside’; maulaunu, phas̄taunu ‘thrive’; parkhanu, pratikcha garnu, kurnu ‘wait’; etc.

39.2 Verbs of Entity-Specific Modes of Being

Class Members: phakrinu ‘bloom’; phulnu ‘blossom’; phuknu ‘blow’; sas phernu ‘breathe’; d̄arkharaunu ‘bristle’; suninu, ganinu ‘bulge’; khiinu ‘corrode/erode’; kuhunu ‘decay’; makkinu ‘decompose’; umlanu ‘effervesce’; paknu, sar̄nu ‘fester’; ph̄iz ūthnu ‘fizz’; bagnu, bah̄anu ‘flow’; phulnu ‘flower’; phoka ūthnu ‘foam’; umrinu ‘germinate’; bar̄thnu ‘grow’; galnu ‘molt’; ph̄izaunu, ph̄ailaunu ‘propagate’; urlanu ‘rage’; dhamilinū ‘roil’; sar̄nu ‘rot’; khija lagnu ‘rust’; paknu ‘seethe’; ph̄ailanu, ph̄id̄zinu ‘spread’; umrinu, t̄usaunu ‘sprout’; d̄amnu ‘stagnate’; bagnu ‘stream’; bar̄arnu ‘sweep’; etc.

39.3 Verbs of Modes of Being Involving Motion

Class Members: katranu ‘bob’; nihurinu, d̄huknu ‘bow’; sarnu, ghas̄ranu ‘creep’; nacnu ‘dance’; bagnu, thuprinu ‘drift’; tairanu ‘float’; pakhēta calaunu ‘flutter’; hallanu ‘hover’; urlanu, d̄har̄kanu ‘pulsate’; hallanu ‘quake’; kamnu ‘quiver’; ghumnu ‘revolve/rotate’; hallanu ‘shake’; calaunu ‘stir’; d̄agmagaunu, d̄hulnu ‘sway’; larkharaunu ‘teeter’; balkanu ‘throb’; d̄ōrnu ‘totter’; tar̄anginu ‘undulate’; d̄hankanu, ghankanu ‘vibrate’; bagnu ‘waft’; ph̄aharaunu, ph̄arpharaunu ‘wave’; hallaunu ‘wiggle’; d̄hulnu ‘wobble’; hat̄pataunu ‘writhe’; etc.

39.4 Verbs of Sound Existence

Class Members: ghankanu, gund̄zanu ‘echo/resonate’; etc.

39.5 Verbs of Group Existence

Swarm Verbs

Class Members: umlanu ‘abound’; calmalinu ‘bustle’; bame sarnu, ghas̄ranu ‘crawl’; uphranu ‘hop’; kudnu ‘run’; cal̄thnu ‘swarm’; plūtanu, tairanu ‘swim’; d̄huttinu, d̄humminu ‘teem’; d̄amma garnu ‘throng’; etc.

Herd Verbs

Class Members: d̄ōrnu, s̄galnu ‘accumulate’; d̄amma garnu ‘aggregate’; san̄graha garnu ‘amass/collect’; d̄ūrnu, d̄ōrnu ‘assemble’; d̄humminu, d̄huttinu ‘cluster’; san̄jod̄zan garnu ‘convene’; thuprinu ‘huddle’; etc.

Bulge Verbs

Class Members: d̄arkharaunu ‘bristle’; suninu, ganinu ‘bulge’; etc.

39.6 Verbs of Spatial Configuration

Class Members: talanu ‘balance’; mōrinu ‘bend’; d̄huknu, nihurinu ‘bow/crouch’; d̄hulaunu ‘dangle’; hallanu ‘flop’; ūrnu ‘fly’; d̄hund̄inu ‘hang’; hallanu ‘hover’; gh̄ūta teknu ‘kneel’; d̄hessinu ‘lean’; d̄halkanu ‘lie’; ghus̄ranu ‘nestle’; khulnu ‘open’; utranu ‘perch’; bisaunu ‘rest’; udaunu ‘rise’; basnu ‘sit’; lat̄rinu ‘slouch’; lat̄nu ‘sprawl’; ubhinu ‘stand’; d̄har̄nu ‘stoop’; maccinu ‘swing’; tersinu ‘tilt’; etc.

39.7 Meander Verbs

Class Members: cal̄thnu ‘climb’; ghas̄ranu ‘crawl’; kātnu ‘cut’; d̄harnu ‘drop’; d̄zanu ‘go’; ghumnu ‘meander’; khasnu ‘plunge’; kudnu ‘run’; charinu ‘straggle’; tankanu

‘stretch’; gulṭinu ‘tumble’; baṭarinu ‘twist’; dṭlṇu, ṭaḥaḷṇu ‘wander’; bunnṇu, baṭṇu ‘weave’; etc.

39.8 Verbs of Contiguous Location

Class Members: dʒoṭṇu, dʒoṭinu, dʒuṭṇu ‘abut/adjoin’; badhṇu ‘bound’; thamṇu, ṭaṭaṇu ‘contain’; dḥakṇu, chopṇu ‘cover’; taṛṇu ‘cross’; kṭḍjaṇu ‘dominate’; ghernu ‘encircle’; chekṇu ‘fence’; bharnu ‘fill’; paḥjaṇu ‘follow’; gaṭṇu ‘frame’; hannu, ṭhokṇu ‘hit’; ṭagalo marnu ‘hug’; kaṭṇu ‘intersect’; milṇu ‘meet’; ucchinṇu ‘precede’; baṭṇu ‘surmount’; ghernu ‘surround’; chunu ‘touch’; etc.

40 Verbs of Appearance; Disappearance; and Occurrence

40.1 Verbs of Appearance

Appear Verbs

Class Members: dekhinu, dekha paṛṇu, udaṇu ‘appear’; uṭṇu, udaṇu, phurnu ‘arise’; uṭṇu, zagnu, biṭḍḥaṇu ‘awake’; biṭḍḥaṇu, dʒaḡaṇu ‘awaken’; toṭṇu, phuṭaṇu, bhācṇu ‘break’; phuṭṇu, paṭṇu ‘burst’; aṇu, paṇu ‘come’; paṇu, utpaṭṭi hunu ‘derive’; phaṭṭaṇu, maṭṭaṇu, vikaṣ gaṛṇu ‘develop’; niṣkaṇu, phuṭṇu ‘erupt’; baḡṇu, baḥaṇu ‘flow’; baṇaṇu ‘form’; baṭṇu, sapraṇu, umraṇu ‘grow’; niṣkaṇu ‘issue’; khulṇu, kholṇu ‘open’; phaḷaṇu, phīzinu ‘spread’; etc.

Reflexive Verbs of Appearance

Class Members: bhannu, kiṭṇu ‘assert’; ghosaṇa gaṛṇu ‘declare’; paṛibhasit gaṛṇu ‘define’; vjaḡṭa gaṛṇu ‘express’; siddha/paṛmanit gaṛṇu ‘manifest’; caṭṭhaṇu, ṭaḡjaṇu ‘offer’; praṭṭu gaṛṇu ‘present’; aṭṭaṇu ‘proffer’; siphariṣ gaṛṇu ‘recommend’; akaṛ diṇu ‘shape’; dekhaṇu, kholṇu ‘show’; sujhaṇu diṇu ‘suggest’; etc.

40.2 Verbs of Disappearance

Class Members: maṛṇu, khaṣṇu, paṛlok hunu, mritju hunu, bitṇu ‘die/expire’; haṛaṇu, aṭṭinu, bilaṇu, lop hunu, viliṇu, aḍriṣṭa hunu ‘disappear/vanish’; khariz hunu, guḍḡaṇu, giriṇu ‘lapse’; naṣ hunu ‘perish’; etc.

40.3 Verbs of Occurrence

Class Members: hunu, ghaṭṇu ‘eventuate/happen’; hunu, painu, raḥaṇu, ghaṭṇu, paṛṇu ‘occur’; etc.

41 Verbs of Body-Internal Motion

Class Members: uphraṇu, laṭṇu ‘buck’; caḡcaḡ gaṛṇu, caṭṭaṭaṇu, caḡmaḡaṇu, caḡṇu ‘fidget’; phukṇu ‘flap’; ghumnṇu, caḡkaṛ laḡṇu ‘gyrate’; laṭ hannu ‘kick’; maṭṭarinu ‘squirm’; dʒmaḡaṇu, haḷḷaṇu, maḡcinu, dʒhukṇu ‘sway’; laṭkhaṛaṇu ‘teeter’; tannu ‘twitch’; etc.

42 Verbs of Assuming a Position

Class Members: dʒhukṇu, niḥurinu ‘bend/bow’; dʒhundinu ‘hang’; ghūḍa ṭekṇu ‘kneel’; dḥessinu ‘lean’; paḷṭaṇu ‘lie’; utraṇu ‘perch’; khaṣṇu ‘plop’; uṭṇu, ubhinu ‘rise’; baṣṇu ‘sit’; laṭrinu ‘slouch’; thaccinu ‘slump’; laṭṇu, leṭṇu ‘sprawl’; ṭukrukkaḡ baṣṇu ‘squat’; ubhṇu, uṭṇu ‘stand’; dʒhaṛṇu, giriṇu ‘stoop’; etc.

43 Verbs of Motion

43.1 Verbs of Inherently Directed Motion

Class Members: *baṛhnu* ‘advance’; *pugnu*, *aipugnu* ‘arrive’; *uklanu*, *caṛhnu* ‘ascend/climb’; *aunu* ‘come’; *taru* ‘cross’; *ḍanu*, *hīṛnu* ‘depart’; *orlanu*, *utranu*, *haru* ‘descend’; *bhitrinu*, *pasnu* ‘enter’; *bhagnu*, *umkanu*, *phutkanu* ‘escape’; *niskanu*, *bahirinu* ‘exit’; *ḍharu*, *giru* ‘fall’; *ḍanu* ‘go’; *choṛnu*, *tjagnu* ‘leave’; *paḥi haṭnu* ‘recede’; *pharkanu* ‘return’; *uṭhnu*, *udaunu* ‘rise’; *khurmurinu*, *gultinu* ‘tumble’; etc.

43.2 Leave Verbs

Class Members: *aphalnu*, *phjāknu*, *milkaunu* ‘abandon’; *choṛnu*, *tjagnu*, *ḍanu* ‘leave’; etc.

43.3 Manner of Motion Verbs

Roll Verbs

Class Members: *uphranu*, *upharnu* ‘bounce’; *bagnu*, *baḥaunu* ‘drift’; *khalsalnu*, *ḍharu* ‘drop’; *utranu*, *tairanu* ‘float’; *sulkanu*, *ciplanu* ‘glide’; *calnu* ‘move’; *guṛnu* ‘roll’; *ciplanu*, *sarnu* ‘slide’; *maccinu* ‘swing’; etc.

MOTION AROUND AN AXIS: *guḍulkinu* ‘coil’; *ghumnu* ‘revolve’; *caḥkar launu* ‘rotate’; *ghumnu* ‘spin’; etc.

Run Verbs

Class Members: *bhagnu* ‘bolt’; *uphranu* ‘bounce/bound’; *hutjaunu* ‘bowl’; *caṛhnu*, *uklanu* ‘climb’; *bame sarnu*, *ghasranu* ‘crawl’; *kamnu*, *thartharaunu* ‘dodder’; *bagnu* ‘drift’; *tairanu* ‘float’; *uṛnu* ‘fly’; *ciplanu* ‘glide’; *hatarinu* ‘hasten’; *hīṛnu* ‘hike’; *larkharaunu* ‘hobble’; *huttinu* ‘hurtle’; *kudnu* ‘jog’; *uphranu* ‘jump’; *khocjaunu* ‘limp’; *saṛkanu* ‘mosey’; *cimoṭnu* ‘nip’; *ghissinu* ‘plod’; *kudnu* ‘run’; *hīṛnu* ‘walk’; *ḍulnu* ‘wander’; etc.

43.4 Verbs of Motion Using a Vehicle

Verbs That Are Vehicle Names

Under this category, Levin (1993) has given following verb forms of English as class members: balloon; cycle; dogsled; ferry; gondola; helicopter; jeep; jet; kayak; moped; motor; motorbike; motorcycle; parachute; punt; raft; rickshaw; rocket; skate; skateboard; ski; sled; sledge; sleigh; taxi; toboggan; tram; trolley; yacht. But no such verbs are found in Nepali under this category.

Verbs That Are Not Vehicle Names

Class Members: *hāknu*, *kudaunu*, *calaunu* ‘drive’; *uṛnu* ‘fly’; *khijaunu* ‘oar/paddle/row’; *calnu* ‘pedal’; *caṛhnu* ‘ride’; etc.

43.5 Waltz Verbs

Class Members: *thunnu* ‘clog’; *nacnu* ‘dance’; *phiṭnu* ‘shuffle’; etc.

43.6 Chase Verbs

Class Members: *khednu*, *lakheṭnu*, *lagarnu*, *dhaṛaunu*, *bhaḡaunu* ‘chase’; *paḥjaunu* ‘follow’; *khoznu* ‘pursue’; *chaja parnu* ‘shadow’; *patto launu* ‘track’; etc.

