Indian Traditional Perspective On Discourse Analysis

A dissertation submitted during 2025 to the University of Hyderabad

in partial fulfillment of the award of the degree of

Doctor of Philosophy

in

Sanskrit Studies

by

Vaze Saee Chandrashekhar

20HSPH03

under the guidance of

Prof. Amba Kulkarni

Department of Sanskrit Studies

Department of Sanskrit Studies

School of Humanities

University of Hyderabad

Hyderabad

2025

CERTIFICATE

This is to certify that the thesis entitled **Indian Traditional Perspective On Discourse Analysis** submitted by **Vaze Saee Chandrashekhar** bearing registration number **20HSPH03** in partial fulfilment of the requirements for the award of **Doctor of Philosophy** in the **School of Humanities** is a bonafide work carried out by her under my 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 journal publications before submission of the thesis for adjudication and has produced evidence for the same in the form of acceptance letter or the reprint in the relevant area of his research:

1. **Indian Perspective on Pronominal Anaphora Resolution**, Ushati, the Ganganatha Jha Campus, Central Sanskrit University, Ganganath Jha Campus, Prayagraj, India, ISSN No: 2277-680X, (Accepted for 22nd issue)
   (Chapter 3 of this dissertation contains the contents predominantly from this article.)

and has made presentations in the following conferences, the presented paper being published in the conference proceedings,

1. **Inter-Sentential Discourse Relations**, In Proceedings of the 7th International Sanskrit Computational Linguistics Symposium, pages 67–83, Auroville, Puducherry, India, Association for Computational Linguistics. (ISBN No: 979-8-89176-027-1), 2024. (Chapter 3 of this dissertation contains the contents predominantly from this article.)

2. **Inter Sentential Discourse Relations: An Overview- IGT Perspective**, 1st Samiksha - National Seminar on Sanskrit-IKS and ICT, National Institute of Advance Studies, Bangalore, Department of Science and Technology, Government of India, and Samskriti Foundation, Mysuru, India, 2022.

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

No.	Course-code	Code-title	Credits	Pass/Fail
1.	SK801	Research Methodology	4	PASS
2.	SK814	Readings in Navya Nyaya	4	PASS
3.	EN875	Research and Publication Ethics	2	PASS
4.	SK813	Indian and Western Logical Systems	4	PASS

Ambapp 15.12.25

Supervisor

Department of Sanskrit Studies

Professor
Department of Sanskrit Studies
University of Hyderabad
P.O. Central University
Hyderabad-500 046,TS

Ambapp 15.12.25

Head of the Department

Department of Sanskrit Studies

HEAD
Department of Sanskrit Studies
School of Humanities
University of Hyderabad
Central University P.O.
HYDERABAD-500 046

15/12/25

Dean of school

School of Humanities,

DEAN
SCHOOL OF HUMANITIES
UNIVERSITY OF HYDERABAD
HYDERABAD-500 046. T.S

DECLARATION

I, **Vaze Saee Chandrashekhar**, hereby declare that the work embodied in this dissertation entitled **"Indian Traditional Perspective on Discourse Analysis"**, is submitted by me under the guidance and supervision of **Prof. Amba Kulkarni**, Sr. Professor, Department of Sanskrit Studies, University of Hyderabad, Hyderabad, is a bonafied research work and has not been submitted for any degree in part or in full to this University or any other University. I hereby agree that my thesis can be deposited in Shodhganga/INFLIBNET.

A report on plagiarism statistics from the University Librarian is enclosed.

Date: 15/12/2025
Place: Hyderabad

Name: VAZE SAEE CHANDRASHEKHAR
Signature of the Student:
Regd. No.: 20HSPH03

Amba P 15.12.25
Signature of the Supervisor
Department of Sanskrit Studies
University of Hyderabad
P.O. Central University
Hyderabad-500 046,TS

INDIAN TRADITIONAL PERSPECTIVE ON DISCOURSE ANALYSIS

by

Vaze Saee Chandrashekhar

20HSPH03

under the guidance of

Prof. Amba Kulkarni

Sr. Professor

Department of Sanskrit Studies

School of Humanities

University of Hyderabad

Hyderabad

Relevance of the dissertation to the United Nations Sustainable Development Goals (SDGs)

This thesis addresses:

SDG 4: Quality Education

SDG 9: Industry, Innovation and Infrastructure.

This thesis, titled INDIAN TRADITIONAL PERSPECTIVE ON DISCOURSE ANALYSIS, contributes meaningfully to several Sustainable Development Goals (SDGs) declared by the United Nations in 2015 by fostering culturally grounded knowledge, promoting linguistic diversity, and enabling inclusive technological development.

SDG 4: Quality Education- focuses on fostering inclusive, contextually relevant, and culturally grounded learning systems. This thesis directly contributes to SDG 4 by rigorously reviving and systematizing Indian grammatical and philosophical frameworks- such as *Mīmāṃsā, Vyākaraṇa,*

Nyāya, and *Sāhityāśāstra-* for application to modern discourse analysis. Through detailed study of concepts like *ekavākyatā, saṅgati, tātparya, vivakṣā,* and *tantrayukti*, the work demonstrates how classical Indian theories address language interpretation across sentences, speaker intention, referent identification, and thematic unity. These insights are organized into computationally usable principles for tasks such as pronoun resolution and discourse-marker interpretation, showing that traditional models can generate a precise, structured understanding of text. In doing so, the research strengthens intellectual continuity, protects linguistic heritage, enriches pedagogy, and enables future scholars to access indigenous linguistic science alongside global paradigms, thereby promoting culturally balanced, high-quality education and research capacity.

SDG 9: Industry, Innovation and Infrastructure- emphasizes sustainable innovation and knowledge infrastructure. The thesis advances this goal by developing transparent, rule-based computational methods grounded in Sanskrit *śāstric* logic and *Pāṇinian* dependency structure. It models anaphora, lexical discourse markers, verbal expectancy, and episode connectivity using traditional interpretive tools, demonstrating their utility for natural language processing in low-resource classical and Indian languages. This supports technological innovation rooted in local intellectual traditions rather than exclusive reliance on data-driven Western models. The work thus contributes to building foundational linguistic technology infrastructure, enabling scalable, interpretable, and culturally relevant language-processing tools for Sanskrit and related Indian languages.

Together, these contributions reinforce educational equity, intellectual self-reliance, and innovation ecosystems by bridging classical linguistic wisdom with computational practice, ensuring sustainable and context-sensitive advancement in language technology and scholarship.

ACKNOWLEDGMENTS

In our coursework, Prof. K. Narayan Chandran, who taught *Research & Publication Ethics*, placed special emphasis on the **Acknowledgement** section. He often spoke of how it reveals the background, intent, journey, and transformation of a scholar. He believed that acknowledgements offer a glimpse into how a student becomes a scholar. Because of him, I devoted time and sincerity to this section, to express my deepest gratitude to all who helped me on this journey, signifying that all this learning and achievement ultimately belongs to those who shaped and supported me.

First and foremost, I extend my heartfelt gratitude to my guide, **Prof. Amba Kulkarni**. She has been much more than a supervisor, she has been my mentor, my support system, and a steady guiding light through all the ups and downs of this journey.

There were long phases when I struggled to make progress, months would pass without substantial work, and then suddenly, I would send her chunks of writing overnight, seeking quick feedback. Not once did she complain. She reviewed everything with care and patience, no matter how unpredictable my pace was. She bore with my chaos, always with kindness.

She supported me not only academically but also mentally and emotionally. She understood the pressure I was under as a young mother pursuing research, and she never made me feel inadequate for not fitting into a "perfect" academic mold. Her trust in me never wavered, even when I doubted myself.

Prof. Kulkarni has been the best mentor I could have asked for. Whatever I have achieved through this thesis, the ideas, the insights, the growth, all of it is deeply rooted in her generous mentorship. This work carries her presence in every line.

I thank **Prof. J.S.R. Prasad** for his insights into research methodology and his constant encouragement. I am grateful to **Prof. Aloka Parasher-Sen** for her kind words and support at crucial moments.

My sincere thanks to all the academic and non-academic staff of the **Department of Sanskrit Studies** for their timely help and administrative support.

To all my teachers who guided me along this path: **Mrs. Gayatri Joshi**, my school teacher who sparked my love for Sanskrit; **Dr. Manjusha Gokhale** and **Dr. Prasad Bhide**, my college professors who opened up dynamic perspectives, and my linguistics mentors from DCP, I owe you all a debt of gratitude.

To **Mahāmahopādhyāya Dr. Pushpa Dixit**, thank you for imparting your knowledge of Vyākaraṇa and inspiring me with your strength and scholarship. I am also deeply grateful to **Mahāmahopādhyāya Devadatta Patil Guruji** for opening the doors of the *Śrividyā Pāṭhaśālā* to me, and to all the teachers there who helped me explore the immense wisdom of Nyāya and Vyākaraṇa śāstras.

I thank **Dr. Pankaj Jaje** and **Dr. Swaraja Salaskar** for teaching me Nyāya and treating me like their own daughter during intense discussions. Dada's perspective on Nyāya texts, really shaped my critical thinking, nevertheless I wouldn't be able to decode the texts used in my research without his help and training. I also thank **Dr. Shridhar Lohakare** for teaching me *Mīmāṃsā-Nyāya-Prakāśa*, a work that greatly enriched my research.

To all my *sahadhyāyī seniors*, thank you for the debates, discussions, and the spirit of collective learning. Special thanks to **Dr. Monali Das** for guiding me with journal publications. I also express my gratitude to **Sanjivji**, **Madhuji**, **Arjun Sir**, and **Pawan Sir** for their continuing support in research.

To my dear friends and fellow scholars, **Malay**, **Sanalji**, **Anil**, **Abir**, **Shriram**, **Dhawalji**, thank you for standing by me. A special mention to **Shriram** for patiently helping me with all things technical, I'm truly a novice, and your help has been invaluable. Thank you **Malay** for always being there to solve my doubts, however silly they were.

To my soul sisters, **Amruta Malavade** and **Amruta Barbadikar**, thank you for your unconditional support, the shared meals, endless care, and emotional strength. Words will never be enough.

Thank you, **Sebanti**, **Prachi**, **Oindrila**, and **Ananya**, for making my hostel stay bearable and joyful. Sebanti, you truly were my "hostel everything."

To my friends beyond campus, your support meant everything. Naming each of you would take pages, but special thanks to **Vaibhav**, **Nayneesh**, **Samidha**, **Manas**, **Dhanashree**, **Rupesh**, **Sanket**, **Uma**, **Shakti**, **Jui**.

To my mother's friends, **Shruti** and **Mitali**, thank you for sharing in our journey and supporting me like family.

They say it takes a village to raise a child. In my case, it took two colonies, *Goregaonkar Lane, Mumbai* and *Hatture Vasti, Solapur*. Thank you to the **Barve** and **Dhanaashri families** for caring for my child like your own. To the **Joshi family**, thank you for offering your office space so I could study and write in peace. I am also grateful to **Mr. Vinayak Devrukhkar**, my father's assistant, for his constant help with typing, printing, and much more.

To our driver **Mahadev** and house help **Buchauna**, your support was small in appearance, but huge in impact.

To my extended family, the **Ponkshe**, **Kelkar**, **Joshi**, **Muttagi**, and **Kulkarni** families, thank you for your unwavering faith in me.

And now, my deepest gratitude goes to the four pillars of my life, my mother, my mother-in-law, my father, and my husband.

This thesis, this entire journey, is not mine alone. It is yours. Society expected me to adjust and focus on the child, and keep academics side-tracked for some time, because I had a child. But you stood by me and refused to let that happen.

Even when you were tired, ill, or had a thousand things of your own to manage, you took over my duties without hesitation. You fed my child, rocked him to sleep, played with him, and distracted

him, just so I could study, even for an hour longer.

In the early days, I would make my baby sleep on my lap and read beside him. Sometimes, I whispered books as lullabies. You all allowed me to continue, not just to survive, but to thrive. You didn't just support me, **you made this happen**.

To my **mother** and **mother-in-law**, thank you for raising my child, and for raising me again through this process. There's no way to truly thank a mother, which is why I dedicate this thesis to you both.

To my **father** and **father-in-law**, thank you for being the kindest grandfather, and for standing silently, but firmly, behind me.

To my **husband Kedar**, you never asked me to choose between being a mother and being a scholar. You loved me through it all, believed in my dreams, and gave me the freedom to pursue them. Even when I couldn't believe in myself, you did.

Every word in this thesis has your support behind it. Every page carries your sacrifices. This achievement belongs to us all.

To my **grand-mother** and **grandmother-in-law** have been beautiful examples of how to love and care selflessly, their lives taught me what it means to work with devotion for the well-being of loved ones.

To my **sister Savani** and **brother-in-law Kadar**, thank you for your unwavering support, steadfast belief in me, valuable perspective, and for always being there when it truly mattered, offering both encouragement and practical help throughout this journey.

To my **brothers-in-law, sisters-in-law**, thank you for your support and care for my child, allowing me to pursue PhD.

And finally, to my favorite little *Bacha-party*, Anagh and Dnyanada, thank you for mesmerizing me with your innocence, your laughter, and your light. You made me smile on days I thought I couldn't.

To my little one, **Raghav**, you may not understand this now, but you shaped my journey in ways no one else ever could. I studied with you in my mind, belly, and lap. You tolerated my struggles,

brought smiles in my darkest moments, and unknowingly gave me strength I didn't know I had.

Finally, my heartfelt gratitude to **Pandurang Shastri Athavale (Dadaji)** for shaping my perspective.

Ultimately, I thank **God** for everything You have given in this life.

"Idam na mama", This is not mine.

This thesis is dedicated to MOTHERHOOD.

To, my mother Mrs. Medha Vaze, mother-in-law Mrs. Sangeeta Muttagi, mother of thesis - my guide - Dr. Amba Kulkarni, and Dear Raghav who understood and adjusted to this struggling mother.