43.7 Accompany Verbs

Class Members: sath lagnu/dinu ‘accompany’; sanchalanu gānu ‘conduct’; rakcha gānu ‘escort’; dōhorjaunu, baṭo dekhaunu ‘guide’; āguvai gānu ‘lead’; goṭhalo gānu ‘shepherd’; etc.

44 Verbs of Linging and Rushing

44.1 Verbs of Linging

Class Members: itrinu, ḍiskinu, chillinu ‘dally’; laṭinu, bāralinu ‘dawdle’; āber hunu, ḍhilo hunu ‘delay’; hickicjaunu ‘dither/hesitate’; vilamba hunu ‘linger’; bhāṭtarinu ‘loiter’; parkhanu ‘tarry’; etc.

44.2 Verbs of Rushing

Class Members: hatarinu, aturinu, atur hunu, haṭpatinu ‘hasten/hurry’; dhakelnu, ghaceṭnu ‘rush’; etc.

45 Measure Verbs

45.1 Register Verbs

Class Members: napnu ‘measure’; paṭhnu ‘read’; paṇḍjikalanu gānu ‘register’; ḍamma gānu, ḍoṭnu ‘total’; taṭlanu ‘weigh’; etc.

45.2 Cost Verbs

Class Members: vāhanu gānu ‘carry’; mol lagnu ‘cost’; khaṭnu ‘last’; paunu ‘take’; etc.

45.3 Fit Verbs

Class Members: boknu ‘carry’; āṭaunu ‘contain’; milnu, ṭhikka hunu ‘fit’; thamnu ‘hold’; basnu ‘seat’; sutnu ‘sleep’; rakhnu ‘store’; linu ‘take’; prājog gānu ‘use’; etc.

45.4 Price Verbs

Class Members: dam launu ‘appraise’; nirupanu gānu ‘assess’; ṭkāl gānu, aknu ‘estimate’; toknu ‘fix’; muljā/mol launu ‘price’; dār toknu ‘rate’; mol gānu ‘value’; etc.

45.5 Bill Verbs

Class Members: baḍji launu ‘bet’; ḍārimana tirnu ‘fine/mulct’; ḍogaunu, bācaunu ‘save’; etc.

46 Aspectual Verbs

46.1 Begin Verbs

Class Members: suru/thalani gānu, thalnu ‘begin/start’; banda/anta hunu ‘cease’; prarāmbha/ārāmbha gānu ‘commence’; nirāntarta dinu, calirāhanu, calu rakhnu ‘continue’; saknu, siddhjaunu, turnu, tungo launu, samapta gānu ‘end/finish’; ṭnu, roknu, rokinu, thaminu ‘halt’; thankjaunu, thanko launu ‘keep’; baṭhnu, calnu ‘proceed’; dōhorjaunu ‘repeat’; pheri thalnu ‘resume’; roknu, banda gānu ‘stop’; etc.

46.2 Complete Verbs

Class Members: pura/purna gānu, saknu, siddhjaunu ‘complete’; choṭnu, roknu ‘discontinue’; suru/thalani gānu, thalnu ‘initiate’; choṭnu ‘quit’; etc.

47 Weekend Verbs

Class Members: cutti/vida hunu ‘holiday/vacation’; bas basnu ‘sojourn’; etc.

48 Weather Verbs

Class Members: calnu, phuknu ‘blow’; sapha hunu ‘clear’; simsim pani parnu ‘drizzle/mizzle’; kuhiro lagnu ‘fog’; kaṭhjägrinu ‘freeze’; batas calnu ‘gust’; asina parnu ‘hail’; bidzuli camkinu ‘lightning’; tūvalo lagnu ‘mist’; panile cutnu ‘pelt’; pani darkanu, barsanu ‘pour/shower’; pani parnu ‘precipitate/rain’; gardzanu ‘roar’; asina parnu ‘sleet’; hiū parnu ‘snow’; charkanu ‘sprinkle’; catjan parnu, gatjanuruj garnu ‘thunder’; etc.

APPENDIX – III

Unique Verb Frames with their Sense IDs

1. k2	ãk%VT%S1%FID1, ãk%VT%S2%FID2 ukas%VT%S2%FID2, ukel%VT%S1%FID1 ukel%VT%S2%FID2, ukhel%VT%S%FID1 ukhel%VT%S2%FID2, ughar%VT%S1%FID1 ucal%VT%S2%FID2, uł%VT%S2%FID2 utar%VT%S1%FID1, ołh%VT%S1%FID1 ołh%VT%S2%FID2, kλmau%VCAUS%S1%FID1 kλmau%VT%S2%FID2, kλmau%VT%S3%FID3 kλrau%VI%S1%FID1, kλs%VT%S1%FID1 kλs%VT%S2%FID2, kał%VT%S1%FID1 kał%VT%S2%FID2, kał%VT%S3%FID3 kał%VT%S4%FID4, kał%VT%S6%FID6 kałh%VT%S2%FID2, kałh%VT%S3%FID3 kałh%VT%S4%FID4, karh%VT%S5%FID5 karh%VT%S6%FID6, karh%VT%S7%FID7 kił%VT%S1%FID1, kił%VT%S2%FID2 kuł%VT%S1%FID1, kuł%VT%S2%FID2 ker%VT%S1%FID1, ker%VT%S2%FID2 khλp%VI%S2%FID2, kha%VT%S1%FID1 kha%VT%S2%FID2, kha%VT%S3%FID3 kha%VT%S5%FID5, kha%VT%S6%FID6 kha%VT%S7%FID7, kha%VT%S8%FID8 kha%VT%S9%FID9, kha%VT%S10%FID10 kha%VT%S11%FID11, khijau%VT%S1%FID1 khijau%VT%S2%FID2, khel%VT%S1%FID1
-------	---

	khel%VT%S2%FID2, khodʒ%VT%S1%FID1 khodʒ%VT%S2%FID2, khol%VT%S1%FID1 khol%VT%S3%FID3, khol%VT%S4%FID4 gAr%VT%S1%FID1, gAr%VT%S2%FID2 cjat%VT%S1%FID1, cjap%VT%S1%FID1 cjap%VT%S2%FID2, chaɾ%VT%S1%FID1 chan%VT%S1%FID1, chan%VT%S2%FID2 dʒan%VT%S1%FID1, dʒan%VT%S2%FID2 dʒogau%VT%S1%FID1, dʒoɾ%VT%S1%FID1 dʒoɾ%VT%S2%FID2, dʒoɾ%VT%S3%FID3 dʒhik%VT%S2%FID2, dʒhos%VT%S1%FID1 dʒhos%VT%S2%FID2, ʈäg%VT%S2%FID2 ʈip%VT%S1%FID1, ʈip%VT%S2%FID2 ʈek%VT%S1%FID1, ʈek%VT%S2%FID2 ʈok%VT%S1%FID1, ʈok%VT%S2%FID2 ʈAg%VT%S1%FID1, ʈAg%VI%S2%FID2 ʈhuŋ%VT%S1%FID1, ʈhuŋ%VT%S2%FID2 ʈhel%VT%S1%FID1, ʈhok%VT%S1%FID1 ʈhok%VT%S2%FID2, ʈhok%VT%S3%FID3 qʰak%VT%S1%FID1, tAr%VT%S2%FID2 tan%VT%S1%FID1, tan%VT%S2%FID2 tok%VT%S1%FID1, tok%VT%S2%FID2 tok%VT%S3%FID3, tham%VT%S1%FID1 tham%VT%S2%FID2, dAbau%VT%S1%FID1 dAbau%VT%S2%FID2, dag%VT%S1%FID1 dag%VT%S2%FID2, dekh%VT%S1%FID1 dekh%VT%S1%FID1, dekh%VT%S2%FID2
--	---

	nagh%VT%S1%FID1, nap%VT%S1%FID1 nap%VT%S2%FID2, pakri%VT%S1%FID1 pakri%VT%S2%FID2, pakau%VT%S1%FID1 pakau%VT%S2%FID2, pacau%VT%S1%FID1 pacau%VT%S2%FID2, pau%VT%S1%FID1 pau%VT%S2%FID2, pal%VT%S1%FID1 puch%VT%S2%FID2, purjau%VT%S1%FID1 purjau%VT%S2%FID2, pokh%VT%S1%FID1 pokh%VT%S2%FID2, phal%VI%S2%FID2 phãt%VT%S1%FID1, phãt%VT%S2%FID2 phuk%VT%S1%FID1, phuk%VT%S2%FID2 pher%VT%S1%FID2, phjãk%VT%S1%FID1 phjãk%VT%S2%FID2, ban%VT%S1%FID1 batau%VT%S2%FID2, bãdh%VT%S1%FID1 bãdh%VT%S2%FID2, bãdh%VT%S3%FID3 bisau%VT%S1%FID1, ban%VT%S1%FID1 ban%VT%S2%FID2, banr%VT%S2%FID2 bhid%VI%S1%FID1, bhakur%VT%S1%FID1 bhakur%VT%S2%FID2, bhãt%VI%S1%FID1 bhãt%VT%S1%FID1, bhãt%VT%S2%FID2 ban%VI%S2%FID2, bhir%VT%S1%FID1 bhir%VT%S2%FID2, bhu%VT%S1%FID1 bhu%VT%S2%FID2, bhu%VT%S1%FID1 bhog%VT%S1%FID1, bhjau%VT%S1%FID1 manau%VI%S2%FID2, math%VT%S1%FID1 math%VT%S2%FID2, markau%VT%S1%FID1 markau%VT%S2%FID2, maski%VT%S2%FID1
--	--

	mAski%VI%S2%FID2, maḍḥ%VT%S1%FID1 man%VT%S2%FID1, man%VT%S2%FID2 mar%VT%S1%FID1, mar%VT%S1%FID2 mas%VT%S1%FID1, mic%VT%S2%FID1 mic%VT%S2%FID2, māl%VT%S1%FID1 muṭ%VT%S2%FID2, meṭ%VT%S1%FID1 meṭ%VT%S2%FID2, rAc%VT%S1%FID1 rAc%VT%S2%FID2, reṭ%VT%S1%FID1 reṭ%VT%S2%FID2, rakh%VT%S2%FID2 rūg%VT%S1%FID1, rūg%VT%S2%FID2 ruc%VI%S1%FID1, ruc%VI%S2%FID2 lAgau%VT%S1%FID1, lAchar%VT%S1%FID1 lAchar%VT%S2%FID2, lAṭhar%VI%S1%FID1 li%VT%S1%FID1, ljau%VT%S1%FID1 ljau%VT%S2%FID2, sAk%VI%S1%FID1 sAk%VI%S3%FID3, sAnki%VI%S1%FID1 sAmāu%VT%S1%FID1, sAmāu%VT%S2%FID2 sAmāu%VT%S3%FID3, sAmhal%VT%S1%FID1 sAmhal%VT%S2%FID2, sAr%VT%S2%FID2 hArāu%VT%S1%FID1, hArāu%VT%S2%FID2 hāk%VT%S1%FID1, hāk%VT%S2%FID2 hāk%VT%S3%FID3, hik%VI%S1%FID1 her%VT%S1%FID1, her%VT%S2%FID2 her%VT%S3%FID3, her%VT%S4%FID4 her%VT%S5%FID5, han%VT%S1%FID1 han%VT%S4%FID4, hAkar%VI%S1%FID1 hAkar%VI%S2%FID2, har%VI%S1%FID1
--	--

	hΛmki%VT%S1%FID1, hũtΛl%VT%S1%FID1
2. k1+verb	ughri%V8%S1%FID1, ughri%V8%S2%FID2 uʈh%VI%S4%FID4, uʈh%VI%S5%FID5 uʈh%VI%S6%FID6, utri%VI%S2%FID2 udau%VI%S1%FID1, umri%VT%S1%FID1 uʈ%VI%S1%FID1, uphri%VI%S2%FID2 khΛs%VI%S2%FID2, khΛs%VI%S2%FID2 kha%VI%S4%FID4, khul%VT%S1%FID1 khul%VI%S3%FID3, khul%VI%S5%FID5 khul%VI%S3%FID6, khol%VT%S2%FID2 gΛl%VI%S1%FID1, gΛl%VI%S2%FID2 ghΛʈ%VI%S1%FID1, cΛmki%VI%S1%FID1 cΛmki%VI%S2%FID2, cΛrki%VI%S1%FID1 cΛrki%VI%S2%FID2, cΛl%VI%S1%FID1 cΛl%VI%S3%FID3, cΛl%VI%S5%FID5 cΛl%VI%S6%FID6, cʃat%VT%S2%FID2 chɿʈau%VI%S2%FID2, chuʈ%VI%S1%FID1 chuʈ%VI%S2%FID2, chuʈ%VI%S3%FID3 dʒΛnmi%VI%S1%FID1, dʒΛnmi%VI%S2%FID2 dʒΛm%VI%S1%FID1, dʒΛl%VI%S1%FID1 dʒag%VI%S1%FID1, dʒag%VI%S2%FID2 dΛʈh %VI%S1%FID1, dΛʈh %VI%S2%FID2 dʰΛlki%VI%S1%FID1, thak%VI%S2%FID2 thun%VT%S2%FID2, pΛgli%VI%S1%FID1 pΛgli%VI%S2%FID2, pΛr%VI%S1%FID1 pΛr%VI%S4%FID4, pΛsau%VT%S1%FID1 pak%VI%S1%FID1, phΛl%VI%S1%FID1