Table of Contents

List of Tables

List of Figures

Abbreviations

NLP	Natural Language Processing
PDTB	Penn Discourse Tree Bank
RST	Rhetorical Structure Theory
Ptr-Ne	Pointer Network
GuiTAR	Graph-based Universal Information and Textual Anaphora Resolution
CRFs	Conditional Random Fields
SARS	Sanskrit Anaphora Resolution System
ISO-DR	Slovenian Institute for Standardization - International Organization for Standardization - Semantic relations in discourse, core annotation schema (24617-8:2018)
IGT	Indian Grammatical Tradition
SCL	Sanskrit Computatinal Linguistics
PR	Pronoun Resolution
DP	Discourse Particles
DM	Discourse Markers
PPN	Personal Pronoun
DPN	Demonstrative Pronoun
RC	Relative-co-relatives
AS	aṣṭādhyāyī
BG	bhagvadgītā
SR	samkṣeparāmāyaṇam
LWG	Local Word Grouping
Tan vā	Tantra vārtika

Articles Published During the Course

Journal Papers:

1. **Indian Perspective on Pronominal Anaphora Resolution,** Ushati, the Ganganatha Jha Campus, Central Sanskrit University, Ganganath Jha Campus, Prayagraj, India. (forthcoming issue, 22nd)

Conference Presenations:

1. **Inter-Sentential Discourse Relations,** In Proceedings of the 7th International Sanskrit Computational Linguistics Symposium, pages 67–83, Auroville, Puducherry, India, Association for Computational Linguistics.

2. **Inter-Sentential Discourse Relations: An Overview- IGT Perspective,** 1st Samiksha - National Seminar on Sanskrit-IKS and ICT, National Institute of Advance Studies, Bangalore, Department of Science and Technology, Government of India, and Samskriti Foundation, Mysuru, India.

CHAPTER 1

Introduction

The discourse analysis study has gained significant attention in recent years, particularly in Natural Language Processing (NLP). Discourse analysis, the study of language in use beyond isolated sentences, is crucial in understanding communication and meaning. It examines how context, purpose, and cohesion influence the interpretation of language in both spoken and written forms. Over the years, discourse analysis has evolved as an interdisciplinary field, gaining insights from linguistics, sociology, psychology, philosophy, and computational sciences.

From an Indian perspective, discourse analysis is not a new endeavor. Ancient Indian traditions, particularly schools like Nyāya, Mīmāṃsā, Vyākaraṇa, and Sāhityaśāstra, have developed sophisticated frameworks to study meaning, context, and coherence. These traditions explore fundamental concepts such as Ekavākyatā (unity of meaning), Tātparya (intended meaning), Vivakṣā (speaker's intent), and Saṅgati (contextual coherence), etc., offering profound insights into the mechanisms of discourse.

This thesis focuses on exploring discourse analysis through the lens of Indian linguistic traditions and their potential applications in Sanskrit Computational Linguistics (SCL). While modern computational approaches often rely on large corpora and machine learning, Indian theories emphasize structured reasoning, regulatory rules, and context, providing complementary methodologies. This research investigates how principles from Indian traditions can address computational challenges such as pronoun resolution, Inter-sentential discourse relationship mapping, etc.

By bridging ancient Indian knowledge with modern computational techniques, this thesis seeks to contribute to both the theoretical understanding of discourse and its practical applications in

computational linguistics. The integration of these perspectives not only highlights the enduring relevance of Indian traditions but also offers innovative pathways to address the complexities of language in diverse contexts.

1.1 What is discourse?

Discourse refers to the use of language in communication, encompassing both written and spoken forms, often studied beyond the sentence level to explore coherence, context, and meaning. Derived from the Latin word *discursus*, meaning 'running off in different directions', it broadly examines how language operates within a given context or situation. Linguists, psychologists, sociologists, philosophers, and computational linguists approach discourse from diverse perspectives, each focusing on its relevance to their field.

Definitions of Discourse in Linguistics:

1. Cambridge Dictionary: The use of language to communicate in speech or writing.[1]

2. Oxford Dictionary: The use of language in speech and writing to produce meaning; language that is studied, usually to see how the different parts of a text are connected.[2]

3. Merriam-Webster Dictionary:[3]

   - Formal and orderly extended expression of thought.

   - Connected speech or writing.

   - A linguistic unit larger than a sentence.

From these definitions, discourse generally involves:

- The functional use of language,

[1] https://dictionary.cambridge.org/dictionary/english/discourse
[2] https://www.oxfordlearnersdictionaries.com/definition/american_english/discourse_1
[3] https://www.merriam-webster.com/dictionary/discourse

- The study of coherence and connection in language units,
- Analysis extending beyond sentence-level communication,
- Examination of written and spoken media of language,

Meaning comprehension and Discourse: Language operates in layers, beginning with sounds (studied in phonology), forming words (in morphology), and building sentences (syntax) and meaning (semantics). Sentence meaning, explored by traditions like Anvitābhidhānavāda[4] often extends beyond compositional word meanings. However, a comprehensive understanding requires both linguistic and non-linguistic factors.

- Linguistic Factors: Include intonation, stress, scripts, style, registers, co-text, references, and discourse extensions like text linguistics and pragmatics.
- Non-Linguistic Factors: Encompass contextual elements like knowledge, situation, emotions, relationships, medium, and cultural factors, all contributing to meaning.

Consider the sentence:

'The first Prime Minister of our nation was Nehru, and now it is Modi, 75 years post-independence.'

The interpretation of this sentence extends beyond the syntactic structure and semantic meaning of individual words or phrases. Understanding it requires extra-linguistic knowledge, such as historical context, political awareness, and numerical reasoning. All of these extralinguistic factors, which are not directly encoded in the lexical or structural components of the sentence, can potentially fall under the scope of discourse. This highlights the interdisciplinary nature of discourse analysis.

[4]Anvitābhidhānavāda is a theory from the Indian linguistic tradition which posits that the meaning of a sentence arises not as a simple sum of the meanings of individual words, but through their syntactic and semantic connections, where each word contributes its meaning in a relational and context-dependent manner to convey a unified sense. As opposed to Abhihitānvayavāda, which holds that individual word meanings are first understood independently and then combined to form sentence meaning.

Computational Perspective on Discourse: As computational linguists, we aim to enable machines to process human language and generate responses using discourse frameworks. We require computational assistance in addressing the following tasks.

- Automation: Handling repetitive linguistic tasks on its own, without human intervention.
- Comprehension and Production: Facilitating two-way language processing by computers.
- Tool Development: Creating modules for tasks such as anaphora resolution, co-reference resolution, discourse tree-banks, segmentation, concept mapping, dialogue systems, summarization, and chat-bots.

Thus, discourse, particularly in computational linguistics, integrates linguistic and contextual studies to address how meaning is constructed and communicated via computation.

Nature of Discourse Markers: Discourse markers are linguistic elements that connect segments of discourse by explicitly signaling the semantic or logical relations between them. They function as cohesive devices that guide the listener or reader through the structure of discourse, indicating relationships such as cause, contrast, elaboration, or consequence. Unlike discourse particles, which primarily serve pragmatic and interactional purposes - managing speaker attitude, emphasis, or conversational flow - discourse markers operate at the level of discourse organization, contributing directly to textual coherence and interpretive clarity. They make the underlying structure of discourse overt, helping the hearer or reader trace how propositions are connected and how ideas progress within a communicative act.

While both discourse markers and discourse particles support coherence, their functions differ in scope and focus: markers link propositions and reveal structural dependencies, whereas particles regulate tone and interaction. In Indian languages, discourse markers often take the form of connectives, conjunctive adverbs, or particles that indicate contrast, for example, through *parantu, kintu*, consequence through *atah̐, tasmāt*, or emphasis through *hi, eva*, serving as explicit indicators

of semantic relations between clauses or sentences. Given their role in mapping the relational architecture of discourse, the present research focuses on discourse markers and not on particles.

1.2 Difference between Western and Indian perspectives on Discourse

As we explore Indian theories of discourse, a key question arises: How do they differ from their Western counterparts? The answer lies in their distinct philosophical foundations and conceptual orientations. Indian and Western approaches diverge not only in method but also in their view of language and meaning, making an independent study of discourse from an Indian standpoint essential.

Western theories emphasize functional aspects such as analysis and comprehension, often without deeper philosophical grounding. In contrast, Indian theories - especially from the Sanskrit tradition - are embedded in extensive philosophical systems. For instance, while the Penn Discourse TreeBank (PDTB) classifies relations between arguments for annotation, the Mīmāṃsā concept of Ekavākyatā analyzes coherence within Vedic texts to preserve the integrity of the Vedas. Such principles have since informed broader linguistic analysis.

Western frameworks are typically goal-oriented and computationally practical. Indian theories, developed for classical texts, require reinterpretation for modern contexts. While Western thought evolved into distinct branches such as pragmatics and discourse studies, Indian traditions integrated these within larger philosophical domains - an approach offering insights particularly relevant to Sanskrit but surely expandable to Indian as well as foreign languages.

Indian frameworks also address linguistic phenomena often missed by Western models. For example, constructions specific to Sanskrit and other Indian languages, such as co-referring sen-

tences, are not fully captured by Rhetorical Structure Theory (RST) or PDTB. Indigenous theories, therefore, provide more nuanced tools for analyzing such structures.

1.3 Need of the Topic

A considerable body of research has examined discourse in Indian languages from both linguistic and computational perspectives.

(Prasad and Strube, 2000) applied Centering Theory to Hindi pronoun resolution, demonstrating that grammatical function (e.g., subjecthood) serves as a stronger indicator of discourse salience than word order. This finding underscores the significance of discourse-level features in morphologically rich languages. The *VASISTH* system proposed by (Sobha and Patnaik, 2000) represents one of the earliest rule-based approaches to anaphora resolution in Malayalam and Hindi, utilizing clause boundaries and morphological agreement features. Further, (Sikdar et al., 2016) introduced a generalized framework for anaphora resolution in resource-scarce Indian languages through multi-objective differential evolution, achieving competitive results for Hindi, Bengali, and Tamil on standard coreference evaluation metrics.

Discourse relation analysis for Indian languages has primarily adapted Western frameworks such as the Penn Discourse TreeBank (PDTB) and Rhetorical Structure Theory (RST) to accommodate linguistic diversity. PDTB-based initiatives include the Hindi Discourse Relation Bank by (Oza et al., 2009), discourse tagging for Hindi (Sweta Khandelwal, 2013), and argument identification (Jain et al., 2016). In Bengali, (Kar et al., 2013) developed a shallow PDTB-style corpus, while (Das and Stede, 2018) constructed a comprehensive RST treebank. For Tamil (Rachakonda and Sharma, 2011) and Malayalam (Kumari and Devi, 2016), early annotation efforts focused on identifying language-specific discourse markers. Telugu research has primarily contributed syntactic treebanks that serve as the foundation for subsequent discourse studies (Vempaty et al., 2010; Nallani et al., 2020).

A review of these studies indicates that most research on discourse in Indian languages remains grounded in Western theoretical frameworks, with adaptations made to address local linguistic features. While such work has advanced computational discourse analysis for Indian languages, it largely involves the application of existing Western methodologies rather than the development of indigenous frameworks rooted in Indian linguistic and philosophical traditions.

Further investigations have explored textual coherence through linguistic and computational approaches, including cohesion and topic modeling in Hindi and Gujarati (Singh and Sharma, 2017; Verma and Gupta, 2019; Joshi et al., 2021), discourse relation and event extraction in Bangla and Hindi (Das et al., 2020; Sahoo et al., 2020), and graph-based summarization and fact linking in Telugu and Bengali (Vakada et al., 2023; Kolluru et al., 2021; Wasi et al., 2024).

India's classical linguistic traditions - Vyākaraṇa (grammar), Nyāya (logic), Mīmāṃsā (coherence), and Sāhitya-śāstra (poetics) - offer profound theoretical insights into language structure, meaning, and rhetoric. Although originally formulated for the interpretation of ancient texts, these frameworks remain highly relevant for understanding discourse phenomena in modern contexts and can significantly contribute to computational discourse modeling.

Integrating such indigenous linguistic knowledge into computational frameworks presents an opportunity to design models better aligned with the structural and cultural characteristics of Indian languages. This approach not only enhances the linguistic accuracy of computational systems but also supports the preservation of the intellectual and cultural heritage inherent in these traditions.

Despite this potential, there has been limited effort to incorporate indigenous linguistic theories into computational discourse analysis for Indian languages. Addressing this gap is crucial for the advancement of linguistically and culturally informed research. The present work aims to bridge this gap by proposing a framework that integrates traditional Indian theories of language and discourse, thereby demonstrating their relevance in contemporary computational contexts and contributing to

the creation of linguistically and culturally grounded NLP systems.

1.4 Aim and Objective

The main aim of this research is to bridge the gap between traditional Indian linguistic theories and modern computational discourse analysis, with a focus on pronoun resolution and inter-sentential relationship mapping.

The specific objectives of the research are:

1. To look at different concepts from tradition and find out how they can contribute to dealing with the problems of discourse. The concepts include *Tātparya, Vivakṣā, Saṅgati, Tantrayukti, Ekavākyatā,* etc.

2. To explore the integration of Indian linguistic theories, particularly from the schools of *Nyāya, Mīmāṃsā, Vyākaraṇa,* and *Sāhitya-śāstra,* into computational models for discourse analysis.

3. To develop a rule-based computational approach for pronoun resolution that leverages principles such as *Ekavākyatā* (unity of meaning) to improve discourse coherence. And to create an empirical model for marking inter-sentential relations using traditional Indian concepts. (Which we have named IGT - Indian Grammatical Traditional Perspective that computationally analyses discourse.)

4. To create a wholesome representation for the sentence and beyond analysis following the dependency framework suggested by *Paṇini*.

5. To contribute to the development of traditionally informed and linguistically accurate Natural Language Processing (NLP) tools for Indian languages.

1.5 Scope of the study

As established in the preceding discussion, discourse analysis is a broad and interdisciplinary field of research. Given the constraints of time and resources, it is not feasible to encompass all aspects of discourse within the scope of this thesis. Therefore, this study is limited to specific issues that align with the research objectives.

According to (Das, 2016), discourse can be categorized into four structural types, among which this study focuses on coherence structure. The coherence structure primarily examines linguistic cues within a text that contribute to its overall meaning and connectivity. Additionally, the study operates at the level of context, focusing on relationships between consecutive sentences. By employing a two-sentence format, the research aims to identify and analyze inter-sentential relationships based on linguistic markers.

Two key challenges in discourse analysis have been identified for investigation, and solutions will be explored through insights derived from ancient and traditional Indian philosophico-linguistic texts.

The first challenge is pronoun resolution, with a specific emphasis on co-referring pronominal resolution. This involves identifying referents of pronouns within nearby sentences, where the antecedents appear in the form of other pronouns. Here, the research focuses on instances where a pronoun form derived from *yad*[5] is referenced by another pronoun form derived from *tad*[6] in adjacent sentences. We also discuss the possible referent of the personal pronouns *Asmad* and *Yuśmad*. We only limit our study to these categories, because other instances of demonstrative pronouns are highly contextual.

[5]A relative root word from the Sanskrit language

[6]A co-relative (as well as a demonstrative) root word from the Sanskrit language

The second challenge pertains to discourse relations, specifically the identification of lexically marked discourse relations at the inter-sentential level. The study aims to detect lexical markers within sentences that signal discourse relations and establish connections with neighboring sentences based on these cues. At this stage, the focus remains on explicit discourse relations, where such markers are overtly present in the text.

In addition to addressing the two primary research problems focused on automation and computational modeling, this study also undertakes a manual exploration of how the entire text can be connected through various discourse-level parameters. This manual analysis aims to complement the computational components by providing deeper qualitative insights into textual coherence and structure.

Currently, our discourse-analysis work focuses exclusively on the Sanskrit language. This allows us to develop and refine our methodological framework within a well-defined linguistic tradition. As our approach matures, we intend to extend it to additional languages.

1.6 Methodology

This thesis adopts an interdisciplinary approach, integrating traditional Indian linguistic theories with rule-based computational techniques to address discourse-level challenges. The theoretical framework focuses on Ekavākyatā, a principle from Indian linguistic traditions, emphasizing the unity of meaning across a sequence of sentences. The research will apply this framework to solve two primary problems in discourse analysis: pronoun resolution and inter-sentential relation marking.

The methodology consists of both theoretical and empirical components. The theoretical component involves exploring Ekavākyatā and related principles from Indian linguistic schools such as Nyāya, Mīmāṃsā, and Vyākaraṇa. These principles will guide the understanding of discourse

structure, coherence, and meaning at the discourse level, going beyond individual sentence analysis. The empirical component will develop a rule-based computational model for pronoun resolution and inter-sentential relation marking, using annotated Sanskrit corpora to test the proposed system's effectiveness.

The ultimate goal is to demonstrate the applicability of Indian linguistic theories in modern discourse processing tasks, improving computational models for discourse analysis, particularly for Indian languages.

1.7 Overview of thesis

This thesis explores the potential of Indian traditional theories in addressing discourse-level comprehension and processing challenges. Focusing on Ekavākyatā, a principle from Indian linguistic traditions, it examines its relevance to discourse processing. Specifically, the study investigates Ekavākyatā's application to pronoun resolution, the marking of inter-sentential relations, and coherence in the text.

The thesis is organized as follows:

Chapter 2 provides a detailed exploration of Indian traditional linguistic and philosophical theories from schools such as Nyāya, Mīmāṃsā, Vyākaraṇa, and Sāhitya. These schools have contributed extensively to understanding language and meaning at various levels. This chapter reviews the tools and frameworks these schools offer for discourse analysis, particularly for units of language that surpass the sentence level. The principle of Ekavākyatā is analyzed in depth, including its definition, the different types of Ekavākyatā, the concept of a macro-sentence (Mahavakya), variation from the perspective of each school of thought, and the application. Additional concepts, such as Tātparya (the intended meaning), Vivakṣa (deliberation or intention in speech), Saṃgati (coherence or connection), and Tantrayukti (techniques for text arrangements), Tātparya-grāhaka (determinants of intended meaning), Anubandha (framework of interpretive relevance), Ṡaktigraha-

upāya (Disambiguation techniques) etc. are also briefly discussed to establish their relevance to discourse comprehension and processing.

Chapter 3 addresses the first major problem tackled in this study: pronoun resolution. The resolution of anaphoric references marked by pronouns is essential for achieving coherence in discourse. This chapter applies the principle of Ekavākyatā to analyze pronouns and their referents, a process traditionally referred to as Sarvānama-vicāra. The analysis focuses on understanding the referential mechanisms for pronouns, the direction of reference, and determining the referents within a discourse. Using the framework of Panini's dependency grammar, this study develops an extension to the existing dependency framework to graphically represent and connect pronouns with their appropriate referents. The Saṃkṣheparāmāyaṇam serves as the primary text for a corpus study. We also developed a preliminary algorithm aimed at identifying referential relationships within referring–coreferring structures, and personal pronouns based on the Sanskrit scriptures.

Chapter 4 investigates the second problem, that is, establishing relations marked by Discourse Marker. These markers, which function as linguistic devices to connect sentences, are analyzed in the context of argument positioning, syntactic structures, and their representation. This chapter compares the proposed framework with widely used discourse annotation systems such as the Penn Discourse Treebank (PDTB) and Rhetorical Structure Theory (RST) and a Standardised framework (ISO-DR) to highlight its distinct advantages. The corpus of the Bhagavadgītā is employed as a case study to apply and validate the theoretical insights from this analysis. We have also devised a primary algorithm to automatically detect these markers and connect sentences with the appropriate relations, and to disambiguate the polysemous markers. The Evaluation of the implemented programme is done using the appropriate matrix.

Chapter 5 expands the scope to examine larger discourse patterns and tools for testing textual coherence at the level of an entire text. It draws parallels between the concept of Prabandhavākya from Sāhityaśastra (theory of literature) and text-linguistics. This chapter identifies and analyzes

tools such as discourse relations (including pronoun resolution and discourse markers), named and co-referring entities, semantic expectancies, and the concept of Kulaka, among others. These techniques are applied to the text of the Saṃkṣeparāmāyaṇam, resulting in the construction of a comprehensive and interconnected graph encompassing 100 shlokas.

Chapter 6 Finally, the Conclusion summarizes the study's key findings, along with the significance, addressing the challenges encountered during the research and providing directions for future work.

With the foundation laid in the introduction, we now proceed to the theoretical framework that forms the basis of this study. The next chapter delves into various tools, discussing discourse, and providing a comprehensive exploration of the concepts and frameworks central to this research.

CHAPTER 2

Tools of Discourse from Indian Tradition

Sanskrit, one of India's most ancient languages, originates in the Vedic tradition and has functioned as a repository of knowledge spanning a vast range of disciplines, from mathematics and medicine to astronomy and economics. Emerging from its rich oral tradition, Sanskrit has developed an extensive textual corpus, the production of which continues to this day. This textual tradition is not only vast but also deeply structured, forming a complex web of interdependence and hierarchical organization.

A distinctive aspect of Sanskrit scholarship is its extensive commentary tradition as mentioned by Das (2016), consisting of various layers of interpretative texts such as *Bhāṣya* (detailed commentaries), *Ṭīkā* (simple and brief meaning), and *Kārikā* (concise aphoristic summaries). These texts do not exist in isolation but rather engage in a dynamic intellectual discourse, building upon and elaborating on pre-existing material. In addition to the commentarial tradition, Sanskrit exhibits a unique practice of *Khaṇḍana-Maṇḍana*, a systematic approach to critique and counter-argumentation, which fosters rigorous academic debates. Another defining characteristic of Sanskrit discourse is its structured meta-language: a specialized system of technical vocabulary that ensures coherence across disciplines. While this meta-language maintains a high degree of uniformity across different fields of study, some discipline-specific terminologies also emerge to cater to specialized areas of inquiry.

Due to its inherently interdependent, layered, and interdisciplinary nature, the study of Sanskrit texts necessitates a thorough understanding of discourse structures. A scholar must not only extract relevant information from texts within the same discipline but also establish connections with related

works across different branches of knowledge. This makes the study of discourse in Sanskrit an essential and complex endeavor. For instance, in the case of commentaries, comprehending isolated sentences is insufficient; a holistic analysis is required to grasp the overarching argumentative and explanatory structures. Without such an approach, critical linkages between texts and ideas might be overlooked, making textual navigation significantly more challenging.

Recognizing this complexity, the Sanskrit tradition has developed theoretical frameworks that explicate the interwoven nature of textual discourse. Although these frameworks were not originally formulated for discourse analysis as an independent discipline, they inherently serve to make textual coherence explicit. These theories enable scholars to discern the intricate web of inter-textual references, dependencies, and logical structures embedded within Sanskrit texts. Their primary purpose is discipline-specific, often directed toward philosophical or doctrinal clarity, yet their underlying principles can be extended to broader discourse analysis. For instance, in the *Vedānta* tradition, the four *Mahāvākya*s (macro-sentence) from the *Upaniṣad*s are integrated through the principle of *Ekavākyatā*, which unifies them into a single conceptual framework explaining the relationship between Brahman (the ultimate reality) and ātman (the individual self). In this case, the primary function of *Ekavākyatā* is not discourse analysis per se but rather the attainment of philosophical realization and ultimately *mokṣa* (liberation). However, since its core mechanism involves the integration of multiple statements into a cohesive conceptual structure, it offers insights that can be projected onto the domain of discourse analysis. Thus, while its intended goal is metaphysical understanding, its functional framework provides a valuable model for studying textual coherence and inter-textuality.

Three Levels of Relational Analysis : We have identified three types of relationships that operate at three different levels within a text. Each level functions at a higher level than the preceding one.

1. Intra-sentential level – This level encompasses relations that exist within a single sentence.[1] Most of these relations occur between nouns and verbs (*kāraka* relations) or between two

[1]We follow the definition of a sentence as '*Eka tiṇ vākyam* ', meaning a unit containing only one finite verb.

nouns (*viśeṣaṇam, ṣaṣṭhī* relations, *upapada* relations, etc.). This level is highly explicit, as such relations cannot be marked without grammatical cues like suffixes or particles. (Kulkarni and Ramakrishnamacharyulu, 2016) have developed a detailed and extensive parser to identify and represent these relations using the *Pāṇinian* Dependency Framework, where they are referred to as grammatical relations. However, some relations are marked between two verbs yet appear within a sentence. Though this may seem contradictory, examples clarify the distinction.

Example 1:

Sanskrit: *Rāmaḥ dugdhaṃ pītvā śālāṃ gacchati.*

Gloss: Rama[m,nom,sg] milk[n,acc,sg] drinking[present.part] School[f,acc,sg] go[present,3p,sg]

English: Rama goes to school after drinking milk.

Here, a temporal relation exists between the two verbs, one finite and non-finite, which aligns with our sentence definition explained further in this chapter, making this an intra-sentential relation. Such relations are marked by non-finite suffixes like *ktvā, lyap, ktavatu, śatṛ, śānac, tumun,* etc.

Example 2:

Sanskrit: *śvaḥ āgamiṣyāmi iti Caitraḥ Maitram vadati.*

Gloss: Tomorrow[indcl] come[present,1p, sg] so[indcl] Chaitra[m, nom, sg] Maitra[m,acc, sg] say[present,3, sg]

English: I will come tomorrow, says Chaitra to Maitra.

In this case, the first sentence serves as a *karman* (object) for the verb *vadati* (says). Despite appearing across two sentences, the relation remains intra-sentential as it follows the *kāraka* theory, marking the *karma* relation.

2. Inter-sentential (across-sentence) level – At this level, relationships exist between two distinct sentences, typically occurring in consecutive sentences through referencing (such as anaphora, co-reference) or connecting (coordination, subordination, etc.). These relations may be explicit or implicit depending on the presence of linguistic cues. This level is the primary

focus of this thesis, where we aim to identify and represent such relations and integrate them into the existing parser based on the *Pāṇinian* Dependency Framework. These relations are referred to as Discourse Relations.

3. Across-text level – This level includes conceptual relations, which extend beyond sentence boundaries to connect broader textual elements such as utterances or groups of sentences. These relations bridge gaps between concepts and meanings represented in the text and are predominantly topical. Due to the absence of explicit linguistic markers, most of these relations are inferred rather than directly expressed. (Das, 2016) initiated research in this area using the concept of *Saṅgati* as a coherence relation. Examples of such relations include coherence and topical relations, which can be analyzed through concepts like *Saṅgati*, *Tantrayukti*, etc.

In the subsequent sections, we examine a set of theoretical constructs from the Sanskrit tradition that can aid in navigating the complexities of discourse analysis. These include *Tātparya* (intended meaning), *Vivakṣā* (speaker's intention), Saṅgati (contextual coherence), and *Tantrayukti* (textual strategies), etc. However, our primary focus will be on the concept of *Ekavākyatā* (homogeneity of sentence meaning), which serves as a foundational principle throughout this thesis, providing a unifying thread that connects various aspects of discourse analysis in the Sanskrit tradition.

2.1 *Ekavākyatā*

Originating from the *Mīmāṃsā* school, *Ekavākyatā* is a foundational concept within *vākyāśāstra*[2] aimed at elucidating the semantics of Vedic texts. This principle is extendable to non-Vedic, natural language sentences, emphasizing the unification of meaning from multiple sentence fragments (words or phrases).

[2]the science of sentences

The core premise of *Ekavākyatā* posits that when two or more sentences or fragmented constructs convey a singular meaning, they must be synthesized. This unified sentence unit is termed *Mahāvākya* (macro-level sentence).

2.1.1 Relation between *Ekavākyatā* and *Mahāvākya*

Ekavākyatā is the essential condition for the formation of a *Mahāvākya*, ensuring that multiple sentences collectively convey a unified meaning. It denotes '*Ekārtha Bodhakatvam*', the ability to express a singular, coherent idea. A *Mahāvākya* arises only when *Ekavākyatā* is established through *Ākāṅkṣā* (expectancy), *yogyatā* (semantic compatibility), and *sannidhi* (proximity). If these conditions are met, individual sentences (*Avāntaravākya*) merge into a *Mahāvākya*, forming a macro-sentence with unified intent. Conversely, in the absence of *Ekavākyatā*, sentences remain distinct and cannot constitute a *Mahāvākya*. As suggested by Subrahmaṇyaḥ (2012), just as a single sentence is comprehended when these three factors align, their coordination (*Samanvaya*) at the macro level results in *Ekavākyatā*, ensuring syntactic and semantic unity in a *Mahāvākya*.

2.1.2 Types of *Ekavākyatā*

Ekavākyatā can be classified into two categories:[3]

1. *Padaikavākytā*: This refers to the scenario where fragments from one sentence (e.g., groups of words) combine with another sentence, facilitating meaning conveyance. This is particularly useful in anaphora resolution. *Padaikavākyatā* is technically defined as: the unity of a sentence achieved by incorporating or introducing a group of words from a subordinate sentence into a main sentence - *pradhānavākyasya padakalpena guṇavākyena saha Ekavākyatā* (Subrahmaṇyaḥ, 2012). The concept of *Padaikavākyatā* has not been extensively discussed

[3]In *Vyākaraṇa*, the treatise *Laghuśabdenduśekhara* introduces a third type of *Ekavākyatā*, termed *Arthaikavākyatā*, wherein only the meaning of certain entities shifts to the concerned sentence rather than the words themselves. However, this concept is not elaborated upon in this thesis for two reasons: first, it is not required for the subsequent discussion, and second, it lacks universal acceptance.

in *mīmāṃsā*, but a case study (which is discussed in Section 2.1.4) within the domain of *vyākaraṇa* that substantiates and elucidates the process of *Padaikavākyatā*. The integration occurs at the lexical level (individual words).