	phuṭ%VI%S1%FID1, phuṭ%VI%S2%FID2 phul%VI%S1%FID1, phul%VI%S2%FID2 pher%VT%S1%FID1, baḡ%VI%S1%FID1 baḍḍ%VI%S1%FID1, baḍḍ%VI%S2%FID2 baṭṭh%VI%S1%FID1, baṭṭh%VI%S2%FID2 baṇ%VT%S2%FID2, bigri%VI%S1%FID1 bigri%VI%S2%FID2, bit%VI%S1%FID1 bit%VI%S2%FID2, bit%VI%S3%FID3 bisau%VT%S2%FID2, bhasi%VI%S1%FID1 bhaṭki%VI%S1%FID1, bhaṭki%VI%S2%FID2 bhaṅki%VI%S1%FID1, maḷci%VI%S1%FID1 maṛ%VI%S2%FID2, maḍḡh%VT%S2%FID2 maś%VT%S2%FID2, mil%VT%S3%FID3 mil%VI%S5%FID5, raṣau%VI%S2%FID2 raḥa%VI%S1%FID1, raṅki%VI%S1%FID1 raṅki%VI%S3%FID3, rāki%VI%S1%FID1 rāki%VI%S2%FID2, rāki%VI%S3%FID3 laṭ%VT%S1%FID2, laṃmi%VI%S1%FID1 laṃmi%VI%S2%FID2, laṭṭhar%VI%S2%FID2 lag%VI%S9%FID9, lag%VT%S12%FID12 lag%VI%S13%FID13, lag%VI%S14%FID14 saṅki%VI%S2%FID2, saṇpri%VI%S1%FID1 saṇpri%VI%S2%FID2, salki%VI%S1%FID1 saṛ%VI%S3%FID3, suk%VT%S1%FID1 haḷli%VI%S1%FID1, hik%VI%S2%FID2 haṃki%VI%S2%FID2
3. k1+k7p+verb	uphri%VI%S3%FID3, kaṛau%VI%S2%FID2

	hΛt%VI%S2%FID2
7. k1+k7+k2+verb	pΛsau%VT%S1%FID1, pirol%VT%S1%FID1 bhΛr%VT%S1%FID1, bhet%VT%S2%FID2 bhog%VT%S2%FID2, much%VT%S2%FID2 rakh%VT%S1%FID1, lΛgau%VT%S1%FID2 lΛgau%VT%S3%FID4, li%VT%S2%FID2 her%VT%S6%FID6, han%VT%S3%FID3
8. k7p+k1+verb	ubhi%VI%S1%FID1, pal%VT%S2%FID2 ubdʒi%VT%S1%FID1, ubhi%VI%S2%FID2 dʒΛm%VI%S2%FID2, niski%VI%S2%FID2 bΛg%VI%S2%FID2, bΛth%VI%S4%FID4 macci%VI%S2%FID2, hal%VT%S1%FID1
9. k4a+k2+verb	au%VI%S2%FID2, pΛr%VT%S2%FID2 pΛr%VT%S3%FID3, pug%VI%S2%FID2 lag%VI%S2%FID2, lag%VT%S5%FID5 lag%VT%S8%FID8, lag%VT%S10%FID10 lag%VT%S11%FID11
10. k1+rd+verb	cΛl%VI%S2%FID2, dʰhΛki%VI%S2%FID2 bΛth%VI%S3%FID3, bhasi%VI%S2%FID2 mΛr%VI%S4%FID4, muɽ%VI%S1%FID1 lag%VT%S1%FID1, lag%VT%S3%FID3
11. k1+k7p+k2+verb	chΛr%VT%S1%FID1, chΛr%VT%S2%FID2 t̪āg%VT%S1%FID1, t̪ās%VT%S1%FID1 t̪hel%VT%S2%FID2, rop%VT%S1%FID1 lΛgau%VT%S3%FID3
12. k1+k5+k2+verb	ukas%VT%S1%FID1, ukas%VT%S3%FID3 ughar%VT%S2%FID2, utar%VT%S2%FID2

	kaɾh%VT%S1%FID1, dʒogau%VT%S2%FID2 dʒhik%VT%S1%FID1
13. k1+k3+verb	pak%VI%S2%FID2, mat%VI%S1%FID1 mat%VI%S2%FID2, rapi%VI%S2%FID2 sʌrki%VI%S1%FID1, har%VI%S2%FID2
14. k1+ras-k+verb	khel%VI%S4%FID4, dʒʌl%VI%S2%FID2 ʈʌs%VT%S2%FID2, tʌrki%VI%S1%FID1
15. k3+k2+verb	mʌr%VI%S3%FID3, mʌr%VI%S3%FID3 rapi%VI%S1%FID1, hʌlli%VI%S2%FID2
16. k1+ras-k+k2+verb	mʌl%VT%S2%FID2, much%VT%S1%FID1 mil%VI%S1%FID1
17. k1+k2+k7p+verb	bhet%VT%S1%FID1, thun%VT%S1%FID1 rʌhʌ%VI%S2%FID2
18. k5+k1+verb	rʌsau%VI%S1%FID1, uɾh%VI%S3%FID3 dʒhar%VI%S1%FID1
19. k1+k2p+verb	au%VI%S1%FID1, pug%VI%S1%FID1 bhjau%VT%S2%FID2
20. k1+k4+k2+verb	dag%VT%S3%FID3, di%VT%S1%FID1 di%VT%S2%FID2
21. k1+k2+k7+verb	nagh%VT%S2%FID2, bʌsau%VCAUS%S1%FID1 bʌsau%VCAUS%S2%FID2
22. k1+rsp+verb	khʌp%VI%S1%FID1, cʌl%VI%S4%FID4 rʌhʌ%VI%S3%FID3
23. k1+k2+rh+verb	tok%VT%S4%FID4, mʌnau%VI%S1%FID1
24. k4+k2+verb	mil%VI%S2%FID1, lag%VT%S5%FID5
25. k7t+k1+verb	thak%VI%S1%FID1
26. rh+k2+verb	mil%VT%S4%FID4

27. k5+k2+verb	chaŋ%VT%S2%FID2
28. k4+k3+verb	pirol%VT%S2%FID2
29. k1+k2+k3+verb	rop%VT%S2%FID2
30. k1+k3+k2+verb	puch%VT%S1%FID1
31. k1+k2+rd+verb	ucal%VT%S1%FID1
32. k1+k7t+k2+verb	uker%VT%S1%FID1
33. ras-k+k2+verb	mil%VI%S1%FID1
34. k3+k7p+verb	q̣hak%VT%S2%FID2
35. k4a+k3+verb	pirol%VT%S2%FID2
36. pk1+mk1+k2+verb	chã̃tau%VCAUS%S1%FID1

APPENDIX –IV **Ambiguous Nepali Verbs with respective number of senses**

NUMBER OF SENSES	AMBIGUOUS NEPALI VERBS WITH DIFFERENT SENSES
Two Senses	<i>au</i> (to come, to know), <i>āk</i> (to estimate, to draw a sketch), <i>uker</i> (to earth up, to instigate), <i>ukel</i> (to vomit, to reveal), <i>uksi</i> (to get rid, to provoke), <i>ukhel</i> (to uproot, to dislocate), <i>ughri</i> (to open, to brighten up), <i>ughar</i> (to open, to unveil), <i>ucal</i> (to lift up, to instigate), <i>ur</i> (to fly, to splurge), <i>utar</i> (to draw a sketch, to unload), <i>utri</i> (to get off, to get rid), <i>udau</i> (to rise, to prevail), <i>ubd̥zi</i> (to grow, to stem), <i>umri</i> (to sprout, to born), <i>orh</i> (to cover, to wear), <i>karau</i> (to shout, to scold), <i>kas</i> (to fasten, to bargain), <i>kit</i> (to determine, to gnash), <i>kuṭ</i> (to beat up, to grind), <i>ker</i> (to interrogate, to erase), <i>khap</i> (to last, to bear), <i>khās</i> (to fall down, to pass away), <i>khijau</i> (to row boat, to become slim), <i>khod̥z</i> (to search, to want), <i>gal</i> (to get tired, to melt down), <i>gar</i> (to do, to perform), <i>ghat</i> (to reduce, to occur), <i>ghum</i> (to rotate, to visit), <i>carh</i> (to climb, to ride), <i>camki</i> (to glitter, to turn up), <i>car̥ki</i> (to crack, to shriek), <i>cjat</i> (to tear, to split apart), <i>cjap</i> (to hold, to favour), <i>chāṭau</i> (to trim, to get mad), <i>char</i> (to scatter, to sow), <i>char̥</i> (to leave, to set free), <i>chan</i> (to filter, to choose), <i>d̥zanmi</i> (to born, to form), <i>d̥zam</i> (to freeze, to gather), <i>d̥zal</i> (to burn, to envy), <i>d̥zan</i> (to know, to understand), <i>d̥zag</i> (to wake up, to become aware), <i>d̥zogau</i> (to spare, to protect), <i>d̥zhar</i> (to fall off, to stoop), <i>d̥zhik</i> (to remove, to expel), <i>d̥zhos</i> (to kindle, to complain), <i>tāg</i> (to hang, to delay), <i>tās</i> (to stick, to cling), <i>tip</i> (to pick up, to pluck), <i>tek</i> (to crush, to lean on), <i>tok</i> (to chew, to bite), <i>thag</i> (to cheat, to swindle), <i>thun</i> (to bite, to peck), <i>thel</i> (to push, to stow), <i>d̥arh</i> (to burn, to ache), <i>d̥hal̥ki</i> (to fall down, to attract), <i>d̥hak</i> (to cover, to spread), <i>tap</i> (to meditate, to burn), <i>tar</i> (to cross, to glare), <i>tarki</i> (to avoid, to drop), <i>tan</i> (to pull, to captivate), <i>thak</i> (to tire, to stop lactating), <i>thun</i> (to imprison, to close down), <i>d̥abau</i> (to press, to exploit), <i>di</i> (to give, to provide), <i>dekh</i> (to see, to foresee), <i>nag</i> (to cross, to outscore), <i>nap</i> (to measure, to become unable), <i>nikal</i> (to extract, to expel), <i>niski</i> (to come out, to release), <i>nihuri</i> (to surrender, to bow), <i>pakri</i> (to catch, to arrest), <i>pakau</i> (to cook, to ripen), <i>pagli</i> (to melt down, to impress), <i>pacau</i> (to digest, to usurp), <i>pau</i> (to catch, to feel), <i>pak</i> (to ripe, to become mature), <i>pal</i> (to rear, to come), <i>pirol</i> (to grill, to pain), <i>pug</i> (to reach, to suffice), <i>puch</i> (to wipe, to erase), <i>purjau</i> (to convey, to maintain), <i>pokh</i> (spill, to express), <i>phal</i> (to yield, to grind), <i>phāṛ</i> (to clear, to make paste), <i>phuk</i> (to blow, to caution), <i>phuṭ</i> (to break down, to split apart), <i>phul</i> (to bloom, to swell), <i>pher</i> (to change, to breathe), <i>phjāk</i> (to throw, to abort), <i>bag</i> (to flow, to float), <i>bad̥z</i> (to ring, to strike), <i>ban</i> (to become, to construct), <i>bas</i> (to sit, to live), <i>batau</i> (to tell, to winnow), <i>basau</i> (to sit, to accommodate), <i>bigri</i> (to spoil, to break down), <i>bisau</i> (to lay down, to ache), <i>bhan</i> (to say, to utter), <i>bhag</i> (to escape, to run), <i>bhid̥z</i> (to drench, to immerse), <i>bhakur</i> (to beat up, to devour), <i>bhar̥ki</i> (to startle, to deviate), <i>bhasi</i> (to subside, to go), <i>bhāṛ</i> (to spoil, to mislead), <i>bhatki</i> (to collapse, to