2. *Vākyaikavākyatā*: This pertains to the integration of two distinct individual sentences to convey a unified meaning, aiding in the discourse of relational semantics. Mīmāṃsā discipline primarily focuses on *Vākyaikavākyatā*. When the primary and subordinate sentences as complete units merge to convey a unified meaning with the subsidiary relation (*śeṣaśeṣibhāva*), it is *Vākyaikavākyatā*. It is technically defined as:
'*aṅgāṅgivākyayoḥ śeṣaśeṣirupayoḥ paprasparamekavākyatā vākyaikavākyatā.* '[4]
The integration occurs at the sentential level (entire sentences).

2.1.3 Elementary Discourse Unit: Sentence

In the Sanskrit tradition, the definition of a sentence (*Vākya*) varies across disciplines (David, 2017). Given these varied perspectives, this study adopts a definition that is both precise and quantifiable for sentence analysis. Specifically, we adhere to the *Vyākaraṇa* school's definition, which states that a sentence is *ekatiṇ vākyam*, meaning it must contain exactly one finite verb form.[5] Some *Mīmāṃsakā*s propose that a sentence must have only one *uddeśya* (subject) and one *vidheya* (predicate), asserting that any multiplicity necessitates the division into distinct sentences.[6] However, this interpretation lacks both clarity and universal acceptance, making it unsuitable for this study.

For example:

Sanskrit: *rāmaḥ grāmaṃ gacchati.*

Gloss: Rama[m,n,sg] village[n,acc,sg] go[present,3p,sg]

[4](Subrahmaṇyaḥ, 2012, p. 02)

[5]A sentence is a unit containing one and only one finite verb form.

[6]This definition, along with the definition *ekatiṇ vākyam*, corresponds to the same sentences. The only difference is that the earlier definition is semantic, whereas the latter one is morpho-syntactic.

English: Ram goes to the village

Such sentences typically convey complete and homogeneous meaning. However, constructions such as below necessitate additional insights from neighboring sentences or broader contextual understanding. Although such constructions contain a single finite verb and are, in technical terms, classified as sentences, they often fail to convey a complete or unambiguous meaning due to their fragmented structure.

Sanskrit: yaḥ kāryaṃ karoti
Gloss : Who[m,nom,sg] work[n,acc,sg] do[present,3p,sg]
English: Who does the work...

In such cases, to solve the problem of comprehension, we reference an expanded definition from *mīmāṃsā śāstra*: A sentence expresses a single idea; upon splitting, the fragments shows expectancies[7] as mentioned by Jaimini (1916, p. 171). Here, a sentence comprises multiple sub-clauses or units of language, facilitating a more nuanced interpretation necessary for achieving cohesive meaning. This definition expects two clauses:

1. *arthaikakatva*: This principle denotes the homogeneity or unity of meaning, asserting that a sentence must convey complete and sufficient meaning.

2. *nirākāṅkṣatva*: This principle states that all expectancies arising from the sentence, whether from individual words or the sentence as a whole, must be resolved, rendering it complete in itself.

For Example:
Sanskrit: yadā rāmaḥ kāryālayaṃ gacchati. tadā sa kāryaṃ karoti.
Gloss: When[indcl] Ram[m,n,sg] Office[n,loc,sg] go[present,3p,sg]. Then[indcl] he[m,nom,sg]

[7] *arthaikaikatvāt ekaṃ vākyam. sākaṅkṣam cet vibhāge syāt.*

work[n,acc,sg] do[present,3p,sg]

English: When Rama goes to the office, he works.

In the example above, both conditions of *arthaikaikatva* and *nirākāṅkṣātva* are satisfied. Despite being distinct sentences (*avāntara-vākya* or part-sentences), together they convey a complete meaning and are contextually incomplete without each other. The conjunctions *yadā* and *tadā* create a mutual expectancy, while the pronoun *saḥ* necessitates *rāmah* as its referent. Therefore, such sentences exemplify the concept of *Ekavākyatā*, enabling their synthesis into a macro-sentence termed *Mahāvākya*, which embodies the properties mentioned above.

To approach this subject from an alternative perspective, consider the analysis of an individual sentence (*avāntara-vākya*) independently, without invoking the broader structure of a macro-sentence or *mahāvākya*. In such a context, the primary objective is to demarcate the syntactic boundaries of a sentence by grouping relevant constituents under the governance of a single finite verb, in accordance with the principles of *kāraka*-theory. This necessitates a more technical and syntactically grounded definition of a sentence. Consequently, the definition *eka-tiṇ-vākyam*- a sentence is that which contains one and only one finite verb- is employed. This definition provides a concrete structural unit to the utterance and serves as the Elementary Unit for establishing discourse-level relations (EDU).

However, when the aim shifts towards the comprehension of meaning, it becomes apparent that these individual sentences may or may not be semantically sufficient or independently complete. In such cases, the fusion of multiple syntactic units is required, leading to the construction of a *mahāvākya*. This shift necessitates a more semantic or meaning-oriented definition, for which the criterion *arthaikatvāt ekam vākyam*- a sentence is that which conveys a single unified meaning- is invoked. This is further supported by the clause *sakāṅkṣam cet vibhāge syāt*, which accounts for the interdependence of syntactic units when semantic expectancy is present.

Mahāvākya - an extension of concept of sentence : *Mahāvākya*, meaning a macro-sentence, extends the fundamental principles governing a conventional sentence to a higher structural level. In the same way that a sentence is formed by the meaningful arrangement of words, a *Mahāvākya* emerges from the syntactic and semantic integration of individual sentences. However, for multiple sentences to collectively function as a *Mahāvākya*, they must satisfy the core linguistic prerequisites of *Ākāṅkṣā* (expectancy), *yogyatā* (semantic compatibility), and *sannidhi* (proximity). *Ākāṅkṣā* ensures that one sentence creates an anticipation that another sentence fulfills, *yogyatā* ensures logical coherence between them, and *sannidhi* ensures that they appear close enough in discourse to be processed as a unified whole. Just as a sentence cannot be formed by randomly placed words, a *Mahāvākya* cannot arise from arbitrarily arranged sentences; rather, it requires an inherent structural and semantic unity.[8]

Consider the utterance:

Sanskrit: *ahaṃ āpaṇaṃ gacchāmi*

Gloss: I[nom,sg] market[n,acc,sg] go[present,3p,sg]

English: I am going to the store.

While this statement conveys an action, its purpose remains unspecified. When followed by

Sanskrit: *mama kācit śākāni kretum āvaśyakāni*.

Gloss: I[gen,sg] some[indef] vegetables[n,acc,pl] buy[purpose] need[present.part]

English: I need to buy some vegetables.

The two sentences together form a coherent unit. The first creates *Ākāṅkṣā* (expectancy) for a purpose, which the second fulfills while maintaining *yogyatā* (semantic compatibility). Their *sānnidhi* (proximity) ensures they are processed as a unified thought, exemplifying *Mahāvākya*tā, where distinct sentences integrate to convey a singular, complete meaning.

[8]Although this example implicitly draws on ākāṅkṣā, yogyatā, and sannidhi - the three factors guiding sentential unity - it is presented here solely for explanatory clarity, as the human mind processes these relations effortlessly.

2.1.4 Different school's perspective on *Ekavākyatā*

Mīmāṃsā

Application of *Ekavākyatā* in *Mīmāṃsā* : The principle of *Ekavākyatā* in *Mīmāṃsā* is applicable in the following cases:

1. *Bhāvanā* (Enablement) plays a crucial role in the interpretation of Vedic injunctions. *Bhāvanā* is of two types: lexical and semantic. Each type involves three components - *sādhya (goal), sādhana (instrument)*, and *itikartavyatā* (additional procedural knowledge). In both kinds of enablement, sentences that establish *sādhya* and *sādhana*, and *itikartavyatā*, together form a macro-sentence (*Mahāvākya*), which ultimately enables an individual to perform the prescribed ritual (*yāga*).

   For example, the sentence: '*darśapūrṇamāsābhyāṃ svargakāmo yajeta* (Bhaskar, 1934) establishes the *sādhya* (goal) as *svarga* (heaven) and the *sādhana* (instrument) as *darśapūrṇamāsa-yāga* (the mentioned sacrifice). However, this information alone is insufficient to execute the ritual. The procedure (*itikartavyatā*) must also be known, which is provided by the sentence: '*svāhākāraṃ yajati* '. stating that the ritual must be performed with *Svāhākā*ra (offering oblations while uttering *Svāhā*). These two sentences, when combined, form a *Mahāvākya* that conveys a complete and coherent directive necessary for ritual performance.

2. Vedic sentences can be categorized into five types: *Vidhi* (injunctions or imperatives), *Mantra* (hymns), *Niṣedha* (prohibitions), *Arthavāda* (praise or censure), *Nāmadheya* (nomenclature).

   Among these, *Vidhi* (imperatives) are considered self-authenticating (*svataḥ pramāṇa*) and semantically independent. The remaining categories derive their significance from *Vidhi*. For instance, *Mantra*s aid in the remembrance of *Vidhi, Niṣedha* serve to restrict or limit *Vidhi, Arthavāda*, which reinforces *Vidhi* through praise and supports *Niṣedha* through censure.

Thus, the formation of a *Mahāvākya* necessitates the integration of these sentence types to ensure a comprehensive and homogeneous meaning.

For example, in the case of the following imperative sentence (*Vidhi*):

'*vāyavyaṃ śvetam ālabheta bhūtikāmaḥ.* '(He who desires prosperity should sacrifice a white animal to *Vāyu.*) (*Taittirīya Saṃhitā* 2.1.1)

An *Arthavāda* statement supplements it:

vāyur vai kṣepiṣṭhā devatā. (Vāyu is the fastest deity.)

Together, these sentences convey a unified meaning:[9]

'*yataḥ śīghrakārisvabhāvatayā śīghraphalapradaḥ vāyuḥ asya śvetapaśoḥ devatā. tataḥ praśastam imaṃ vāyudevatākaṃ paśuṃ aiśvaryakāmaḥ saṃspṛśet.* '

(Since Vāyu, due to his swift nature, bestows quick results and is the presiding deity of the white animal, one desiring prosperity must consecrate this animal to the deity *Vāyu.*)

This exemplifies how *Ekavākyatā* integrates *Vidhi* and *Arthavāda* into a *Mahāvākya*, ensuring the completeness of the directive.

3. In *Mīmāṃsā*, the implementation of Vedic rituals (*viniyoga-Vidhi*) follows six sources of authentication, wherein *aṅgatva* (subsidiarity) is determined. Among these, *vākya* is the third criterion after *śruti* (direct scriptural statements) and *liṅga* (contextual markers).

The concept of *vākya*[10] must be understood in terms of *Mahāvākya*. The definition of vākya as samabhivyāhāraḥ (syntactic proximity) implies that individual sentences, when serving a unified purpose and found in close textual proximity, should be treated as a single *Mahāvākya*. For instance, consider the statement:

[9]The implicit meaning, acting as a causal link, integrates these two sentences into a *śeṣa-śeṣībhāva* relationship, wherein one functions as the principal (*śeṣī*) and the other as the subordinate (*śeṣa*).

[10]Though *vākya* can be classified into two types - *āvāntara* (individual) and mahā (macro) - the author of Śrī Mādhavācārya (1989) asserts that in this context, vākya should be interpreted as*Mahāvākya*.

‘*yasya parṇamayī juhurbhavati, na sa pāpaṃ ślokaṃ śṛṇoti.* ’(*Taittirīya Saṃhitā* 3-5-7) (One whose ladle is made of *palāśa* (flame of the forest tree) does not hear inauspicious words.)

Here, due to textual and semantic proximity, *palāśa* becomes a subsidiary (*śeṣa*) of the ladle (*śeṣī*) through the principle of *śeṣa-śeṣī-bhāva* (subordination relation). This integration exemplifies how *Mahāvākya* structures arise in Vedic discourse to establish ritual significance.

4. The determination of a text's intended meaning relies on seven key indicators: *upakrama* (beginning), *upasaṃhāra* (conclusion), *abhyāsa* (repetition),*apūrvatā* (novelty), *phalam* (result), *arthavāda* (praise or censure), and *upapatti* (reasoning).

   For example, in the chapter *Vaiśvānareṣṭi*, a set of sentences elaborate on the expected results or what one should wish for according to the number of *kapāla* (earthen pot fragments) offered in a sacrifice. The concerned sacrifice requires only one sentence specifying twelve *kapāla* pieces. However, the chapter contains several sentences describing the use cases of different numbers of *kapāla* pieces. In this case, the other sentences appear futile. Therefore, these sentences must be understood through the lens of *upakrama, upasaṃhāra*, and other such tools. The intended meaning of these sentences is to praise the prescribed number twelve, as stated in:

   ‘*yad dvādaśakapālo bhavati jagatyevāsmin paśūn dadhāti* ’

   (Twelve pieces of *kapāla* bless one with many cattle). These sentences together form an *arthavāda* (a group of praise sentences), establishing *Ekavākyatā* with the main imperative sentence:

   ‘*vaiśvānaraṃ dvādaśa-kapālaṃ nirvapet putre jāte.* ’

   (One blessed with a son must offer twelve *kapāla* pieces to the *Vaiśvānara* fire deity).

5. *Ekavākyatā* is also seen in the pair of *prakṛtiyāga* and *vikṛtiyāga*. *Prakṛtiyāga* refers to a prototype or primary sacrifice explicitly prescribed in the Vedic texts, serving as the standard ritual framework with all the necessary instructions and nuances for its performance.

In contrast, *vikṛtiyāga* denotes a derivative sacrifice, which is a modified version of the *prakṛtiyāga*, incorporating variations in elements such as offerings, procedural details, or deities while preserving the fundamental structure and intent of the original ritual. However, *vikṛtiyāga* lacks all independent instructions on execution and must derive essential details from the *prakṛtiyāga*. This process of borrowing is only possible through the assumption of *Ekavākyatā*, ensuring that the instructions of the *prakṛtiyāga* and the requirements of the *vikṛtiyāga* form a unified and coherent directive. Consequently, this alignment results in the formation of a *Mahāvākya*, integrating the two sets of sentences into a singular, operative instruction. For instance, *Darśapūrṇamāsa* is a *prakṛtiyāga*, a fundamental Vedic sacrifice performed on new moon and full moon days, while *Agrayāga* is a *vikṛtiyāga*, derived from *Darśapūrṇamāsa*, where the modification involves offering the first portion of the new grain instead of the standard offerings.

***Prāmāṇyam* (Authentication) :** According to *Mīmāṃsā*, if *Mahāvākya* exists, it should be considered ultimate, independent, and authentic, and not the individual sentences it comprises:
‘ *na ca mahāvākye sati, āvāntaravākyānām pramāṇam bhavati* ’(*śābarabhāṣya* 4.7.25).

However, when sentences stand alone without further expectancy, they must be considered independent and authentic (*pramāṇa*):
‘*śeṣaśeṣivākyato anyatarasya na svātantryamasti* ’((Subrahmaṇyaḥ, 2012, p.33)).[11]

***Viśeṣaṇa–Viśeṣya Sambandhaḥ* (Modified–Modifier Relation):** This syntactic-semantic relationship manifests in three primary forms:

1. Between co-referential words and adjectives, known as *sāmānādhikaraṇya* (shared referentiality).

2. Between verbs and their corresponding nouns, termed *kriyā-kārakavatī* (verb-case role association).

[11]Neither the matrix nor the subordinate sentences/clauses are independent alone.

3. Between subordinate and matrix clauses, forming a *śeṣa–śeṣī-bhāva* (dependency relation).

In the third case, despite the absence of a common case-ending (*vibhakti*), *Ekavākyatā* (sentence unity) is established as the constituents integrate to form a *Mahāvākya* (macro-sentence), provided that the essential syntactic and semantic prerequisites - *Ākāṅkṣā* (expectancy), *yogyatā* (semantic compatibility), and *sānnidhi* (proximity) - are satisfied.

Formational process of Māvākya : There are two distinct theoretical perspectives regarding the comprehension and formation of a group of sentences. *Abhihitānvayavāda* asserts that individual words (*pada*) in a sentence (*vākya*) fully convey their meaning independently, and the sentence is merely an arrangement (*anvita*) of these meaningful units. This view aligns with the principle of compositionality and is also referred to as *padavāda*, as advocated by *Kumārila*. When extended to the level of *Mahāvākya*, this perspective implies that individual sentences serve as meaningful inputs, and their combination results in a larger coherent unit. On the other hand, *Anvitābhidhānavāda* maintains that individual words and sentences do not inherently carry complete meaning or are insufficient to do so. Instead, meaning emerges primarily from their arrangement and interrelation. A *Mahāvākya*, under this perspective, is fundamentally the structured composition of its constituent sentences. This view, also known as *vākyavāda*, was propagated by *Prabhākara Miśra* . The principles of *Abhihitānvayavāda* and *Anvitābhidhānavāda* are projected onto the *Mahāvākya* level in the same manner as they apply to the formation of a sentence from individual words.

Thus, *Ekavākyatā* remains a foundational principle in *Mīmāṃsā* discourse analysis. (Subrahmaṇyaḥ, 2012) extensively examines various instances wherein *Ekavākyatā* is identified within Vedic sacrificial injunctions, as analyzed in *Mīmāṃsā* exegesis. This analysis underscores that *ekavākyatā* serves as the fundamental criterion for the constitution of a *mahāvākya*. In the absence of *ekavākyatā*, the given sentences must be interpreted as independent linguistic units, devoid of syntactic and semantic integration.

Vyākaraṇa

In *Mahāvākyavicāra*, Subrahmaṇyaḥ (2012) postulates that *Mīmāṃsā-śāstra* is formulated based on the principle of *Ekavākyatā*, originally conceptualized by the *Mīmāṃsā* school for Vedic language. From a grammatical standpoint, *Ekavākyatā* can be examined through two distinct perspectives: natural language (*vyāvahārika bhāṣā*) and meta-language (*pāribhāṣika bhāṣā*).

Pāṇini incorporated *Ekavākyatā* within different types of *sūtra*s (such as *saṃjñā, paribhāṣā*, etc.) while formulating the *Aṣṭādhyāyī*. A distinguishing feature of the *Aṣṭādhyāyī* is lāghava (brevity), achieved by avoiding unnecessary repetition through the mechanism of *Ekavākyatā*. A *sūtra*, once stated, is not reiterated but recalled whenever necessary. The *Aṣṭādhyāyī* prescribes six types of *sūtra*s:

1. *Saṃjñā* (Nomenclature)
2. *Paribhāṣā* (Meta-rule)
3. *Niyama* (Restrictive Rule)
4. *Vidhi* (Imperative or Core Instruction)
5. *Atideśa* (Transfer of Property)
6. *Adhikaraṇa* (Topical Property or Domain)

These *sūtra*s, functioning in an interdependent manner, ensure grammatical coherence. Consequently, *pada-ekavākyatā* or *vākya-ekavākyatā* is systematically achieved within grammar.

Kumārila Bhaṭṭa, in *Tantravārttika* , states:
sarvāṇyeva hi śāstrāṇi svapradeśāntaraissaha.
ekavākyatayā yuktamupadeśaḥ pratanuvate.. (Tan. Vā. 3.44.13)

Translation: All the sūtras, which operate both in their stated locations and in other pertinent contexts, collectively establish their own unity and thereby prescribe the requisite function.

This implies that all *sūtra*s, irrespective of their placement, collectively form a unified instruction (*upadeśa*) through *Ekavākyatā*. Commenting on this, *Kumārila* highlights the Significance and role of *Ekavākyatā* in the linguistic processes described by *Pāṅini*:
viśeṣaṇa tu vyākaraṇe. tatra hyekasmin pade prāyeṇa Aṣṭādhyāyī vyāpīyate.
Translation: Ekavākyatā is of particular significance in Vyākaraṇa, where even a single word encapsulates the essence of the entire Aṣṭādhyāyī

This highlights that in *Vyākaraṇa*, the entire *Aṣṭādhyāyī* is implicitly embedded within a single word, meaning that uttering a grammatical unit activates all associated rules necessary for interpretation.

In *Mahāvākyavicāra*, Subrahmaṇyaḥ (2012) extensively illustrates how different *sūtra*s interact to establish *Ekavākyatā*. Consider the following example from the *Aṣṭādhyāyī*:

Example 1 : In the work *Paribhāṣenduśekhara* by (Viśvanātha, 1990), while discussing *yathoddeśaḥ* and *kāryakāla pakṣa* in the context of *saṃjñā, paribhāṣā sūtra* (*sūtra*s 2 and 3), the example of '*tasmīniti nirdiṣṭa pūrvasya* 'is introduced. This *sūtra* serves as a technical aphorism, meaning that when a term is presented in the seventh case (locative), the operation is to be understood as affecting the state of what immediately precedes the term denoted.

This *sūtra* is instrumental in the application of *Vidhi sūtra*s, such as '*iko yaṇaci*.' Translation: the semivowels (*yaṅ - y, v, r, l*) serve as substitutes for the corresponding four vowels (*ik - i, u, ṛ, ḷ*) when followed by a vowel (*ac*). The synergy of the above-mentioned two *sūtra*s conveys the intended meaning. Both terms *aci* and *tasmīn* are in the locative case, which triggers the concept of *Ekavākyatā* between the two *sūtras*. Here, *tasmīn* acts as the referring expression, while *aci* denotes the intended referent. The terms *pūrvasya* (corresponding to *ikaḥ*) and *nirdiṣṭa* (corresponding to *aci*) together create a combined meaning: because of the locative case, the vowels *i, u, ṛ, ḷ* are replaced by their respective semivowels (*y, v, r, l*). This concept of *Ekavākyatā* illustrates how

nirdiṣṭa and *pūrvasya* combine with the main *sūtra - iko yaṇaci*.

Through this example, we can extend the concept and functioning of *Padaikavākyatā* to other scenarios. Applying this analytical framework to a common example:

S1: *rāmaḥ grāmaṃ gacchati.*
Gloss : Ram[m,nom,sg] village n,acc,sg] go[present,3p,sg]
English: Rama goes to the village

S2: *saḥ mṛgaṃ paśyati.*
Gloss : He[m,nom,sg] animal[n,acc,sg] see[present,3p,sg]
English: He sees an animal.

In this instance, the word *saḥ* from S2 serves as a referring pronoun. It relies on context to derive its meaning, seeking a suitable antecedent. Given that *rāma* reflects the same category in terms of person, number, gender, and case, it becomes the most eligible candidate. Thus, *rāma* from S1 engages in *Ekavākyatā* with S2, culminating in the construction of a *Mahāvākya*:

Sanskrit: *saḥ rāmaḥ mṛgaṃ paśyati.*
Gloss: He[m,nom,sg] Rama[m,nom,sg] Animal[n,acc,sg] See[present,3p,sg]
English: He that is Rama sees an animal.
This newly formed sentence is fully comprehensible in its own right.

Example 2 : The *sūtra - na ajjhalau* (*Aṣṭādhyāyī* 1.1.10) states that vowels (ac) and consonants (hal) are not to be considered savarṇa (similar kind of sounds). This *sūtra* appears to contradict the *saṃjñā* rule '*tulyāsya-prayatnaṃ savarṇam* '(*Aṣṭādhyāyī* 1.1.9), which defines *savarṇa* as sounds that share similar place and manner of articulation. Without reference to the propagation *sūtra* (1.1.9), the refutation *sūtra* (1.1.10) lacks coherence. Thus, both *sūtra*s naturally exhibit *Ekavākyatā*.

However, even this pair of *sūtra*s does not suffice in isolation. The propagation rule '*aṇuditsavarṇasya ca apratyayaḥ* '(*Aṣṭādhyāyī* 1.1.69) specifies when the *savarṇa* classification should be applied. Additionally, understanding '*na ajjhalau* 'requires other auxiliary *sūtra*s, such as the *Māheśvara-sūtra*s (*a, i, u, ṇ. ṛ, ḷ, k.* etc.) and phonetic rules like '*upadeśe'janunāsika it* '(*Aṣṭādhyāyī* 1.3.2) and halantyam (*Aṣṭādhyāyī* 1.3.3), which define nasal vowels and terminal consonants. Another essential *sūtra*, ādira-antyena saheta (*Aṣṭādhyāyī* 1.1.71), determines the phonological grouping of sounds (*pratyāhāra*). All are essential for comprehension.

These interdependent *sūtra*s collectively form a *Mahāvākya*, ensuring that a comprehensive meaning is conveyed without redundancy. Instead of restating information at multiple places, these *sūtra*s are recalled when necessary. Each *sūtra* acts as an *avāntara-vākya* (individual sentence), which collectively establishes *vākya-ekavākyatā* - a fundamental phenomenon, particularly in *Vidhi-Pratiṣedha* (propagation-refutation) *sūtra* pairs.

Similar to *Mīmāṃsā*, *Vyākaraṇa* considers *Vidhi-sūtra*s (imperatives) as *pradhāna-sūtra*s (primary), while auxiliary rules (*paribhāṣā, saṃjñā, pratiṣedha, apavāda,* etc.) function as subordinate (*gauṇa-sūtra*s). Through *pada-ekavākyatā* and *vākya-ekavākyatā, gauṇa-sūtra*s integrate with *Vidhi-sūtra*s to form a *Mahāvākya*, ensuring the completeness of a grammatical process (*prakriyā*).

Applying this principle to *vyāvahārika bhāṣā* (natural language), words within a sentence exhibit expectancy (*Ākāṅkṣā*), thereby forming a coherent meaning unit. Likewise, in a *Mahāvākya*, individual sentences that share mutual expectancy contribute to a unified interpretation. The meaning of a *Mahāvākya* is holistic (*akhaṇḍa*) and follows a process of synthesis, whereas its components can be mentally separated through analysis (*apoddhāra*). However, comprehension occurs through synthesis rather than decomposition.

Consider the following example:

Sentence 1:

Sanskrit: *Gauḥ duhyatām.*

Gloss: cow[f,nom,sg] milk[imp,3p,sg]

English: Milk the cow.

Sentence 2:

Sanskrit: *ācāryaḥ payasā odanaṃ bhuṅktvā māṃ adhyāpayiṣyati.*

Gloss: teacher[m, nom,sg] milk[imp,3p,sg] rice[n,acc,sg] eat[present.part] I[acc,sg] Teach [fut,3p, sg]

English: The teacher, having eaten rice with milk, will teach me.

These two sentences exhibit expectancy. The act of milking the cow (S1) serves the implicit purpose of providing milk for the teacher's meal (S2), which is a precondition for his teaching. This implicit expectancy establishes their *Mahāvākyatā*.

Two perspectives exist regarding the structure of *Mahāvākya*:

1. *Pratyeka-parisamāpti*: The *Mahāvākya* aligns with each individual part sequentially.

2. *Samudāya-parisamāpti*: The *Mahāvākya* is a unified whole formed by the collective contribution of its parts.

Similarly, two perspectives exist for comprehension:

1. *Khale-kapota-nyāya*: An analogical rule explaining how pigeons of all kinds gather where grains are available. Similarly, the entire Mahāvākya - all its parts together - is grasped instantaneously as a unified, consolidated whole.

2. *Rāja-puruṣa-nyāya*: An analogical rule illustrating how the king's men enter after the king, establishing a sequence and hierarchy. Similarly, meaning in language is derived sequentially in hierarchical order, ultimately forming the complete interpretation of the text.

This detailed discussion demonstrates how Ekavākyatā functions within Vyākaraṇa, with Pada-ekavākyatā and Vākya-ekavākyatā contributing both to the comprehension of meaning and to the

technical construction of grammatical processes.

Nyāya

The *Nyāya* school of thought primarily focuses on the means of knowledge (*pramāṇa*), whereas the *Vaiśeṣika* school emphasizes true understanding (*pramā*). Among the various means of knowledge, inference (*anumāna*) holds a central position in *Nyāya*. Inference is fundamentally based on the knowledge of concomitance (*vyāpti*) between two events, concepts, or entities. Naturally, inference is a cognitive process when one engages in reasoning personally (*svārthānumāna*). However, when inference is communicated to others, requiring them to derive the same conclusion, it is articulated through a set of five structured statements (*pañcāvayavī vākyāni*).[12] This structured inference is known as *parārthānumāna* (inference for others) as described by Annambhatta (1917).

According to the *Nyāya-Vaiśeṣika* traditions, these five statements must be comprehended as a coherent whole to enable valid inference. The integration of these five components relies on the principles of *Ekavākyatā* and *Mahāvākya*.[13]

The Structure of *Parārthānumāna* : To understand the concept of *parārthānumāna*, consider the following classical example:

Explanation of the Five Components :

1. *Pratijñā* (Proposition): The first statement asserts the existence of the predicate (*sādhya*, i.e., fire) in the subject (*pakṣa*, i.e., hill). The notion that the *sādhya* (fire) is to be proven in the *pakṣa* (hill) is termed *pakṣadharmatā (sādhyasya pakṣavṛttitvam)*.

[12]The pañcāvayavī vākyāni in the Indian logical tradition is often condensed into three statements in Western logic:
(i) Wherever there is smoke, there is fire.
(ii) This hill has smoke.
(iii) Therefore, this hill has fire.