	ruin), <i>bhanki</i> (to buzz, to get angry), <i>bhir</i> (to carry, to wear), <i>bhut</i> (to fry up, to shoot down), <i>bhul</i> (to forget, to indulge), <i>bhet</i> (to meet up, to become equal), <i>bhog</i> (to experience, to enjoy), <i>bhjau</i> (to finish, to manage), <i>macci</i> (to swing, to be), <i>manau</i> (to agree, to celebrate), <i>math</i> (to churn, to investigate), <i>markau</i> (to twist, to sprain), <i>maski</i> (to get cooked, to blush), <i>madzh</i> (to scrub, to improve), <i>man</i> (to accept, to worship), <i>mat</i> (to intoxicate, to spoil), <i>mar</i> (to kill, to beat), <i>mas</i> (to spend, to extinct), <i>mur</i> (to turn around, to shave), <i>much</i> (to mix up, to involve), <i>munṭi</i> (to fall down, to go), <i>meṭ</i> (to erase, to quench), <i>rach</i> (to create, to plan), <i>rasau</i> (to leak, to weep), <i>reṭ</i> (to cut, to play), <i>rakh</i> (to keep, to spare), <i>rapi</i> (to heat, to enrage), <i>rūg</i> (to wait, to watch), <i>ruch</i> (to like, to choose), <i>rop</i> (to plant, to stab), <i>lachar</i> (to pull, to drag), <i>laṛ</i> (to fall down, to fight), <i>laṭhar</i> (to join, to fight), <i>lanḍi</i> (to fall, to get cheated), <i>lammi</i> (to prolong, to delay), <i>li</i> (to get, to hold), <i>ljau</i> (to bring, to marry), <i>sanki</i> (to get angry, to get mad), <i>sapri</i> (to flourish, to prosper), <i>sābhal</i> (to care, to preserve), <i>salki</i> (to burn, to spread), <i>sarki</i> (to choke up, to boast), <i>suk</i> (to dry up, to wither), <i>sulki</i> (to slip, to pass away), <i>halli</i> (to dangle, to become afraid), <i>harau</i> (to lose, to defeat), <i>haṭ</i> (to separate, to shift), <i>hakar</i> (to yell, to abuse), <i>har</i> (to lose, to get upset), <i>hamki</i> (to fan, to bellow), <i>hūṛal</i> (to mix, to fight)
Three Senses (16)	<i>ukas</i> (to pull out, to provoke, to give bail), <i>uphri</i> (to jump, to get angry, to frisk), <i>kamau</i> (to make shiver, to earn, to cultivate), <i>chut</i> (to separate, to miss, to get rid), <i>dṣoṛ</i> (to repair, to patch up, to add), <i>thok</i> (to strike, to hit, to beat), <i>dag</i> (to like, to aim, to shoot), <i>bādh</i> (to tie up, to unite, to handcuff), <i>bit</i> (to pass time, to pass away, to ruin up), <i>raha</i> (to remain, to stay, to last), <i>ranki</i> (to heat, to get angry, to resonate), <i>rāki</i> (to flare, to rage, to fatten), <i>sak</i> (to finish, can, to spoil), <i>samau</i> (to hold, to catch, to arrest), <i>sar</i> (to shift, to transmit, to glide), <i>hāk</i> (to drive, to boast, to challenge)
Four Senses (9)	<i>khel</i> (to play, to act, to muse, to play with fire), <i>khol</i> (to open, to set free, to establish, to start), <i>tok</i> (to determine, to fix, to identify, to appoint), <i>par</i> (to fall/drop, to feel, to lit, to come), <i>baḥ</i> (to increase, to spread, to move forward, to grow), <i>mar</i> (to die away, painstaking, to dry up, to want), <i>lagau</i> (to wear, to apply, to sow, to fence), <i>hal</i> (to store, to insert, to mix, to spend), <i>han</i> (to throw, to strike, to lose, to hit)
Five Senses (2)	<i>uṭh</i> (to wake up, to get up, to arise, to build, to collect, to pass away), <i>mil</i> (to mix, to get, to settle down, to arrange, to agree)
Six Senses (4)	<i>kaṭ</i> (to refute, to issue, to cross, to pass, to slice, to book), <i>khul</i> (to open up, to get free, to establish, to brighten up, to reveal, to turn clear), <i>cal</i> (to move, to go, to start, to blow, to tease, to go), <i>her</i> (to look at, to check, to read, to look after, to wait, to study)
Seven Senses	<i>kaḥ</i> (to borrow, to call out, to outcaste, to uproot, to dribble, to flay, to extract)

Eleven Senses (1)	<i>kha</i> (to eat, to drink, to smoke, to survive, to gobble, be, to get beat, to bribe, to get into trouble, to get cheated, to kiss)
Fifteen Senses (1)	<i>lag</i> (to start up, to feel, to support, to go, to utilize, to stain, to win, to possess, to ring, to inebriate, to hit, to lean on, to rise, to close down, to chase)

APPENDIX V

Sample Verb Frames

A complete database is available at <https://github.com/KrishnaMManger/Nepali-Verb-Frames>

1.

```
Verb::au
SID::au%VI%S1
Verb_Sense::agaman_hu
Eng_Gloss::Come
Verb_class::Verb of motion
Verbs_in_Same_Class::Synonyms>
Example::u      ghar      a-jo
           he.NOM  home.LOC  come-3.SG.PST
           'He came home'
Theta_Roles::AGENT GOAL VERB
Frame_ID::au%VI%S1%FID1
```

```
-----
arc-label necessity vibhakti  lextype  ontology
-----
k1          m          0          p/n      [+ani/+hum]
k2p         m          0          n         [+plc]
-----
```

```
SID::au%VT%S2
Verb_Sense::gan
Eng_Gloss::Know
Verb_class::Verb of perception
Verbs_in_Same_Class::Synonyms>gan%VT%S1%FID1
Example::us-lai  dherai  kura      au-cha
           He-DAT  many   things.ACC  know-3.SG.NPST
           'He knows many things'
Theta_Roles::EXPERIENCER TOPIC VERB
Frame_ID::au%VI%S2%FID2
```

```
-----
arc-label necessity vibhakti  lextype  ontology
-----
k4a          m          lai      n/p      [+ani/+hum]
k2           m          0         n         [+top]
-----
```

2.

Verb::āk
SID::āk%VT%S1
Verb_Sense::anuman_gar
Eng_Gloss::Estimate
Verb_class::Measure verb
Verbs_in_Same_Class::Synonyms>
Example::mistri-le rukh-ko nap āk-jo
carpenter-ERG tree-GEN size.ACC estimate-3.SG.PST
'Carpenter estimated the size of a tree'
Theta_Roles::AGENT ATTRIBUTE VERB
Frame_ID::āk%VT%S1%FID1

arc-label	necessity	vibhakti	lextype	ontology
k1	m	le/0	p/	[+ani/+hum]
k2	m	0	n	[+msmt]

SID::āk%VT%S2
Verb_Sense::chitra_banau
Eng_Gloss::Draw sketch
Verb_class::Performance verb
Verbs_in_Same_Class::Synonyms>utar%V%S1%FID1
Example::ram-le mero chitraāk-jo
Ram-ERG my sketch.ACC draw-3.SG.PST
'Ram drew my sketch'
Theta_Roles::AGENT THEME VERB
Frame_ID::āk%VT%S2%FID2

arc-label	necessity	vibhakti	lextype	ontology
k1	m	le/0	p/n	[+ani/+hum]
k2	m	0	n	[+art, +obj]

3.

Verb::ukas
 SID::ukas%VT%S1
 Verb_Sense::uthau
 Eng_Gloss::Pull out
 Verb_class::Verb of exerting Force
 Verbs_in_Same_Class::Synonyms>ukas%VT%S1%FID1,tan%VT%S1%FID1
 lAchar%VT%S1%FID1

Example::us-le khaḍal-baṭṭa ḍhungo ukas-jo
 he-ERG pit-ABL stone.ACC pull-3.SG.PST
 'He pulled out the stone from the pit'

Theta_Roles::AGENT SOURCE THEME VERB

Frame_ID::ukas%VT%S1%FID1

arc-label	necessity	vibhakti	lextype	ontology
k1	m	le/0	p/n	[+ani/+hum]
k5	d	baṭṭa/dekhi	n	[+plc]
k2	m	0/lai	p/n	[+con]

SID::ukas%VT%S2
 Verb_Sense::utteḡit_par
 Eng_Gloss::Provoke
 Verb_class::Verb of psychological State
 Verbs_in_Same_Class::Synonyms>uker%VT%S2%FID2,uksi%VT%S6%FID6
 ucal%VT%S2%FID2

Examlpe::us-le tjaṣ keṭa-lai ukas-jo
 he-ERG that boy-ACC provoke-3.SG.PST
 'He provoked that boy'

Theta_Roles::AGENT PATIENT VERB

Frame_ID::ukas%VT%S2%FID2

arc-label	necessity	vibhakti	lextype	ontology
k1	m	le/0	p/n	[+ani/+hum]
k2	m	lai	p/n	[+ani/+hum]

```

SID::ukas%VT%S3
Verb_Sense::mukta_gar
Eng_Gloss::Bail out
Verb_class::Verb of removing
Verbs_in_Same_Class::Synonyms>uker%VT%S2%FID2
Examlpe::us-le  Øel-batΔ  Δparadhi-lai  ukas-jo
           he-ERG  jail-ABL  culprit-ACC  bail-3.SG.NPST
           'He bailed the culprit out from the jail'
Theta_Roles::AGENT SOURCE PATIENT VERB
Frame_ID::ukas%VT%S3%FID3
-----
arc-label  necessity      vibhakti      lextype      ontology
-----
k1          m              le/0          p/n          [+ani/+hum]
k5          d              batΔ/dekhi    n            [+plc]
k2          m              lai           p/n          [+ani/+hum]
-----
*****

```


5.

Verb::ukel
 SID::ukel%VT%S1
 Verb_Sense::vanta_gar
 Eng_Gloss::Vomit
 Verb_class::Verb involving the body
 Verbs_in_Same_Class::Synonyms>
 Example::us-le khaeka kura ᄃᄃᄃᄃᄃᄃᄃ ukel-jo
 He-ERG eatery thing.ACC all vomit-3.SG.NPST
 'He vomited everything he ate before'
 Theta_Roles::PATIENT MATERIAL VERB
 Frame_ID::ukel%VT%S1%FID1

```
-----
arc-label necessity vibhakti  lextype  ontology
-----
k1          m          le/0      p/n      [+ani/+hum]
k2          m          0         n        [+any]
-----
```

SID::ukel%VT%S2
 Verb_Sense::bhan
 Eng_Gloss::Reveal
 Verb_class::Verb of communication
 Verbs_in_Same_Class::Synonyms>khul%VT%S5%FID5
 Example::us-le mera rahasja ukel-jo
 he-ERG my secret.ACC reveal-3.SG.NPST
 'He revealed all my secrets'
 Theta_Roles::AGENT TOPIC VERB
 Frame_ID::ukel%VT%S2%FID2

```
-----
arc-label necessity vibhakti  lextype  ontology
-----
k1          m          le/0      p/n      [+ani/+hum]
k2          m          0         n        [+top]
-----
```

```
Verb::uksi
SID::uksi%VI%S1
Verb_Sense::mukta_hu
Eng_Gloss::Rid
Verb_class::Verbs of removing
Verbs_in_Same_Class::Synonyms>utri%VI%S2%FID2,chut%VI%S3%FID3
Example::mā      rin-baṭa ũksi-ẽ
          I.NOM  loan-ABL  rid-1.SG.NPST
          'I got rid of the loan'
Theta_Roles::AGENT SOURCE VERB
Frame_ID::uksa%VI%S1%FID1
```

7.

Verb::ukhel
SID::ukhel%VT%S1
Verb_Sense::nikal
Eng_Gloss::Uproot
Verb_class::Verb of removing
Verbs_in_Same_Class::Synonyms>kaṛh%VT%S4%FID4
Example::us-le ghāś ukhel-jo
 he-ERG grass.ACC uproot-3.SG.M.PST
 'He uprooted the grass'
Theta_Roles::AGENT MATERIAL VERB
Frame_ID::ukhel%VT%S%FID1

arc-label necessity vibhakti lextype ontology

k1 m le/0 p/n [+ani/+hum]
k2 m 0 n [+con]

SID::ukhel%VT%S2
Verb_Sense::uṭhibas_lau
Eng_Gloss::Dislocate
Verb_class::Verb of removing
Verbs_in_Same_Class::Synonyms>
Example::sarkar-le avaidh ghar-haru ukhel-jo
 Government-ERG illegal house-PL.ACC dislocate-3.SG.PST
 'The Government dislocated illegal houses'
Theta_Roles::AGENT MATERIAL VERB
Frame_ID::ukhel%VT%S2%FID2

arc-label necessity vibhakti lextype ontology

k1 m le/0 n/p [+ani, +org]
k2 m 0 n [obj]

8.

Verb::ughri
SID::ughri%VI%S1
Verb_Sense::khul
Eng_Gloss::Open up
Verb_class::Verb of appearance
Verbs_in_Same_Class::Synonyms>khul%VI%S1%FID1
Example::qhokā ughri-jo
 door.NOM open-3.SG.PST
 'The door got opened'
Theta_Roles::THEME VERB
Frame_ID::ughri%V8%S1%FID1

arc-label necessity vibhakt lextype ontology

k1 m 0 n/p [+con]

SID::ughri%VI%S2
Verb_Sense::uḡjalo_hu
Eng_Gloss::Brighten up
Verb_class::Verb of change of state
Verbs_in_Same_Class::Synonyms>
Example::din ughri-jo
 day.NOM brighten-3.SG.PST
 'The day has brightened up'
Theta_Roles::THEME VERB
Frame_ID::ughri%V8%S2%FID2

arc-label necessity vibhakti lextype ontology

k1 m 0 n/p [+abs]

9.