[13]Despite some terminological differences, the *Nyāya-Vaiśeṣika* schools share common conceptual foundations; hence, their perspectives are discussed together.

	Nomenclature	Sentence
1	*pratijñā* (the proposition to be proven)	*parvato vahnimān.* (The hill has fire.)
2	*hetu* (the reason or indicator)	*dhūmāt.* (Because of the smoke.)
3	*udāharaṇa* (the example)	*yaḥ yaḥ dhūmavān.* *saḥ saḥ vahnimān.* *yathā mahānasaḥ.* (Wherever there is smoke, there is fire, as observed in the kitchen.)
4	*upanaya* (application or leading statement)	*tathā ca ayam,* *-parvato vahnivyāpya dhūmavān.* (Similarly, this hill has smoke, which is always accompanied by fire.)
5	*nigamana* – conclusion	*tasmāt tathā. - parvato vahnimān.* (Therefore, the hill has fire.)

Table 2.1: The Structure of *Parārthānumāna*

2. *Hetu* (Reason or Indicator): The second statement presents the logical basis for the inference - here, the presence of smoke (*dhūmaḥ*) serves as an indicator (*hetuḥ*) of fire (*vahniḥ*).

3. *Udāharaṇa* (Example Statement): The third component illustrates the invariable concomitance (*vyāptiḥ*) between smoke and fire. *Vyāpti* can also be expressed negatively: Wherever there is no fire, there is no smoke, as seen in a lake (*yatra dhūmaḥ na asti, tatra vahniḥ api na asti, yathā hṛdaḥ*), where *hṛdaḥ* (lake) serves as the *vipakṣa* (counterexample). In this relation, smoke (which occurs in fewer locations than fire) is *vyāpya* (pervaded), while fire (which occurs in more locations than smoke) is *vyāpaka* (pervader). This statement is also known as *vyāpti-smaraṇam* (remembrance of concomitance).

4. *Upanaya* (Application of the Rule): The fourth statement applies the general rule established in the previous step to the specific case. This step, also called *parāmarśa* (recollection), synthesizes *pakṣadharmatā* and *vyāptiḥ* within a single assertion, making it the crucial step leading to inference.

5. *Nigamana* (Conclusion): The final statement, which reiterates the *pratijñā*, signifies the completion of the inferential process. The *nigamana* itself is the *anumāna-pramāṇa* (valid

inferential cognition).

Ekavākyatā in Parārthānumāna : The term *avayava* in *pañcāvayavī vākyāni* suggests that these five statements function as integral components of a unified inference. Their synthesis forms a homogeneous unit of meaning, which is treated as a *Mahāvākya* (macro-sentence). Unlike individual sentences that stand independently, these inferential components are mutually dependent, ensuring that inference operates as a logically coherent and comprehensive process.

Beyond these structural principles, the *Nyāya-Vaiśeṣika* tradition does not provide further theoretical elaboration on *Ekavākyatā*. However, the outlined mechanism demonstrates how the synthesis of logically connected statements forms a valid inferential argument, reinforcing the necessity of *Ekavākyatā* in inferential reasoning.

Sāhitya

Sāhitya-śāstra, also referred to as *Kāvyā-śāstra* or *Alaṅkāra-śāstra*, encompasses the allied disciplines that analyze literary style and figures of speech. It systematically studies and examines the composition of poetry and literature. Any literary work, particularly poetry, must possess an inherent cohesion, ensuring connectivity through sentences or thematic concepts such as *Rasa* and *Alaṅkāra*. These individual components, when synthesized, create a homogeneous text. Thus, the concept of *Ekavākyatā*, which ultimately forms a *Mahāvākya*, can also be studied from a literary perspective.

In this section, we examine various literary phenomena through which *Ekavākyatā* is established based on different underlying principles.

1. **Group of Sentences or Verses (*Kulaka*):** A *kulaka* consists of a group of verses that together form a single sentence, typically facilitated by *anvaya* (syntactical re-arrangement). For instance, in *Raghuvaṃśa*, verses 5 to 9 of *Sarga* 1 collectively form a single sentence:
   '*So'haṃ - raghūṇām anvayaṃ vakṣye.* '

Here, a sequence of *viśeṣaṇa*s (descriptors) of *Raghuvaṃśa* (such as *ājanmaśuddha*, etc.), along with the poet's limitations (*tanuvāgvibhavaḥ*), are interwoven between them. This entire segment conveys a unified meaning, primarily structured with a single finite verb. Hence, it possesses *Ekavākyatā*, forming a cohesive *Mahāvākya*.

2. ***Artha-alaṅkāra*(Figures of Speech Based on Meaning)**: Figures of speech that rely on sense (*artha*) generally contribute to the formation of a *Mahāvākya*. They introduce a stylized or embellished mode of expression where the intent (*Vivakṣā*) and contextual meaning (*tātparya*) determine the presence of *Ekavākyatā* within interconnected sentences. A few notable examples include:

   (A) *Arthāntaranyāsa* (Supplementary Proposition): This figure of speech presents a proposition followed by its justification or illustration in subsequent text. In such instances, the primary proposition serves as the main sentence, while the justification is subordinate - mirroring the *Vidhi-arthavāda* structure in *Mīmāṃsā*.

   Example:

   S1: '*Yaḥ svabhāvo hi yasyāsti sa nityaṃ duratikramaḥ.* '

   S2: '*śvā yadi kriyate rājā tatkiṃ nāśnātyupānaham.* '

   Translation:

   S1: The inherent nature of a being remains unchangeable.

   S2: If a dog is made a king, will it cease eating footwear?

   Here, S1 is the main proposition, while S2 illustrates and reinforces it. Since both sentences convey a unified meaning, they form a *Mahāvākya* by possessing *Ekavākyatā*.

   (B) *Vyāja-stuti* (Concealed Praise through Apparent Criticism): This figure of speech presents praise indirectly by superficially expressing criticism (*nindā*). Even though the apparent meaning conveys censure, it remains subordinate, as the true intent is to offer praise (*stuti*).

   Example:

   S1: ' *Dhigananyopamāmetāṃ tāvakīṃ rūpam padam.* '

S2: '*Trailokye'pyanurūpo yadvarastava na labhyate.*'[14]

Translation:

S1: Fie upon this unmatched beauty of yours

S2: Even in the three worlds, a suitor worthy of you is not to be found.

In this verse, the word *dhik* appears to criticize the subject's beauty by asserting that it has no comparison. However, the intended meaning is to highlight its uniqueness and unparalleled nature, which is explicitly stated in the subsequent sentence. Since both sentences ultimately serve the same function- glorifying the subject- they form a *Mahāvākya* with *Ekavākyatā*. The key determinant in such instances is *Vivakṣā*, which governs how different sentences coalesce into a singular thematic unit.

3. ***Rasa* or core-sentiment:** The *Rasa* Theory is one of the most fundamental frameworks for understanding emotions evoked in poetry and literature. Scholars such as Ānandavardhana (1990), Bharata (1961), and Mammaṭa (1928) assert that a literary work maintains a single primary sentiment (*aṅgīrasa*). For example, *Vīra* (Heroic sentiment) is the dominant *Rasa* in *Mahābhārata*, and *Karuṇa* (Pathos) is the dominant *Rasa*in *Rāmāyaṇa*.

   While other subordinate sentiments appear throughout the text, they serve only as temporary emotional shifts and do not permeate the entire narrative. Additionally, descriptive passages, dialogues, and subsidiary stories function primarily to enhance and express the core sentiment. These secondary elements are considered subordinate since they serve the overarching emotional essence.

   The relationship between the *aṅgīrasa* and its supporting elements aligns with the *aṅga-aṅgi-bhāva* (principal-subordinate relation) as discussed in the *Mīmāṃsā* tradition. Consequently, the entire text attains *Ekavākyatā* with the primary sentiment binding its structure into a coherent *Mahāvākya*.

4. **Implicit Theme:** Every literary composition possesses an implicit core theme that binds and connects the text, supported by structural devices such as *upakrama* (introduction) and

[14]Kumārasambhava - 3.20

upasaṃhāra (conclusion) - paralleling similar constructs in *Mīmāṃsā*. Since this underlying theme permeates the entire composition, it is regarded as the central concept that governs all sentences.

Ānandavardhana (1990) in *Dwānyāloka* and Viśvanātha (1957) in *Sāhityadarpaṇa* assert that the core theme, in conjunction with all individual sentences or verses, constitutes a *Mahāvākya*. Since this *Mahāvākya* spans the entire text, it is referred to as *prabandhavākya*. For example,

The central theme of *Mahābhārata*:

'*Dharmarājādivat vartitavyam na Duryodhanādivat* '

(One should act like *Dharmarāja (Yudhiṣṭhira)*, not like *Duryodhana.*)

The central theme of Rāmāyaṇa is:

'*Rāmādivat vartitavyam na Rāvaṇādivat* '

(One should act like *Rāma*, not like *Rāvaṇa*).

This implicit concept functions as a *Vidhi* (imperative) that unifies all other sentences within the text. Similar to *Mīmāṃsā*, where all components of the Vedas *(nāmadheya, mantra, arthavāda, Niṣedha)* contribute to *Ekavākyatā* with the *Vidhi-vākya*, literary texts also achieve *Ekavākyatā* through thematic cohesion.

Thus, the principle of *Ekavākyatā* in *sāhitya* is realized through various mechanisms, such as *anvaya* (syntactical cohesion), *Vivakṣā* (intent), and *aṅga-aṅgi-bhāva* (principal-subordinate relationship). The notion that an entire text can be considered a single sentence due to its unified theme and interdependent components is a unique and significant perspective in literary analysis.

2.1.5 Application of *Ekavākyatā*

Padaikavākyatā

Padaikavākyatā requires the relocation of words within a sentence or across sentences to ensure the completion of meaning. This principle is naturally applicable in contexts where such a shift is anticipated or where meaning completion depends on it. One such domain is anaphora resolution. Anaphora[15] refers to instances where the interpretation of an expression relies on its antecedent. Anaphora can manifest in various forms, such as pronominal anaphora, verb phrase anaphora, one-anaphora, adjectival anaphora, and adverbial anaphora. Among these, pronominal anaphora is the focus of this study, wherein a pronoun derives its referential meaning through the process of *parāmarśa* (referencing). A detailed discussion on this topic is provided in Chapter 3.

Vākyāikavākyatā

Vākyāikavākyatā pertains to the unification of two distinct and independent sentences, such that they collectively convey a singular, cohesive meaning. This connection can either be explicitly marked through linguistic cues or inferred implicitly from the combined meaning of both sentences. Regardless of how the linkage is established, the two sentences function together as a unit, conveying a unified concept. Further elaboration on this subject is presented in Chapter 4.

Prabandhavākyatā

Prabandhavākyatā[16] refers to textual coherence, wherein an entire text is treated as a single unit. This implies that every sentence or *śloka* within the text must be interlinked, forming a structured web of meaning. Such coherence ultimately contributes to the realization and enjoyment of a unified *Rasa* (aesthetic essence) or core sentiment of the text. A comprehensive discussion on this principle is provided in Chapter 5.

[15] `http://dictionary.cambridge.org/dictionary/english/anaphora`

[16] *Prabandhavākyatā* is an extended form of *Vākyāikavākyatā* rather than a distinct category of *Ekavākyatā*

2.2 *Tātparya* and *Vivakṣā*

***Vivakṣā* or Speaker's Intent** *Vivakṣā* refers to the intent of the speaker in communication. A speaker, when conveying opinions, experiences, and ideas, deliberately selects a specific arrangement of words to express a particular meaning. The fundamental assumption underlying this selection is that the chosen linguistic expression is both sufficient and appropriate to convey the intended meaning to the listener. Raghavan (1963) in his treatise *śṛṅgāraprakāśa*, defines *Vivakṣā* as '

śabdeṣu arthābhidhāna-abhiprāyaḥ Vivakṣā,'

Translation: The intention to denote or express meaning by a particular word.

further elaborating that:

'*vaktuḥ vivakṣitapūrvikā vṛttiḥ* ',

Translation: The speaker's mindset before constructing an utterance.

which implies that *Vivakṣā* pertains to the intended meaning that the speaker seeks to infuse into the words used. It represents the cognitive process that precedes linguistic expression, encapsulating the speaker's strategic planning in the articulation of an utterance. In essence, *Vivakṣā* governs how an utterance is structured and how the speaker ensures its effectiveness in communication. The conceptual counterpart of *Vivakṣā* corresponds to *śuśrūṣā*, which denotes the listener's intent in interpreting the utterance.

The expression of *Vivakṣā* can be observed through multiple linguistic mechanisms:

1. Suffixes – The intent of the speaker significantly influences the selection of grammatical suffixes, particularly in the case of *kāraka* roles. For example, when describing a father-son relationship, '*ayam asya* '(this belongs to him) employs the *ṣaṣṭhī* (genitive) suffix to indicate possession, whereas '*ayam asmāt* '(this originates from him) uses the *pañcamī* (ablative) suffix to signify source. The choice between these suffixes is dictated by the speaker's intended emphasis.

2. Syntactic Features – The arrangement of words, structural dependencies, ellipses, and sequencing within a sentence all contribute to the explicit realization of *Vivakṣā*. This is particularly evident in cases where implicit relationships exist between sentences. For instance, in the sequence

   Sanskrit: '*Rāmaḥ prātaḥ kāle uttiṣṭhati. Devaṃ namati. Dugdhaṃ pītvā śālāṃ gacchati.* '

   Gloss: Rama[m,nom,sg] Morning[m,loc,sg] Wake-up[present,3p,sg]. God[m,acc,sg] pray[present,3p,sg]. milk[n,acc,ag] drink[present-part] go[present,3p,sg]

   English: Rama wkaes up in the morning. Praises God. And having drunk milk goes to school. The ordering of these sentences suggests a temporal progression, revealing an underlying intention of sequence.

3. Context (*Prakaraṇa*) – The contextual setting plays a crucial role in determining meaning. In Indian linguistic traditions, *prakaraṇa* is a fundamental aspect of interpretation, as emphasized in *Mīmāṃsā*. For instance, in the utterance '*Saindhavaṃ ānaya* ', the meaning of *Saindhava* is context-dependent: in the context of a meal, it refers to "salt,", whereas in a battlefield scenario, it denotes "horse."The speaker's *Vivakṣā* is thus clarified by the surrounding context.

4. Conventions (*Rūḍhī*) – Established idiomatic expressions, metaphors, and conventional phrases also serve as markers of *Vivakṣā*. For example, in the phrase '*Punar-āgamanāya ca* (literally meaning "and for returning"), the implicit intention behind the statement is *gacchatu* (go "). The intended meaning is conveyed not by the literal semantics of the words but by their conventionalized usage in discourse. This principle applies broadly to commonly understood phrases and idioms.

***Tātparya* or Expressed meaning** *Tātparya* signifies the intended meaning or purpose underlying an already articulated sentence. Some scholars categorize *Tātparya* as the fourth type of meaning, following *Abhihitārtha* (primary denotative meaning), *Lakṣyārtha* (indirect or connotative meaning), and *Vyāṅgyārtha* (suggestive meaning). *Sāhityikas* (literary theorists) often integrate *Tātparya* within *Lakṣyārtha* or *Vyāṅgyārtha*, whereas *Mīmāṃsaka*s view it as a distinct component,

particularly in the interpretation of *arthavāda* statements, which convey persuasive or motivational intent beyond the conventional denotative and connotative meanings. Such statements are employed in *Vidhi* (imperative) contexts to encourage specific actions.

According to (Raghavan, 1963), the interpretation of *Tātparya* varies across different domains. In literary and poetic contexts, where aesthetic embellishment is paramount, such intended meaning is classified as *Vyāṅgyārtha*. Conversely, in natural language and practical communication, where the speaker's specific intent governs interpretation, the same concept is designated as *Tātparya*. The responsibility of discerning *Tātparya* rests primarily with the listener, who must decode the intended meaning based on contextual and linguistic cues. Understanding the speaker's communicative intent is essential for accurate interpretation, ensuring that the meaning perceived aligns with the speaker's original intent.

Comparing *Vivakṣā* and *Tātparya* : *Vivakṣā* and *Tātparya* are two sides of the same coin, depending on whether they are viewed from the perspective of generation or comprehension. The key distinction between them lies in their focus: *Vivakṣā* is speaker-centric, whereas *Tātparya* is meaning-centric. *Vivakṣā* is a cognitive process that occurs before the utterance or usage, representing the speaker's intent in formulating an expression. In contrast, *Tātparya* is linguistic and exists even after the utterance, as it pertains to the meaning that needs to be interpreted by the listener. The responsibility of decoding *Tātparya* falls on the listener, who must derive the intended meaning based on the given linguistic and contextual cues.

Importance in Discourse Analysis : *Vivakṣā* and *Tātparya* play a crucial role in discourse, as effective communication depends on both the speaker's ability to plan and express their intent accurately and the receiver's ability to comprehend and recognize the intended meaning. Without this alignment, the process of communication remains incomplete, leading to possible misinterpretations or loss of meaning.

Application in Discourse Analysis : *Vivakṣā* and *Tātparya* have various applications, ranging from language generation and comprehension to discourse analysis. In discourse analysis, when moving beyond the word and sentence level, understanding the speaker's intent and how it is realized in an utterance becomes crucial. An application of this principle is examined in Chapter 3, in the context of identifying the referents of the pronouns *asmad* and *yuṣmad*.

2.3 *Adhyāhāra* and *Anuṣaṅga*

***Adhyāhāra* – Implicit Inclusion :** *Adhyāhāra* refers to the process of supplying unexpressed words or phrases that are either omitted in prior discourse or not explicitly mentioned later in the text. This process is inherently implicit, relying purely on the imaginative reconstruction of what must be included to complete the meaning and context of a sentence.

The application of *Adhyāhāra* is most evident in implicit discourse relations, where a logical connector or meaning must be inferred to establish coherence between two sentences.
For example:
Sanskrit : '*na te doṣo ayam. (kintu) svāmino doṣaḥ. '*
English Translation: This is not your fault. (But) It is the master's fault.
In this case, the connective *kintu* (but) is implicit, and its presence needs to be imagined to convey the intended contrast or antithesis. The process of *Adhyāhāra* is used here to infer the missing logical marker. Another common scenario where *Adhyāhāra* is applied is when a sentence lacks an explicit verb, and one must be mentally supplied to complete its meaning.
For example: When someone shouts: "*jalam* "(Water!). The implied meaning is: "*jalam (ānaya)*" (Bring water!). Here, the verb *ānaya* (bring) is unexpressed but naturally understood through *Adhyāhāra*.

***Anuṣaṅga*–Explicit Repetition** *Anuṣaṅga* refers to the repetition or reconsideration of words or phrases that have appeared earlier in the text. Unlike *Adhyāhāra*, this process is mechanical and

explicit, involving the deliberate reuse of previously mentioned linguistic elements.
The two primary purposes of *Anuṣaṅga* are:

- Textual economy (*lāghava*): To shorten the text and avoid redundancy.
- Aesthetic refinement (*sauṃdarya*): To maintain the stylistic elegance of the text.

One of the most prominent applications of *Anuṣaṅga* is in *Anuvṛtti* as seen in texts like *Pāṇini*'s *Aṣṭādhyāyī*, where words or grammatical rules from preceding aphorisms (*sūtra*s) are carried forward and reused. Another key area where *Anuṣaṅga* finds crucial application in anaphora or co-reference resolution, where previously mentioned entities are systematically recalled within the discourse. This process of retrieval operates through the mechanism of reference, known in the Indian grammatical tradition as *parāmarśa*.

Both *Anuvṛtti* (semantic or syntactic carry-over) and *parāmarśa* (referential recall) are regarded as subtypes of *Anuṣaṅga*, collectively contributing to the continuity and coherence of textual interpretation.

In summary, while *Adhyāhāra* involves imaginative reconstruction of missing words or meanings, *Anuṣaṅga* is a deliberate repetition of previously stated elements, serving textual and linguistic efficiency.

2.4 Saṅgati and Tantrayukti

This section examines two foundational frameworks from Indian discourse theory that offer complementary insights into the structure and coherence of texts: *Saṅgati* and *Tantrayukti*. *Saṅgati* primarily functions as an analytical instrument, illuminating the internal relational dynamics of a text. It helps the reader trace the logical, thematic, and rhetorical connections between units of discourse, making explicit how individual elements cohere within the whole. In contrast, *Tantrayukti* operates

as a prescriptive and constructive framework, oriented toward synthesis. It provides strategies and principles for assembling ideas into a coherent textual unit, guiding authors in the systematic organization and presentation of complex material. Whereas *Saṅgati* emphasizes interpretation and the uncovering of latent structural patterns, *Tantrayukti* emphasizes composition and the deliberate orchestration of content. Together, these frameworks offer a comprehensive paradigm: *Saṅgati* equips scholars to read and understand the intricate interplay of textual elements, while *Tantrayukti* equips writers to generate well-structured, cohesive discourses.

2.4.1 *Saṅgati*

In the Indian discourse tradition, *Saṅgati*s denote structured relational mechanisms that ensure coherence and logical consistency within a text. These relations play a pivotal role in linking different sections of a discourse, thereby facilitating the systematic progression of ideas and reinforcing the integrity of arguments. The tradition recognizes distinct sets of *Saṅgati*s across different philosophical schools, each contributing to the interpretative framework of texts.

***Mīmāṃsā Saṅgati*s:** [17]

1. *ākṣepa* (Objection) – Introduces a counterpoint to a previously stated argument, necessitating further clarification or refutation.

2. *Dṛṣṭānta* (Example) – Provides illustrative instances to substantiate claims, aiding in the conceptualization of abstract ideas.

3. *Pratyudāharaṇa* (Counter Example) – Challenges an existing assertion by presenting counter-evidence, thereby testing the robustness of the argument.

4. *Prasaṅga* (Corollary) – Serves as an incidental extension or consequence of the main argument, revealing underlying implications.

[17](Śāstrī, 1916)

5. *Upodghāta* (Pre-requisite) – Introduces essential background information, setting the foundation for comprehending the core argument.

6. *Apavāda* (Exception) – Identifies exceptions to a general rule, refining the scope of an argument and addressing potential contradictions.

***Nyāya Saṅgatis*:** [18]

1. *Prasaṅga* (Corollary) – Similar to the *Mīmāṃsā* definition, it logically follows from an argument and strengthens its implications.

2. *Upodghāta* (Pre-requisite) – Provides necessary context or foundational details for subsequent discussions.

3. *Hetūtā* (Causal Dependence) – Establishes a causal relationship between sections, indicating how one influences or leads to another.

4. *Avasara* (Opportunity for Further Inquiry) – Suggests that a particular argument or point provides scope for additional exploration or questioning.

5. *Nirvāhakaikya* (Common End) – Highlights that adjacent sections share a common conclusion, reinforcing textual unity.

6. *Kāryaikya* (Joint Causal Factors) – Demonstrates that multiple sections contribute to a shared outcome, emphasizing interdependence.

***Vyākaraṇa Saṅgatis*:** [19]

1. Praśna (Question) – Marks the initiation of inquiry or a thematic opening.

2. Uttara (Answer) – Provides a direct response to the raised question or problem.

3. Ākṣepa (Objection) – Poses a challenge or counterpoint to a stated view or answer.

[18](Śrī Mādhavācārya, 1989)

[19]As mentioned in (Das, 2016) derived from the commentaries of *Mahābhāṣya*

4. Samādhāna (Resolution) – Offers resolution to an objection, restoring argumentative continuity.

5. Bādhaka (Rejection) – Marks the refutation or invalidation of prior argument or resolution.

6. Sādhaka (Reaffirmation) – Reasserts a previously rejected view with additional justification.

7. Udāharaṇa (Example) – Provides a textual or linguistic example to support a theoretical point.

8. Dūṣaṇa (Criticism) – Indicates analytical criticism of a stated view or rule.

9. Vyākhyā (Explanation) – Elaborates on a term, view, or response, contributing to interpretive clarity.

10. Vārttika (Supplementary Rule) – Supplies additional rule or reflection that complements or challenges preceding statements.

These *Saṅgati*s ensure logical cohesion within a text, guiding its structural and argumentative integrity. Their systematic application facilitates rigorous textual analysis, allowing for a nuanced understanding of argument construction. Moreover, they serve as foundational principles in computational discourse analysis, aiding in the automation of textual parsing.

Recent studies highlight the significance of *Saṅgatis* in textual interpretation and computational linguistics. Das (2016) explored their automation in the *Mahābhāṣya*, while Subalalitha and Parthasarathi (2012) integrated them with Rhetorical Structure Theory (RST) to refine discourse parsing methodologies.

2.4.2 *Tantrayukti*

The word *Tantrayukti* is composed of two components: *Tantra*, referring to a science or specific discipline, and *Yukti*, meaning tools or techniques. *Tantrayukti* denotes scientific tools that help

eliminate imperfections in a treatise. These techniques unify and enhance coherence in a text. Despite being specific to a given work, they can be adapted for other texts with minimal modifications. *Tantrayuktis* serve as guiding principles for structuring a treatise.

As accumulated in Lele (2006), various scholars have discussed *Tantrayuktis*. *Kauṭilya* applied them in the *Arthaśāstra*, while *Caraka, Suśruta*, and *Vāgbhaṭa* incorporated them into their *Ayurvedic* treatises. Although there are variations in number and content among different traditions, the underlying descriptions remain largely consistent.

***Tantrayuktis* and Their Examples :**

1. *Adhikaraṇam*: *adhikaraṇaṃ nāma prastāvaḥ*. It refers to different chapters based on context. *Adhikaraṇa* can further be divided into *prakaraṇa*, etc. For example, *adhikaraṇaka pṛthivyāḥ lābha pālana-upayayoḥ śāstram* (the science of acquiring and maintaining land) - a very definition of *Arthaśāstra* is a *prastāva* (introduction) of the entire treatise.

2. *Vidhānaṃ*: *śāstrasya prakaraṇānupūrvīḥ*. It refers to the structured flow of content (*viśiṣṭa padaracanā*) followed by the author. It is a systematic presentation of the subject matter. Kauṭilya lists *adhikaraṇa* and *prakaraṇa* sequentially at the very beginning of the *Arthaśāstra*.

3. *Yoga*: *vākyayojanā*. It refers to the relationship (*sambandha*) between two statements. For example, *caturvarṇāśramo lokaḥ* (the world consists of four varṇas) and *rājā daṇḍena pālyata* (the king rules with law) - these two partial statements are combined to convey the complete meaning.

4. *Padārthaḥ*: *padāvadhikaḥ arthaḥ*. It refers to the specific meaning imposed on a word by a particular *śāstra* (discipline). For example, *mūlahara* is used in a specialized sense, meaning "one who takes away the principal amount." Additionally, technical terms such as *vṛddhi guṇa saṃjñā* (terms from Sanskrit grammar) serve as examples.

5. *Hetvarthaḥ*: *hetuḥ* (or) *artha-sādhakaḥ*. It refers to the justification of reason (*kāraṇa*). For example, *arthamūlau hi dharmakāmau* (wealth is the foundation of both righteousness and desire).

6. *Uddeśaḥ*: *samāsavākyam*. A concise statement. It also refers to the mere mention of concepts (*arthānāṃ śabdamātraṃ kīrtanam*). For example, in *Tarkasaṅgraha*, *dravya-guṇa-karma-sāmānya-viśeṣa-samavāya-abhāva*, referring to the seven categories (*sapta padārthāḥ*) in brief.

7. *Nirdeśaḥ*: *vyāsavākyaṃ*. A detailed elaboration of a statement. It refers to the specification of previously mentioned concepts. For example, in *Tarkasaṅkraha*, an in-depth discussion of the nature and types of *sapta padārtha* follows.

8. *Upadeśaḥ*: *evaṃ vartitavyam iti*. It refers to the instruction based on the wisdom of an expert (*āptavacanaśya kīrtanam*). For example, from the *Taittirīya Upaniṣad*, "Respect your teacher as a deity" (*ācārya devo bhava*) and "Do not criticize food" (*annaṃ na nindyāt*).

9. *Apadeśaḥ*: *evaṃ asau āha iti*. It denotes the opinions of others. For example, according to Kauṭilya, a prince must learn four disciplines, whereas Manu prescribes three, Bṛhaspati two, and Uśanas only one. Such differing views are also found in *Kauṭilya's* work, often cited as quotations.

10. *Atideśaḥ*: *uktena atideśaḥ sādhanam*. It refers to the extension of already established rules to a new context. For example, the rules of debt repayment (*ṛṇadāna*) stated in the chapter on gifts (*dattasyāpradāna*) can also be applied to state loans provided to merchants.

11. *Pradeśaḥ*: *vaktavyena sādhanam*. It involves referencing something that will be elaborated upon later in the text. For example, in the discussion of emergency provisions (*āpattyadhyāya*), details on fulfilling deficiencies in resources (*hināśaktipūraṇam*) are mentioned.