Verb::ughar
 SID::ughar%VT%S1
 Verb_Sense::khol
 Eng_Gloss::Open
 Verb_class::Verb of change of state
 Verbs_in_Same_Class::Synonyms>khol%VT%S1%FID1
 Example::nani-le akha ughar-jo
 Baby-ERG eyes.ACC open-3.SG.PST
 'The baby opened the eyes'
 Theta_Roles::AGENT THEME VERB
 Frame_ID::ughar%VT%S1%FID1

```
-----
arc-label necessity vibhakti lextype ontology
-----
k1          m          le/0          n/p          [+ani/+hum]
k2          m          0              n            [+bdp, +mensta]
-----
```

SID::ughar%VT%S2
 Verb_Sense::khol_haṭau
 Eng_Gloss::Unveil
 Verb_class::Verb of communication
 Verbs_in_Same_Class::Synonyms>
 Example::us-le rahasjā-baṭa parda ughar-jo
 he-ERG secret-ABL screen.ACC unveil-3.SG.PST
 'He unveiled the secret'
 Theta_Roles::AGENT SOURCE TOPIC VERB
 Frame_ID::ughar%VT%S2%FID2

```
-----
arc-label necessity vibhakti lextype ontology
-----
k1          m          le/0          p/n          [+ani/+hum]
k5          m          baṭa/dekhi   n            [+top]
k2          m          0              n            [+con]
-----
```

11.

```

Verb::uṭh
SID::uṭh%VI%S1
Verb_Sense::bjũḡhΛ
Eng_Gloss::wake up
Verb_class::Verb of change of state
Verbs_in_Same_Class::Synonyms>
Example::u      bihaṇ-Λi      uṭh-cha
           he.NOM  morning-EMPH.LOC wake-3.SG.M.NPST
           'He wakes up early in the morning'
Theta_Roles::ACTOR TIME VERB
Frame_ID::uṭh%V11%S1%FID1
-----
arc-label necessity vibhakti  lextype  ontology
-----
k1          m          0          p/n          [+ani/+hum]
k7t         d          0/ma      adv          [+tme]
-----
*****

SID::uṭh%VI%S2
Verb_Sense::ubhi
Eng_Gloss::Get up
Verb_class::Verb of motion
Verbs_in_Same_Class::Synonyms>ubhi%VI%S1%FID1
Example::u      chaṭki-baṭΛ      uṭh-jo
           he.NOM  chair-ABL      [get up]-3.SG.M.PST
           'He got up from the chair'
Theta_Roles::AGENT SOURCE VERB
Frame_ID::uṭh%V11%S2%FID2
-----
arc-label necessity vibhakti  lextype  ontology
-----
k1          m          0          p/n          [+ani/+hum]
k5          d          baṭΛ/dekhi  n          [+plc, +obj]
-----
*****

```

SID::uṭh%VI%S3
 Verb_Sense::birodh_gar
 Eng_Gloss::Arise
 Verb_class::Verb of communication
 Verbs_in_Same_Class::Synonyms>
 Example::ḡanta-baṭa avaḡ uṭh-jo
 public-ABL voice.NOM arise-3.SG.PST
 'The voice arose from the public'
 Theta_Roles::SOURCE TOPIC VERB
 Frame_ID::uṭh%VI%S3%FID3

 arc-label necessity vibhakti lextype ontology

 k5 m baṭa/dekh n [+plc]
 k1 m 0 n [+top]

SID::uṭh%VI%S4
 Verb_Sense::ban
 Eng_Gloss::Build
 Verb_class::Verb of creation
 Verbs_in_Same_Class::Synonyms>ban%VI%S2%FID2, hu%VI%S2%FID2
 Examlpe::ghar-ko dhuri uṭh-jo
 house-GEN axis.NOM build-3.SG.PST
 'The axis of the house is built'
 Theta_Roles::MATERIAL VERB
 Frame_ID::uṭh%VI%S4%FID4

 arc-label necessity vibhakti lextype ontology

 k1 m 0 n [+con]


```
Verb::ur
SID::ur%VI%S1
Verb_Sense::ghum
Eng_Gloss::Fly
Verb_class::Verb of motion
Verbs_in_Same_Class::Synonyms>
Example::caro      ur-jo
          Bird.NOM  fly-3.SG.P
          'The bird flew away'
Theta_Roles::AGENT VERB
Frame_ID::ur%VI%S1%FID1
```

arc-label	necessity	vibhakti	lextype	ontology
k1	m	0	n/p	[+ani]

```

SID::ur%VT%S1
Verb_Sense::siddhi/sΔk
Eng_Gloss::Splurge
Verb_class::Aspectual Verb
Verbs_in_Same_Class::Synonyms>sΔk%VT%S2%FID1
Example::chora-le paisa          ur-a-jo
          Son-ERG  money.ACC      splurge-CAUS-3.SG.PST
          'The son splurged the money'
Theta_Roles::AGENT ASSET VERB
Frame_ID::ur%VT%S2%FID2

```

arc-label	necessity	vibhakti	lextype	ontology
k1	m	le/0	n/p	[+ani/+hum]
k2	m	0	n	[+cur]

.....

13.

```

Verb::utar
SID::utar%VT%S1
Verb_Sense::chitra kor
Eng_Gloss::Draw Sketch
Verb_class::Performance verb
Verbs_in_Same_Class::Synonyms>āk%VT%S2%FID2
Example::hari-le      mero naksā      utar-jo
           Hari-ERG    my picture.ACC draw-3.SG.M.PST
           'Hari drew my sketch'
           Theta_Roles::AGENT PRODUCT VERB
Frame_ID::utar%VT%S1%FID1
-----
arc-label necessity vibhakti  lextype  ontology
-----
k1          m          le/0      p/n      [+ani/+hum]
k2          m          0         n        [+art]
-----
*****

SID::utar%VT%S2
Verb_Sense::oral
Eng_Gloss::Unload
Verb_class::Verb of removing
Verbs_in_Same_Class::Synonyms>
Example::hari-le      gaḍi-baṭṭa      saman      utar-jo
           Hari-ERG vehicle-ABL goods.ACC unload-3.SG.M.PST
           'Hari unloaded goods from the vehicle'
           Theta_Roles::AGENT SOURCE MATERIAL VERB
Frame_ID::utar%VT%S2%FID2
-----
arc-label necessity vibhakti  lextype  ontology
-----
k1          m          le/0      p/n      [+ani/+hum]
k5          d          baṭṭa     n        [+con]
k2          m          0/lai     n        [+con]
-----
*****

```

14.

Verb::utri
 SID::utri%VI%S1
 Verb_Sense::ḡhar
 Eng_Gloss::Get down
 Verb_class::Verb of motion
 Verbs_in_Same_Class::Synonyms>
 Example::hari gadi-baṭṭa utri-jo
 Hari.NOM vehicle-ABL [get down]-3.SG.M.PST
 'Hari got down from the vehicle'
 Theta_Roles::AGENT SOURCE VERB
 Frame_ID::utri%VI%S1%FID1

```
-----
arc-label necessity vibhakti    lextype    ontology
-----
k1            m            0            n/p            [+ani/+hum]
k5            m            baṭṭa/dekhi    n            [+con]
-----
```

SID::utri%VI%S2
 Verb_Sense::ḡhaṭ
 Eng_Gloss::Decrease
 Verb_class::Verb of change of state
 Verbs_in_Same_Class::Synonyms>uksi%VI%S1%FID,chuṭ%VI%S3%FID3
 Example::ama-ko piṛa utri-jo
 mother-GEN sorrow.NOM decrease-3.SG.PST
 'Mother's sorrow decreased'
 Theta_Roles::THEME VERB
 Frame_ID::utri%VI%S2%FID2

```
-----
arc-label necessity vibhakti    lextype    ontology
-----
k1            m            0            n            [+abs]
-----
```

15.

```
Verb::udau
SID::udau%VI%S1
Verb_Sense::ḡhulki
Eng_Gloss::Rise up
Verb_class::Verb of appearance
Verbs_in_Same_Class::Synonyms>lag%VI%S13%FID13
Example::gham    uḡa-jḡ
          sun.NOM rise-3.SG.PST
          'The sun rose'
Theta_Roles::THEME  VERB
Frame_ID::udau%VI%S1%FID1
-----
arc-label necessity vibhakti  lextype  ontology
-----
k1          m          0          n          [+cels]
-----
*****

SID::udau%VI%S2
Verb_Sense::dekhḡ_pḡr
Eng_Gloss::Appear
Verb_class::Verb of appearance
Verbs_in_Same_Class::Synonyms>
Example::mero    jiwan-mḡ    khusijḡli    uḡa-jḡ
          my    life-LOC    happiness.NOM appear-3.SG.M.PST
          'Happiness appeared in my life'
Theta_Roles::LOCATION THEME VERB
Frame_ID::udau%VI%S2%FID1
-----
arc-label necessity vibhakti  lextype  ontology
-----
k7          m          mḡ          n          [+abs]
k1          m          0          n          [+abs]
-----
*****
```

APPENDIX - VI
Research Papers Published

Argument Structure of Nepali Verbs: A Study on Lexico-Semantic Ambiguity

KRISHNA MAYA MANGER
University of Hyderabad, Telengana, India

ABSTRACT

Ambiguity is a common phenomenon for any natural language. It occurs in lexical, syntactic, lexico-semantic and pragmatic levels. The issue of ambiguity in any linguistic levels has proved to be a biggest challenge in the field of Natural Language Processing (NLP). In this context, this paper is an analysis of lexico-semantic ambiguity with a particular reference to Nepali language. It is an initial attempt to enhance Nepali verb sense disambiguation task following knowledge based approach. To understand a word in its context is a first step of disambiguation task and understanding a verb with its argument structure supplies basic knowledge about its context as well as its selectional preference. In this paper, an attempt has been made to analyze argument structure of 10 most ambiguous Nepali verbs in order to provide linguistic insight for the disambiguation of those verbs.

Keywords: Ambiguity, verbs, Nepali, argument structure

1. INTRODUCTION

Being attributed with expressions which possess more than one sense or meaning is a common phenomenon for every natural language. Natural language by the virtue of such attribute is an organised system of “finite set of rules with infinite number of expressions.” It is quite fascinating that a language having limited number of lexical items with handful of sentence formation strategies is capable of representing unlimited and

uncountable expressions made by creative human mind. One of the significant factors which equipped a language with such a generative property is the ambiguity of words. Ambiguity can be understood as a sense relation in which an expression either in a lexical or syntactical level tend to have two or more than two interpretations of meaning in different contexts. Ambiguity for a human interpreter is not a big deal but is a major challenge in the field of Artificial Intelligence (AI). Human with a vast and conscious understanding of real world can effortlessly interpret any kind of ambiguous expressions within a fraction of time but the same expression become tedious in the case of computer programming.

Natural Language Processing (NLP) is one of the sub-fields of AI which deals with the automatic processing or programming of a language with a view to enhance computer to conduct various automatic tasks like parsing, semantic interpretation, machine translation, speech synthesis, information retrieval, question answering, text mining, etc. In this context, Word Sense Disambiguation (WSD) is one of the NLP motivated task by which proper sense of a word in various contexts is determined. It involves the computational program of picking up the intended meaning of a word with the help of predefined set of information derived from different knowledge bases like Machine Readable Dictionaries, Thesauri, WordNet etc. It deals basically with the open class category of words like “noun,” “verb,” “adjective” and “adverb” and each of these categories need different treatment for determining their multiple senses. For instance, contextual information of a word is foremost disambiguation cue and in order to understand a “noun” in its context, one needs to look upon the modifiers of that noun whereas in order to carry same task for verb, argument structure is necessary.

Literature shows that disambiguation of “noun” category has acquired good attention in the field of WSD than the “verb.” Despite of the fact that there is consensus among NLP researchers that disambiguating a verb can assist the disambiguation process of other categories WSD as well. Perhaps, only in the recent years some efforts for verb disambiguation has been initiated.

Verbs are more ambiguous in nature rather than any other word class because verbs in a natural language are somehow finite in number and they possess more than one meaning while using in different context with different nominal. Moreover, use of verbs in phrasal verbs, proverbs and metaphors make them more deviating in their meaning.

2. RELATED WORKS

Till date only a few scarce works are available on Nepali word sense disambiguation. Perhaps, the plenty of works are there on Nepali verbs from traditional and descriptive perspectives. Handful of works are also there on Nepali verbs from the computational perspective.

Firstly, Shrestha, Hall & Bista (2008) is an investigation on the impact of NLP resources like Morphological Analyzer and Machine Readable Dictionary in ambiguity resolution. It also presents an approach of Lesk Algorithm for Nepali WSD using Nepali WordNet containing few sets of noun. It compares accuracy of 50-70% using Lesk Algorithm with the accuracy of 80% through the manual morph analysis. Dhungana, Shakya, Baral & Sharma (2014) proposes a new knowledge based approach for disambiguating Nepali words i.e. using clue words defined as “related context words for each word.” It claims of having 91% of accuracy for disambiguating small scale corpora (201 sentences with 201 polysemous words). Dhungana & Shakya (2014a) is similar study as previous one. It seems an organized projection of same WSD approach using clue words for Nepali. Dhungana & Shakya (2014b) proposes modified adapted Lesk Algorithm for WSD in Nepali. It includes synset, gloss, example and hypernym as well as context window containing all the words except articles, prepositions and pronouns. It claims 88% accuracy while testing 201 sentences using a small scale WordNet consisting of 348 words. Secondly, Adhikary (1980), Pant (1990), Acharya (1991), Pokharel (1991, 1992, 1997), Lohani (1999), Pant (2001), Sharma (2005), Prasain (2008), Pokharel & Adhikari (2010) etc. are descriptive study available on Nepali verbs in general.

3. METHODOLOGY

The primary data for this study has been manually collected from four Nepali dictionaries: *Nepali kriyaaharuko kosh* (dictionary of Nepali verbs) by Mahdav Prasad Poudel; *Nepali brihat sabdakosh* (Nepali Comprehensive dictionary) developed by Nepal Academy; *Ekta Concise Nepali-to-English Dictionary* edited by Shridhar Prasad Lohani and Rameshwar Prasad Adhikari; and *Pragya Nepali-to-English Dictionary* developed by Nepal Academy. These dictionaries are helpful in finding Nepali verbs with multiple senses. Firstly, all the polysemous verbs found in these dictionaries were noted down with their different senses. Secondly, the verbs with more than one sense were listed down and finally, out of those verbs, 10 most frequently used verbs has been sorted out with the help of corpus based frequency analysis. The secondary data for this study has been collected from relevant literatures available in libraries and web sources available.

In this paper, analysis of argument structure of a verb has been carried out on the basis of theoretical framework provided by Murthy & Kesava (2003). It has provided a framework for the information regarding the NOUN arguments in terms of features like +/-h (HUMAN), +/-a (ANIMATE), c (CONCRETE), ca (COMBUSTIBLE ARTICLE), +/-ed (EDIBLE), +/-bp (BODY PART) etc. which are encoded in boxes. According to which, “each sense of the verb and their arguments occurring in the example sentences along with their feature codes in the boxes and their meanings are given as a frame. Each argument with its set of features arranged in the frame looks like {[N<>()]} or {[P<>()]} . The first element inside the brackets can be either N (Noun) or P (Pronoun), the second element enclosed in the <> brackets is the Thematic Relation of the argument with the particular sense of the verb, the next element in the () brackets is a set of features like (+h) which are assigned to that argument. The number of arguments in any frame is dependent on the verb sense. Finally, the meaning or the sense of the verb in ‘s’ has been provided at the end of the frame” (Murthy & Kesava 2003).

4. ANALYSIS AND DISCUSSION

There are almost around 200 verb forms with two or more than two interpretation of meaning in the above mentioned four dictionaries. Among those, ten most frequently used Nepali ambiguous verbs has been analysed here in this paper.

1. **Kaṛnu**

It is a transitive verb which can act as predicator with two or three place arguments. It is a class member of the verbs of obtaining according to Levin (1993: 141). The verb *kaṛnu* is most ambiguous Nepali verb with seven different senses such as “to burrow” (loan), “to call out” (someone’s name in disrespect), “to outcaste” (one from his/her community), “to uproot” (sapling), “to fill” (mouth with saliva), “to take out skin” (especially animals), “to extract” (juice/honey). The sentences below shows the different senses implied by this verb.

- (1a) *hāri-le bānkbāṭā lon kaṛ-jo*
 Hari-ERG bank-ABL loan burrow-3.SG.M.PST
 “Hari burrowed loan from the bank”
- (1b) *hāri-le babu-ko nam kaṛ-jo*
 Hari-ERG father-GEN name call out-3.SG.M.PST
 “Hari called out the name of his father”
- (1c) *samaj-le hāri-lai pani-ma kaṛ-jo*
 society-ERG Hari-ACC water-LOC outcaste-3.SG.M.PST
 “Society outcasted Hari”
- (1d) *hāri-le dhan-ko biu kaṛ-jo*
 Hari-ERG paddy-GEN sapling uproot-3.SG.M.PST
 “Hari uprooted the sapling of paddy”
- (1e) *hāri-le ral kaṛ-jo*
 Hari-ERG saliva fill-3.SG.M.PST
 “Hari filled his mouth with saliva”
- (1f) *hāri-le bakhra-ko chala kaṛ-jo*
 hari-ERG goat-GEN skin take out-3.SG.M.PST
 “Hari took out the skin of a goat”
- (1g) *hāri-le māuri-ko maha kaṛ-jo*
 Hari-ERG honeybee-GEN honey extract-3.SG.M.PST
 “Hari extracted honey of honeybee”

In order to disambiguate each senses of this verb, a closer examination of argument structure is necessary. Argument Structure of each senses of the verb *kaɽjo* are as follows:

kaɽjo, {[N<Ag>(+h)]# [N<So>(-a,+c)]# [N<Th>(-a,+c)]}; “to burrow”
kaɽjo, {[N<Ag>(+h)]# [N<Pt>(+h)]# [N<Th>]}; “to call out”
kaɽjo, {[N<Ag>(+h)]# [N<Pt>(+h)]# [N<Lo>(-a,-c)]}; “to outcaste”
kaɽjo, {[N<Ag>(+h)]# [N<Th>(-a,+c)]}; “to uproot”
kaɽjo, {[N<Ag>(+h)]# [N<Th>(-a,-c)]}; “to fill”
kaɽjo, {[N<Ag>(+h)]# [N<Th>(+a, bp)]}; “to take out skin”
kaɽjo, {[N<Ag>(+h)]# [N<Th>(-a,+c)]}; “to extract”

The verb *kaɽjo* in the sentences (1a), (1d), (1f), (1g) have the meaning of “take out” in the deep level and the difference in the senses (to burrow, to uproot, to extract) are there due the extension of meaning. Whereas the meaning of verb *kaɽjo* in sentences (1b), (1c), and (1e) have completely different senses (“to call out,” “to fill,” and “to exclude”).

2. Kaɽnu

Another most ambiguous verb in Nepali which is also a transitive verb, a predicator with two place arguments. It has four different senses: “to cut,” “to book” (ticket), “to reject,” and “to issue.” The sentences given below exemplify those senses.

- (2a) *bhai-le aiphɒl kaɽ-jo*
 brother-ERG apple cut-3.SG.M.PST
 “Brother cut down an apple”
- (2b) *bhai-le tren-ko tikɒt kaɽ-jo*
 brother-ERG train-GEN ticket cut-3.SG.M.PST
 “Brother booked the train ticket”
- (2c) *bhai-le mero kura kaɽ-jo*
 brother-ERG My talk reject-3.SG.M.PST
 “Brother rejected my talk”
- (2d) *sarkar-le notis kaɽ-jo*
 government-ERG notice issue-3.SG.PST
 “Government issued a notice”

The Argument Structure of each senses of this verb are as follows:

katjo, {[N<Ag>(+h)]#[N<Th>(-a,+c)]}; “to cut”
katjo, {[N<Ag>(+h)]#[N<Th>(-a,+c)]}; “to book”
katjo, {[N<Ag>(+h)]#[N<Th>(-a,-c)]}; “to reject”
katjo, {[N<Ag>(+h)]#[N<Th>(-a,-c)]}; “to issue”

3. *kamaunu*

It is another verb with ambiguity in Nepali. This verb as seen in (3a) below is causative form of the verb root *kamjo* whereas the root of *kamaunu* in (3b) and (3c) is *kamau*. Hence, there is sense relation of homonymy of *kamjo* in (3a) and (3c). The verb *kamjo* is an intransitive verb whereas *kamau* is transitive one. The causativization process has increased the valency of verb *kamjo* as seen in the (3a). It has three senses: “to make shiver,” “to earn,” and “to lease out” (especially land).

Examples:

- (3a) *ḍaṛo-le hari-lai kama-jo*
 cold-ERG Hari-ACC shiver-3.SG.M.PST.CAUS
 “Cold made Hari to shiver”
 (3b) *hari-le pāisa kama-jo*
 Hari-ERG money earn-3.SG.M.PST
 “Hari earned money”
 (3c) *hari-le mero khet kama-jo*
 Hari-ERG my [paddy field] lease-1.SG.M.PST
 “Hari leased out my paddy field”

The Argument Structure of this verb are as follows:

kamajo, {[N<Fo>(-a)]#[N<p>(+h)]} “to make shiver”
kamajo, {[N<Ag>(+h)]#[N<Th>(-a,+c)]} “to earn”
kamajo, {[N<Ag>(+h)]#[N<Th>(-a,+c)]} “to lease out”

4. *khəsnu*

One of the important disambiguation cue for the two senses of verb *khəsnu* is their transitivity. *khəsjo* in the (4a) is a transitive verb with two place argument. Whereas *khəsjo* in (4b) is an intransitive verb with only one place argument. It has two senses like “to fall down” and “to die.”

Examples:

- (4a) *dhunga kha_{AS}-jo*
 Stone fall-3.SG.M.PST
 “Stone fell down from the above”
- (4b) *us-ko bhai kha_{AS}jo*
 He-GEN brother die-3.SG.N.PST
 “His brother died”

The Argument Structure of this verb are as follows:

kha_{AS}jo, {[N<Th>(-a,+c)]} “to fall down”
kha_{AS}jo, {[N<Pt>(+h)]} “to die”

5. **udaunu**

The verb *udaunu* has two senses like “to rise up” and “to appear.” It also exhibits difference in the transitivity. *udajo* in (5a) is an intransitive verb whereas it is a transitive verb in the sentence (5b).

Examples:

- (5a) *gham uda-jo*
 sun-NOM rise-3.SG.PST
 “Sun rised”
- (5b) *ram-ko jiwana-ma khusijali uda-jo*
 Ram-GEN life-LOC happiness appear-3.SG.M.PST
 “Happiness appeared in Ram’s life”

The Argument Structure of this verb are:

udajo, {[N<Th>(-a,-c)]}
udajo, {[N<Pt>(-a,-c)#[N<Th>(-a,-c)]}

6. **haraunu**

As in the case of the verb *kamaunu*, there is difference in the root form of the verb *haraunu*. The root form of *haraunu* in (6a) is *haraui* whereas the verb *haraunu* in (6b) is derived from the noun stem i.e. *har*: Hence it is also a case of homonymy. It has two senses: “to lose something,” and “to defeat someone.”

Example:

- (6a) *hΛri-le bhai-ko kitab hΛra-jo*
 Hari-ERG brother-GEN book lose-3.SG.PST
 “Hari lost brother’s book”
- (6b) *ram-le rΛvan-lai yuddhΛ-ma hΛra-jo*
 Ram-ERG ravan-ACC war-ABL defeat-3.SG.PST
 “Ram defeated Ravan in the war”

The Argument Structure of this verb are given below:

hΛra-jo, {[N<Ag>(+h)]#[N<Th>(-a,+c)]} “to lose”
hΛra-jo, {[N<Ag>(+h)]#[N<Pt>(+h)]#[N<Lo>(-a,-c)]} “to defeat”

7. **utarnu**

The verb *utarnu* has two completely different senses: “to draw a picture” and “to unload.” It is a transitive verb with three place arguments.

Examples:

- (7a) *hΛri-le mero nΛksa utar-jo*
 Hari-ERG my picture draw-3.SG.M.PST
 “Hari drew my picture”
- (7b) *hΛri-le gaḍi-baṭΛ saman utar-jo*
 Hari-ERG vehicle goods unload-3.SG.M.PST
 “Hari unloaded goods from the vehicle”

The Argument Structure of this verb is as follows:

utar-jo, {[N<Ag>(+h)]#[N<Th>(-a,+c)]} “to draw”
utar-jo, {[N<Ag>(+h)]#[N<So>(-a,+c)]#[N<Th>(-a,+c)]} “to unload”

8. **sarkinu**

The verb *sarkinu* is another example of different transitivity. *sarkijo* in (8a) is an intransitive whereas in (8b), it is a transitive verb. It has two senses: “to cough” and “to become sarcastic.”

Examples:

- (8a) *bhaṭ khΛḍa hΛri sΛrki-jo*
 rice eat-CONJ Hari-NOM choke-3.SG.M.PST
 “Hari choked up while eating rice”

- (8b) *aphno vijajA-prati ram sarki-jo*
 own victory-About Ram boast-3.SG.M.PST
 “Ram boasted about own victory”

The Argument Structure of this verb is given below:

sarkijo, {[N<Th>(-a,-c)]#[N<Ex>(+h)]} “to choke up”
sarkijo, {[N<Th>(-a,-c)]#[N<Ag>(+h)]} “to boast”

9. *bhjaunu*

The verb *bhjaunu* is a transitive verb with two different collocative senses: “to manage time” and “to finish”

Examples:

- (9a) *hari-le sabai kam_ bhja-jo*
 Hari-ERG all work finish-3.SG.M.PST
 “Hari finished all works”
 (9b) *hari-le bAdgar dganA bhja-jo*
 Hari-ERG market go-INF manage-3.SG.PST
 “Hari managed to go market”

The Argument Structure of this verb are as follows:

bhja-jo, {[N<Ag>(h)]#[N<Th>(-a,-c)]} “to finish”
bhja-jo, {[N<Ag>(h)]#[N<Th>(-a,+c)]#[N<Lo>(-a,+c>)]} “to manage”

10. *khijaunu*

The verb *khijaunu* is a pure example of ambiguous verb with completely two different meanings. It is a transitive verb with two senses: “to row boat” and “to make thin.”

- (10a) *hari-le nau khija-jo*
 hari-ERG boat row-3.SG.M.PST
 “Hari rowed the boat”
 (10b) *hari-le jiu khija-jo*
 Hari-ERG body [make thin]-3.SG.M.PST.CAUS
 “Hari made his body thin”

The Argument Structure of this verb are as follows:

khijajo, {[N<Ag>(+h)#[N<Th>(-a,+c)]]} “to row boat”
khijajo, {[N<Ag>(+h)#[N<Th>(+bp)]]} “to make thin”

5. CONCLUSION

Ambiguity is one of the major issues discussed in the literature of Natural Language Processing. It occurs in grammatical as well as non- grammatical levels in a language. The lexical level of ambiguity with due reference to the Nepali verbs has been analysed here in this paper. The Nepali verbs like *kaṭnu*, *kaḍnu*, *kəmaunu*, *khəsnu*, *udaunu*, *həraunu*, *utənu*, *sarkinu*, *bhjaunu* and *khijaunu* have been identified as the ten most frequently used ambiguous verbs in Nepali. They are provided with their argument structure based on the syntactic and semantic features in this paper which can pave the way for further process of Nepali Verb Sense Disambiguation.

REFERENCES

- Acharya, J. 1991. *A Descriptive Grammar of Nepali and an Analyzed Corpus*. Washington, D.C: Georgetown University Press
- Adhikary, R. P. 1980. ch- as a non-past marker in Nepali. *Nepalese Linguistics*. Vol. 1. Kathmandu, Nepal: Linguistic Society of Nepal.
- Basnet, S. B. & Pandey, S. B. 2009. Morphological analysis of verbs in Nepali. *Nepalese Linguistics*. Vol. 24. Kathmandu, Nepal: Linguistic Society of Nepal
- Dhungana, U. R., Shakya, S., Baral, K. & Sharma, B. 2014. Word sense disambiguation using WSD specific WordNet of polysemy words. *International Journal on Natural Language Computing*, 3/4. Available online: <<http://airccse.org/journal/ijnlc/papers/3414ijnlc05.pdf>>.
- . & Shakya, S. 2014a. Word sense disambiguation using clue words. *Journal of the Institute of Engineering*, 10/ 1, 192-198. Available online: <<http://www.nepjol.info/index.php/JIE/issue/view/716>>.
- . & Shakya, S. 2014b. Word sense disambiguation in Nepali language. In *Proceedings of Fourth International Conference on Digital Information and Communication Technology and its Applications*, University of Thai Chamber and Commerce, Bagkok, Thailand, 6-8 May. Available online: <<http://ieeexplore.ieee.org/document/6821655/>>.

- Levin, B. 1993. *English Verb Classes and Alternations: A Preliminary Investigation*. Chicago: The University of Chicago Press
- Lohani, R. R. 1999. Permissive constructions in Nepali: Evidence from complex predicates. *Nepalese Linguistics*. Vol. 16. Kathmandu, Nepal: Linguistic Society of Nepal
- . 2001. Valency increase in Nepali predicates: The processes of transitivity and causativization. *Nepalese Linguistics*. Vol. 18. Kathmandu, Nepal: Linguistic Society of Nepal
- Lohani, S. P. & Adhikari, R. P. (eds.) 2010. *Ekta Concise Nepali-to-English Dictionary*. Kathmandu, Nepal: Ekta Books.
- Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D. & Miller, K. 1990. Introduction to WordNet: An on-line lexical database. *International Journal of Lexicography*, 3/4, 235-244.
- Murthy, M. & Kesava, C. 2003. *The Argument Structure of Telugu Verbs*. Hyderabad: The University of Hyderabad Dissertation
- Pant, G. R. 2001. Logical structure of Nepali verbs. *Nepalese Linguistics*. Vol. 18. Kathmandu, Nepal: Linguistic Society of Nepal
- Parajuli, K. P. (ed.) 2010. *Nepali Brihat Sabdakosh*. 7th ed. Kathmandu, Nepal: Nepal Academy.
- Pokharel, M. P. 1992. Moods in Nepali. *Nepalese Linguistics*. Vol. 9. Kathmandu, Nepal: Linguistic Society of Nepal.
- . 1991. Compound verbs in Nepali. *Contribution to Nepalese Studies*. Vol.18:2. Kathmandu, Nepal: CNAS, Tribhuvan University
- . 1997. *Nepali Vakya Vyakaran*. Kathmandu, Nepal: Ekta Publishing House
- . 2015. *Nepali Kriyadaruko Kosh*. Kathmandu, Nepal: Vidhyarthi Prakashan Private Limited.
- Poudyal, N. P. 2009. The syntactic alignments across three-argument verbs in Nepali. *Nepalese Linguistics*, Vol. 24. Kathmandu, Nepal: Linguistic Society of Nepal
- Prasain, B. 2008. Computational analysis of Nepali basic verbs (written forms). *Nepalese Linguistics*, Vol. 23. Kathmandu, Nepal: Linguistic Society of Nepal
- . 2011. Computational Analysis of Nepali Morphology: A Model for Natural Language Processing. Unpublished PhD Dissertation, Central Department of Linguistics, Tribhuvan University, Kathmandu, Nepal.
- Roy, A., Sarkar, S. & Purkayastha, B. S. 2014. Knowledge based approaches to Nepali Word sense disambiguation. In *International Journal on Natural Language Computing*, 3/3. Available online: <<http://airccse.org/journal/ijnlc/papers/3314ijnlc05.pdf>>.

- Shrestha, N. P., Hall, A. V. & Bista, S. K. 2008. Resources for Nepali word sense disambiguation. In *Proceedings of International Conference on Natural Language Processing and Knowledge Engineering*, Beiyou Building of Science and Technology, Beijing, China, 19-22 Oct.
- Upreti, G. P., Bandhu, C., Yadava, Y. P., Pokharel, M. P. & Sharma, K. C. (Eds.), *Pragya Nepali-to-English Dictionary*. Kathmadu, Nepal: Nepal Academy

KRISHNA MAYA MANGER
CENTER FOR APPLIED LINGUISTICS
AND TRANSLATION STUDIES,
UNIVERSITY OF HYDERABAD, HYDERABAD,
TELENGANA 500046, INDIA,
E-MAIL: <KRISHNAMANGER@GMAIL.COM>
PH: 9593977637

Development of Verb Frames for Nepali

Krishna Maya Manger, M.A. (Nepali), M.A. (Linguistics)

Assistant Professor
Department of Nepali
University of North Bengal,
Siliguri, West Bengal, India, 734013
krishnamanger@gmail.com

=====

Abstract

This paper describes the method and procedures of building Verb Frames for the Nepali language which is developed as a part of doctoral research. A total of 486 Verb Frames have been developed for 200 ambiguous Nepali verbs following the theoretical framework provided by Begum (2017) with some modifications. Our Verb Frame captures lexical, syntactic and semantic information about each sense of a particular verb with due focus on its argument structure and ontology. The significance of Verb Frames is seen in the area of NLP, especially in Parsing and Word Sense Disambiguation. It provides a detailed description of the linguistic attributes of Nepali verbs for scholars who are interested in studying verbs in the language.

Keywords: Nepali, Verbs, Ambiguity, Verb Frames, Argument Structure

1. Introduction

The verb is a core grammatical category of a language that plays a pivotal role in determining the functions of each argument in a sentence. It is a central element in a sentence without which the structure of a sentence seems impossible. That is why the database of verbs which provides linguistic information about a verb has become a major focus in the field of automatic processing of a language at present. There are different approaches for building linguistic resources which could capture all the necessary information about a verb 'ranging from phonological and morphological to syntactic, semantic and pragmatic criteria' (Walde, 2009). In

this context, Verb Frame is a kind of knowledge base which can be used as a linguistic resource, particularly in the field of computational linguistics and natural language processing.

The origin of the **concept of verb frame** is presumed to be started from the work of Chomsky (1965) which presented the idea of a sub-categorization frame for the first time. The sub-categorization according to Chomsky denotes 'the ability/necessity for lexical items (usually verbs) to require (allow the presence and types of the syntactic arguments with which they co-occur' (Chomsky, 1965). The notion of sub-categorization looks similar to the notion of valency in the sense that both account for the number and status of arguments in a sentence. But the two are different since sub-categorization in its original meaning did not include the subject.

Though modern theories include the subject in the sub-categorization frame as well, whereas valency was perceived as the number of arguments including the subject from the very beginning. However, this study takes the Verb Frame in a particular sense which has its root in the work of Levin's classification of English verbs (Levin, 1993) which has inspired later development of the concept in many different ways. VerbNet (Schuler, 2005), FrameNet (Baker et al, 1998), PropBank (Palmer et al, 2005), and WordNet (Miller, 1995) are the works which have contributed in the development of the concept, especially for natural language processing. Verb Frame in this study, is a tabular representation that 'captures linguistic information about the syntactic distribution of a verb in a language' (Begum, 2017).

2. Related Work

Levin (1993) presents a large-scale classification of English Verbs which is based on the syntactico-semantic correlation of verbal behavior. It was developed under the Lexicon project of the Centre for Cognitive Science, MIT. Levin assumes that 'the behavior of a verb, particularly for the expression and interpretation of its arguments is to a large extent determined by its meaning' and 'the verb meaning is a key to verb behavior'.

VerbNet (Schuler, 2005) is an online lexicon of English verbs developed under a research project led by Martha Palmer at the University of Colorado. It is a hierarchical domain-independent broad coverage verb lexicon with mappings to other lexical resources such as WordNet, Xtag and FrameNet'. It aims to refine and add subclasses to Levin's verb classes in order 'to achieve syntactic

and semantic coherence among members of a class'. Each class of VerbNet contains two types of information about a verb. They are i) syntactic description that depicts thematic roles and selectional restrictions of a verb; and ii) syntactic frames which represent syntactic description of a verb with semantic predicates that shows temporal function.

FrameNet (Baker et al, 1998) is a human and machine-readable lexical database of English developed by the International Computer Science Institute, Berkeley. It is a kind of semantic role labelling that attempts to represent roles in the frame. It contains more than 13,000-word senses with annotated examples. It has more than 200,000 manually annotated sentences linked to more than 1,200 semantic frames which provide a unique training dataset.

PropBank (Palmer et al (2005) is an acronym for 'Proposition Bank' which contains a corpus annotated with verbal propositions and their arguments. It provides resources of sentences annotated with semantic roles which are defined for an individual verb sense where each sense of each verb has a specific set of roles such as Arg0, Arg1, Arg2 and so on.

WordNet (Miller, 1995) is a lexical database of English which groups different words (nouns, adjectives, verbs, adverbs) into sets of cognitive synonyms called synsets. Each synset expresses a distinct concept and is interlinked through conceptual semantic and lexical resources which are navigated through the browser. Synonymy is the main relation among words in the WordNet. However, it also encodes the sense relation of hyperonymy, meronymy, troponymy and antonymy.

Begum (2017) is a doctoral dissertation entitled *Developing a Pilot Hindi Treebank based on Computational Paninian Grammar* submitted at IIIT Hyderabad. In a chapter of her thesis, Rafiya Begum has developed a database of 486 Verb Frames for 300 verbs for Hindi which is one of the pioneering works in the field in India. The model of verb frame presented by her is based on the Paninian Grammatical Framework which has represented the linguistic information (description and dependency relation) about a verb in tabular form. Her work is the main motivation behind this research. The theoretical framework and the methodology devised here followed her work in particular.

3. Methodology

The corpus of 200K sentences was collected using various written sources from personal effort as well as from various NLP projects like the Indian Languages Corpora Initiative conducted in consortium mode led by Jawaharlal Nehru University, New Delhi and Shallow Parser Tool for Indian Languages led by Center for Applied Linguistics and Translation Studies, the University of Hyderabad funded by MeiT, Government of India. The written corpus from different domains like entertainment, health, sports, news, literature, tourism, and agriculture was collected from both projects. The corpus for the domain of literature is also collected from the website of samakalinsahitya.

The data for this study were extracted from the corpus of the aforementioned sources. The first step of extracting data was to sort out all verb occurrences from 200K sentences. Secondly, unique verb forms looking at their roots were separated. Thirdly, looking at the frequencies of their occurrence, 200 verbs were chosen for the study. A total of 486 unique verb frames have been developed for each sense of 200 ambiguous verbs. The verb senses are verified using the corpus and Nepali dictionaries available in print forms like Poudel (2015), Parajuli (2010) and Lohani & Adhikari (2010). The example sentences for each sense of verbs are extracted from the corpus and Verb Frames are developed following the theoretical model presented by Begum (2017) with some modifications. They contain the dependency relations of the mandatory and desirable arguments with theta role and ontological information.

Ambiguity is the primary criterion for the selection of data which means Nepali verbs having two to 15 different senses are selected for the study. Frequently used verbs with more obvious ambiguous senses are preferred for making verb frames. Verb forms with simple past tense are incorporated in this study since each tense and aspect shows different morphology which can be resulted in different senses of a verb. Also, the motive behind this study is to present a small database of verb frames for Nepali which would help to build a large-scale database in future.

4. Components of Verb Frames

The Verb Frame in this study contains two kinds of information about a particular verb such as description and the verb frame. These two pieces of information are provided in a data file

which is referred to as a 'verb entry' by Begum (2017: 95). The description part of the verb frame includes the information like Verb name, Sense-id, Verb Sense, English gloss, Verb in the same class, Example sentence, Theta roles and Frame ID whereas verb frame consists the information like Arc-label, Necessity of the argument, *Vibhakti*, Lexical Type and Ontology. Since, Indian languages like Nepali are Subject Object Verb (SOV) dominant phrase order language and the position of arguments to their predicate always remains left in almost all cases and since the dependency grammar treated arguments as children of a verb, their relation with arguments also remain constant in all cases, the fields 'position' and 'relation' are removed in study which are incorporated in the verb frame developed by Begum (2017). Also, new fields like 'verb class' and 'ontology' are added for this kind of information about a verb found to be helpful in identifying a particular sense of a verb in the context.

Figure 4.1 represents the sample of verb frame developed for the Nepali verb *ukas* which has three senses: 'pull out', 'provoke' and 'bail out'. The figure below demonstrates each component of the verb frame for the third sense of the verb *ukas*.

Figure 4.1 Sample of Verb Frame

Verb:: <i>ukas</i> SID:: <i>ukas%VT%S3</i> Verb_Sense:: <i>mukta_gar</i> Eng_Gloss:: Bail Verb_class:: Verb of putting with specified direction Verbs_in_Same_Class:: Synonyms> <i>uker%VT%S2%FID2</i> Frames:: Frame_Name_3:: Ex:: <i>us-le dʒel-baʈa ʌparadhi-lai ukas-jo</i> He-ERG jail-ABL culprit-DAT bail-3.SG.NPST 'He bailed the culprit out from the jail' Theta_Roles:: AGENT SOURCE PATIENT VERB Demand_Frame:: Frame_ID_3 Frame ID:: <i>ukas%VT%S3%FID3</i>				
arc-label	necessity	vibhakti	lextype	ontology
k1	m	0	p	[+hum]
k5	d	baʈa	n	[+artfplc]
k4	m	lai	n	[+hum,+rol]

As we see in Figure 4.1 the first field is **Verb**, the name of a particular verb in Nepali which is written in IPA for its appropriate pronunciation.

The second field is **Sense ID (SID)** which is represented as *ukas%VT%S3*. It is a unique identification number for the particular sense of a verb. It consists of verb name, verb type and sense number which are separated by the symbol of percentage (%). In the verb frame above, *ukas* is the verb name, 'VT' is the 'transitive verb' and 'S3' is the 'Sense 3' of the verb respectively. The verb type is of three kinds: Transitive Verb (VT), Di-transitive Verb (VDT) and Causative Verb (VCAUS). Different Senses of a verb are denoted by the convention of S1, S2, S3 and so on according to the number of senses a verb can have.

Verb Sense is the third field which represents each sense of a given verb in Nepali. For example, in Figure 4.1 the third sense of the verb *ukas* is *muktA_gAr* 'bail out'.

The fourth field is **English Gloss** which gives the meaning of a verb in the English language. It is represented as 'Eng_Gloss' in the verb frame. In Figure 4.1, 'Bail out' is an English gloss for a particular sense of the verb *ukas*.

The fifth field represents **Verb Class** which is a semantic class of a verb based on Levin (1993). In Figure 4.1 above, 'bail out' falls under the semantic class of 'Verb of putting with a specified direction'. This kind of information is included because it helps to provide insights into the disambiguation of Nepali verbs since the verbs having the same semantic class behave similarly and have similar argument structures. It also helps to categorize Nepali verbs in certain semantic classes.

The **Verb in the Same Class** is the sixth field which provides information about the synonymous verb frame found in the lexicon. It means that there are several verb frames which have similar senses which are considered synonymous with each other or the verbs in the same class. The verbs in the same class are represented by the unique frame ID they possess. The information about the same class is necessary to see whether or not the verbs in the same class behave similar way.

The seventh field is **Example** which is an example sentence for each sense of a verb which is denoted by the convention of 'Ex'. The example is given in Nepali which is transcribed in IPA. The conventions and the methods of glossing are written following interlinear morpheme-by-

morpheme glossing provided by Leipzig Glossing Rules (2015). The following example is given for the verb *ukas* in Figure 4.1 above:

<i>us-le</i>	<i>ɖʒel-baʈʌ</i>	<i>ʌpʌradhi-lai</i>	<i>ukas-jo</i>
He-ERG	jail-ABL	culprit-DAT	bail-3.SG.NPST

'He bailed the culprit out from the jail'

The eighth field in the verb frame is **Theta Role** which refers to the 'role performed by each argument of a predicate, defined regarding a restricted universal set of thematic functions' (Crystal, 2003: 463). The definitions provided by VerbNet (2006) have been taken as the model for annotating theta roles of each argument.

Frame Name, **Demand Frame** and **Frame ID** are the fields which are all about naming and giving a unique ID to each verb frame for each sense of a particular verb. In Figure 4.1, *Frame_Name_3* is the frame name, *Frame_ID_3* is Demand Frame and *ukas%VT%S3%FID3* is frame ID for the third sense of the verb *ukas*. The frame ID consists of two parts: Sense Id (e.g. *ucal%VT%S1*) and the Frame Number (e.g. *FID1*) which are separated by a percentage sign (%). It helps to identify a particular frame for a particular sense of a verb.

The term '**verb frame**' is used in two senses in this study as in Begum (2017). In a broad sense, it refers to the whole data, i.e., the verb entry (both 'description of the verb' and 'verb frame') and in a particular sense it refers to the tabular form which is a verb frame. 'The actual verb frame is the table given in the verb entry' (Begum, 2017: 99) which represents five pieces of information such as Arc Label, Necessity, Vibhakti, Lexical Type and Ontology.

The first field in the verb frame is the **Arc Label** which represents the dependency relation or *karaka* relation of arguments. Conventions used for each type and subtypes of *karakas* are based on 'AnnCorra: TreeBanks for Indian Languages Guidelines for Annotating Hindi Tree Bank' developed by IIIT, Hyderabad (Bharati et al, 2012). Figure 4.1 demonstrates the method of representing *karaka* information in the verb frame. The sentence given in the figure has three arguments: *k1 (us-le)*, *k5 (ɖʒel-baʈʌ)* and *k2 (ʌpʌradhi-lai)* respectively and they are captured in the first field i.e. Arc Label.

The second field is **Necessity** which is the information about the necessity of argument in the sentence. In the example sentence given in Figure 4.1, *us* 'he', *dʒel* 'jail' and *aparadhi* 'culprit' are three arguments of the predicate *ukas-jo* 'bailed out' which have theta roles of 'agent', 'source' and 'patient' respectively.

The third field is **Vibhakti** which provides information about the type of case marker or any other postpositional element that comes with a particular argument. In the example sentence, as given in figure 4.1, there are three vibhaktis –*le* (ergative), –*baʈa* (ablative), –*lai* (dative) attached with each argument (*us*, *dʒel*, *aparadhi*) respectively.

The fourth field is **Lexical Type** which contains information about parts of speech of an argument.

Finally, the fifth field consists of **Ontology** which captures information about the semantic properties of the concepts and their relationship with each other. It is useful in determining selectional restrictions of a verb which further helps to disambiguate words having more than one sense. The tag set prepared for 'Shared Task cum Workshop on OntoLex in Indian Languages' by the Center for Applied Linguistics and Translation Studies, University of Hyderabad which was held from 29th November to 1st December 2017 has been used for annotating ontologies of arguments in this study. The ontology for the arguments of the example sentence given in Figure 4.1 are marked as +hum (human) for *us* 'he', +artcplc (artefact place) for *dʒel* 'jail' and +hum, +rol (human, role) for *aparadhi* 'culprit' respectively.

5. Conclusion

Nepali is in the primary stage of natural language processing and is considered one of the least-resourced languages in terms of NLP research and development. Thus, the development of linguistic resources in every aspect is necessary for this language. In this regard, this study is a small step towards building a knowledge base of verbs which can pave the way for the larger database in future.

Verb frames as a linguistic resource have four basic implications such as i) it is useful as a database for knowledge-based NLP applications; ii) it is helpful to understand the verbal behavior of Nepali verbs; iii) it can be used as a tool for annotating (especially parsing) Nepali verbs; and

iv) it can be used as a tool for word sense disambiguation. Though, this study embodies a very less and limited number of verbs, the verb frames developed through this study can be used as a model of a larger database of verb frames for Nepali which would contribute to the field of NLP tasks and applications like parsing, word sense disambiguation and machine translation.

=====

References

- Baker, Collin F, Charles J. Fillmore and John Lowe. (1998). 'The Berkeley FrameNet Project'. In *COLIN '98*. Proceedings of the conference. Retrieved from <https://aclanthology.org/P98-1013.pdf>
- Begum, Rafiya. (2017). *Developing a Pilot Hindi Treebank Based on Computational Paninian Grammar*. An unpublished PhD Dissertation. International Institute of Information Technology, Hyderabad, Telangana
- Chomsky, N. (2015). *Syntactic Structures* (2nd Edition). New York: Mouton De Gruyter
- Crystal, David. (2003). *A Dictionary of Linguistics and Phonetics* (5th Edition). Hoboken, New Jersey: Blackwell Publishing
- Levin, Beth. (1993). *English Verb Classes and Alternations*. Chicago: University of Chicago Press
- Lohani, Shridhar Prasad and Rameshwar Prasad Adhikari (Eds.). (2010). *Ekta Concise Nepali-to-English Dictionary*. Kathmandu, Nepal: Ekta Books
- Palmer Martha, Dan Gildea, Paul Kingsbury. (2005). 'The Proposition Bank: A Corpus Annotated with Semantic Roles'. In *Computational Linguistics Journal*, Vol. 31:1. Retrieved from <https://aclanthology.org/J05-1004.pdf>
- Parajuli, Krishna Prasad (Ed.). (2010). *Nepali Brihat Sabdakosh*, (7th Edition). Kathmandu, Nepal: Nepal Academy
- Poudel, Madhav Prasad. (2015). *Nepali Kriyahaaruko Kosh*. Kathmandu, Nepal: Vidhyarthi Prakashan Private Limited

=====

Language in India www.languageinindia.com ISSN 1930-2940 23:6 June 2023

Krishna Maya Manger, M.A. (Nepali), M.A. (Linguistics)

Development of Verb Frames for Nepali

Schuler, Karin Kipper. (2005). *VerbNet: A broad-coverage, comprehensive verb lexicon*.
Retrieved from <https://repository.upenn.edu/dissertations/AAI3179808>

=====

Verb Sense Disambiguation using Verb Frames

by Krishna Maya Manger



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

Submission date: 25-Sep-2023 04:41PM (UTC+0530)

Submission ID: 2176307258

File name: Krishna_Maya_Manger.pdf (2.05M)

Word count: 25140

Character count: 137253

Verb Sense Disambiguation using Verb Frames

ORIGINALITY REPORT

8%

SIMILARITY INDEX

8%

INTERNET SOURCES

2%

PUBLICATIONS

1%

STUDENT PAPERS

PRIMARY SOURCES

1

www.languageinindia.com

Internet Source

4%

2

mafiadoc.com

Internet Source

1%

3

www.tdil-dc.in

Internet Source

<1%

4

www.dravidianuniversity.ac.in

Internet Source

<1%

5

docnum.univ-lorraine.fr

Internet Source

<1%

6

link.springer.com

Internet Source

<1%

7

verbs.colorado.edu

Internet Source

<1%

8

hdl.handle.net

Internet Source

<1%

9

specrada.chnu.edu.ua

Internet Source

<1%

10	aclweb.org Internet Source	<1 %
11	basiceducation.uomosul.edu.iq Internet Source	<1 %
12	propertibazar.com Internet Source	<1 %
13	repository.smuc.edu.et Internet Source	<1 %
14	www.cram.com Internet Source	<1 %
15	repository.lib.cuhk.edu.hk Internet Source	<1 %
16	www.ldcil.org Internet Source	<1 %
17	Submitted to University of Hyderabad, Hyderabad Student Paper	<1 %
18	"Data Science", Springer Science and Business Media LLC, 2021 Publication	<1 %
19	tuprints.ulb.tu-darmstadt.de Internet Source	<1 %
20	docplayer.net Internet Source	<1 %

21	esd.mit.edu Internet Source	<1 %
22	en.wikipedia.org Internet Source	<1 %
23	"Handbook of Linguistic Annotation", Springer Science and Business Media LLC, 2017 Publication	<1 %
24	Niladri Sekhar Dash. "Language Corpora Annotation and Processing", Springer Science and Business Media LLC, 2021 Publication	<1 %
25	etd.aau.edu.et Internet Source	<1 %
26	ojs.ub.uni-konstanz.de Internet Source	<1 %
27	stars.library.ucf.edu Internet Source	<1 %
28	Lecture Notes in Computer Science, 2014. Publication	<1 %
29	M. Sreenivasa Rao, Arun K. Pujari. "Word Sense Disambiguation Using the Hopfield Model of Neural Networks", Journal of Intelligent Systems, 1999 Publication	<1 %

30

Submitted to University of London External System

Student Paper

<1%

Exclude quotes On

Exclude matches

< 14 words

Exclude bibliography On