12. *Upamānam*: *dṛṣṭena anuṣṭhita sādhanam*. It explains an unknown concept through a known

example. For instance, in Vedanta, the analogy of a dancer (*nartakī*) is used to describe the behavior of *prakṛti* (nature).

13. *Arthāpattiḥ*: *yatra anuktaṃ arthaṃ āpadyate*. This refers to inferred meaning. *Mīmāṃsakas* consider it a separate form of valid knowledge (*pramāṇa*). A classic example is: "Devadatta, who is overweight, does not eat during the day," implying that he eats at night.

14. *Saṃśayaḥ*: *ubhayato hetumānaḥ arthaḥ*. It denotes doubt arising from two opposing possibilities. For example, "Is it a pillar or a person?" (*sthāṇur vā puruṣo vā*).

15. *Prasaṅgaḥ*: *prakaraṇāntareṇa samānaḥ arthaḥ*. It highlights the commonality between the two topics. For example, the duties of a household spy (*gṛhapatika vyanjana*) are stated to be similar to those of a spy stationed outside (*udyāsthit*).

16. *Viparyayaḥ*: *pratilomena sādhanam*. A concept is explained by its opposite. For instance, knowledge (*vidyā*) is explained through ignorance (*avidyā*) in Vedanta.

17. *Vākyaśeṣaḥ*: *yena vākyam samāpyate*. It refers to a fragment that completes a sentence. For example, the fragment *paśukāmaḥ* completes the entire context of the sentence *somena yajeta* (One who desires cattle should perform the Soma sacrifice).

18. *Anumatam*: *paravākya apratiṣedhaḥ*. It refers to the acceptance (*aṅkīkāra*) of an opposing viewpoint (*parapakṣa*). For example, in *Nirukta*, Yāska accepts the etymology given by Aupamanyava: "Since they are solely compilations (*nighantava*), they are called *nighaṇṭavāḥ* based on compilation" (*nighantava eva santo nigamanāt nighaṇṭavāḥ ucyante*).

19. *Vyākhyānam*: *atiśayavarṇanā*. It refers to the detailed explanation (*vistaraṇākhyānam*) of a briefly stated meaning. For example, Kauṭilya's elaborate discussion on *tantradaurbalya* (the weakening of administration) in the context of *vyasana* (negligence of wealth), where he explains how a king loses everything due to gambling (*dyūta*).

20. *Nirvacanam*: *guṇataḥ śabdanirṇayaḥ*. It refers to determining the meaning of a term based on its characteristics. Additionally, it denotes the functional definition of a technical term (*saṃjñayā uktasya tadarthena niyojanam*). For example, in *Tarkasaṅgraha*, *ayutasiddhaḥ* is further explained as "that which remains dependent even when one of the two elements is destroyed" (*yayormadhye ekam avinaśyoparāśritam eva tiṣṭhati iti ayutasiddho*). This refers to the actual place of application of a concept.

21. *Nidarśanam*: *dṛṣṭānta-dṛṣṭāntayuktaḥ*. It refers to an example or illustration (*sādhyasya ekadeśadṛṣṭāntaḥ*). For instance, "A king who goes to war without preparation suffers like a person who fights an elephant while standing on the ground" (*hastinā pāśayuddham iva*).

22. *Apavargaḥ*: *abhiplutavyapavarṣaṇam*. It refers to an exception carved out of a general category (*vyāpta*). For example, the *Manusmṛti* lays down strict rules regarding purity but also provides exceptions in cases of necessity. While consuming leftover food is generally prohibited, it is permitted during a famine (*Apad-dharma*). This is an example of *Apargaḥ*, where an exception is made to a universal rule.

23. *Svasañjñā*: *paraiḥ asaṃjñita śabdaḥ*. It refers to technical terms used exclusively by a particular system. For example, terms such as *ṭi, ghu, bha* introduced by Pāṇini in his grammatical meta-language, which were not used by others.

24. *Pūrvapakṣaḥ*: *pratiṣeddhavyaṃ vākyaṃ*. A statement or argument that is to be refuted by the text or author. For example, *anarthakāḥ vaḥ mantrāḥ* (Your mantras are meaningless) - a claim made by Kautsa, which is later refuted by Yāska in *Nirukta*.

25. *Uttarapakṣaḥ*: *nirṇayavākyaṃ*. Also known as *siddhānta* (established doctrine). For example, Yāska's counterargument in response to Kautsa, asserting that mantras do have meaning and significance.

26. *Ekāntam*: *sarvatra āyattam*. A universal truth or principle that applies in all cases. For example, *tasmād utthānam ātmanah kurvīta* (Therefore, one must make an effort by oneself),

as stated by Kauṭilya.

27. *Anāgtāvekṣyaṇam*: *paścād evam vihitam*. Also known as *anāgatāpekṣā*, it refers to a mention of something that will be explained later in the text. For example, sentences like *agre vakṣyate* (This will be stated ahead).

28. *Atikrāntāvekṣaṇam*: *purastād evaṃ vihitam*. Also known as *atītāpekṣā*, it refers to a reference to something that has been previously stated. For example, phrases like *pūrvamuktam* (This has been said before).

29. *Niyoga*: *evaṃ na anyathā iti*. A rule regarding an obligatory action (*kriyā* or *vyāpāra*). For example, the process of *kāṇḍana kriyā* (pounding) is prescribed for husking rice (*vrīhi-tuṣa-vimocana*), rather than alternative methods like *nakhaiḥ tuṣavimocana* (removal of husk by nails).

30. *Vikalpaḥ*: *anena anena vā*. "This or this" - indicating an optional choice between two alternatives. For example, in Pāṇini's grammar, the root *veṭ* allows for the optional inclusion of *iṭ āgama*.

31. *Samuccayaḥ*: *anena anena ca*. "This and this" - where two or more things are applicable simultaneously. For example, in the inheritance section (*dāyāda*), the term *svayamjāta* can be applied both to *pitā* (father) and *pitṛbandhu* (paternal relative).

32. *ūhyam*: *anuktakaraṇam*. This refers to an implied or hypothesized action. For example, in a legal context, if it is stated that neither the giver (*dātā*) nor the receiver (*pratigrahītā*) should suffer, it implies that a particular kind of transaction must be carried out. The specific mode of transaction remains *ūhya* (implied or unsaid), though general guidelines are mentioned.

Application of *Tantrayukti* in Discourse Analysis :

Tantrayukti plays a crucial role in structuring and analyzing discourse by ensuring logical coherence, clarity, and systematic argumentation. It aids in organizing textual content through

principles such as *Adhikaraṇa* (subject classification) and *Vidhāna* (structured arrangement), which establish a logical flow in a treatise. The technique of *Pūrvapakṣa* (presentation of opposing views) and *Uttarapakṣa* (resolution of arguments) enhances critical engagement by allowing a comprehensive examination of different perspectives before concluding. Furthermore, *Nidarśana* (illustration) and *Upamāna* (analogy) make abstract concepts more accessible by providing concrete examples, facilitating better comprehension. The use of *Vikalpa* (alternative options) and *Samuccaya* (simultaneous applicability) allows for nuanced interpretations, accommodating multiple viewpoints within a discourse. Additionally, *Tantrayukti* ensures precision and depth by defining terms explicitly through *Padārtha* (restricted meaning) and reinforcing reasoning with *Hetvartha* (justification of cause). By employing these techniques, discourse analysis becomes more structured, coherent, and methodologically rigorous, making *Tantrayukti* a valuable tool for both ancient and modern textual studies. Recently Panchukrishnan (2024) has worked on this principle and its application on finding the coherence in a partial portion of *Rāmāyaṇa*.

2.5 Tātparya-grāhaka and Anubandha

Both *Tātparya-grāhaka* and *Anubandha-catuṣṭaya* serve as critical frameworks in classical Indian philosophy for discerning the intentional structure and purposive orientation of a text. Together, they contribute significantly to the tātparya-nirṇaya (determination of intended meaning or purpose) in discourse analysis. These tools not only guide the exegetical reading of śāstric texts but also offer methodologically rigorous parameters for understanding the internal coherence, relevance, and communicative design of philosophical discourse.

Tātparya-grāhaka-ṣaḍ-liṅga: The *Ṣaḍ-liṅga*, or sixfold set of determinants of intended meaning, is a foundational apparatus in Sanskrit hermeneutics, particularly within the Mīmāṃsā and Vedānta traditions. Comprising *upakrama* (beginning), *upasahāra* (conclusion), *abhyāsa* (repetition), *apūrvatā* (novelty), *phala* (result), and *arthavāda* (commendatory or illustrative statement), this framework is articulated in classical texts such as the *Mīmāṃsā Sūtras* of Jaimini and expounded

by commentators like Śabara and Śaṅkara. These six indicators are employed to identify the central intention (*tātparya*) of scriptural passages, especially when interpretative ambiguity arises. In the context of discourse analysis, the *ṣaḍ-liṅga* facilitates the recognition of thematic continuity, argumentative emphasis, and rhetorical structuring within a text.

Anubandha-catuṣṭaya: The *Anubandha-catuṣṭaya*, or fourfold framework of interpretive relevance, is another pivotal model in classical Indian discourse theory, particularly in the Vedānta and Nyāya traditions. It consists of four interrelated components: *adhikārī* (qualified recipient), *viṣaya* (subject matter), *prayojana* (intended purpose or benefit), and *sambandha* (relation between subject and purpose). Found in seminal works like the *Tarka-saṅgraha* and *Vedānta-sāra*, this model functions as a pre-discursive analytical lens that determines a text's scope, applicability, and coherence. Within discourse analysis, the *Anubandha-catuṣṭaya* helps in mapping the communicative intent of the text by clarifying for whom, about what, why, and in what relational context a discourse is composed.

Taken together, *Tātparya-grāhaka* and *Anubandha-catuṣṭaya* provide a complementary hermeneutic apparatus for purpose-driven discourse analysis. While the former focuses on intra-textual markers that signal the author's intent, the latter frames the interpretive environment within which that intent operates. This synthesis enables a nuanced understanding of both the internal structure and the external communicative function of classical texts. However, this area remains largely unexplored in existing scholarship.

2.6 Śaktigrahopāya

The *Śaktigrahopāya*, or the methodological apparatus for grasping the specific contextual meaning (*śakti*) of a word, forms a crucial component in classical Sanskrit semantic theory. It outlines how meaning is not fixed by lexical definition alone but is dynamically determined by a range of contextual and cognitive factors. Traditional sources enumerate multiple such factors- *saṃyoga*

(co-occurrence), *viprayoga* (disassociation), *sāhacarya* (habitual association), *virodhitā* (semantic opposition), *prakaraṇa* (discourse context), *liṅga* (indicatory signs), *saṃnidhi* (proximity), *sāmarthya* (semantic compatibility), and *aucitya* (appropriateness), among others- as the *viśeṣa-smṛti-hetavaḥ* (cognitive prompts for distinguishing intended sense). These mechanisms are treated in texts such as the *Tarka-saṅgraha*, *Nyāya* works, and grammatical commentaries, which emphasize that meaning is a function of both linguistic structure and contextual inference.

Within the domain of discourse analysis, *Śaktigrahopāya* serves as a classical framework for word-sense disambiguation. It accounts for how interpreters negotiate polysemy, metaphor, and contextual ambiguity by relying on syntactic, semantic, and pragmatic cues. Particularly in philosophical and literary texts where precision of interpretation is critical, these factors function as decision-making criteria for selecting the intended referent or meaning of a term. Thus, *Śaktigrahopāya* supports disambiguation at the lexical level, complementing broader discourse-level coherence strategies. It also aligns with modern concerns in lexical semantics and pragmatics by foregrounding how linguistic meaning is context-sensitive, inferential, and contingent on interpretive competence. There is a noticeable absence of systematic research in this field.

2.7 Summary

This chapter examines the structured nature of Sanskrit discourse, emphasizing interpretative frameworks that shape scholarly engagement. It highlights the commentary tradition's role in enriching textual analysis and coherence.

A central principle, *Ekavākyatā*, ensures thematic unity by linking meaning across sentences and texts. *Tantrayukti* offers generative and organizational principles that aid in structuring coherent argumentation and synthesizing textual content. *Saṅgati* provides analytical tools for identifying logical and thematic connections across sections, thereby facilitating coherence in interpretation. The study also explores *Tātparya* and *Vivakṣā*, which clarify the speaker's intent and intended meaning.

Tātparya-nirṇaya offers systematic criteria for determining textual purpose through markers such as *upakrama* and *phala*, while the *Anubandha-catuṣṭaya* framework situates discourse about its audience, subject, purpose, and internal linkage. *Śaktigrahopāya* provides context-sensitive strategies for word-sense disambiguation based on co-occurrence, proximity, and semantic appropriateness. It also discusses *Adhyāhāra*, the supplementation of implied meaning, and *Anuṣaṅga*, the progression of thought within discourse. Together, these analytical tools demonstrate the sophistication of Indian discourse traditions and their relevance for contemporary textual analysis.

Ekavākyatā provides the central framework for addressing the problems examined in the following chapters, while other related concepts highlight how principles from classical Indian treatises can inform and facilitate discourse analysis.

CHAPTER 3

Pronoun Resolution

3.1 Introduction

The task of pronoun resolution has gained considerable attention in recent years within computational linguistics. Pronoun resolution refers to determining the specific entities to which pronouns refer within a given discourse. Unlike general anaphora, which involves a broader range of referential expressions, pronoun resolution focuses on resolving the referents of pronominal forms such as 'he', 'she', 'you', etc. These pronouns inherently depend on antecedent expressions, often pointing to entities introduced earlier in the text. While humans resolve such references effortlessly using contextual and cognitive cues, computational systems require explicitly designed algorithms to perform this task. Accurate pronoun resolution is essential for enabling sentence-level understanding and ensuring coherence across larger spans of text.

Scope of the study From the perspective of Western linguistics, anaphora and pronouns are not treated as fully synonymous concepts. However, they often overlap, particularly in cases involving pronouns such as he, they, her, etc. However, specific pronouns such as I, you, this, that, and what are generally not categorized as anaphoric, despite their referential nature. Consequently, within the Western theoretical framework, pronoun resolution and anaphora resolution are approached as distinct, with analytical emphasis typically placed on the anaphoric dimension. The literature review presented in this thesis reflects this orientation, drawing primarily from work in anaphora resolution.

In contrast, this chapter and the thesis as a whole adopt an Indian theoretical perspective. Within this framework, all pronouns possess inherent referring power, regardless of whether the referent is explicitly stated in the text. As such, all pronouns are potentially anaphoric in function (termed *parāmarśaka* in the Indian Grammatical Traditions). Therefore, while this chapter engages in a discussion parallel to that of anaphora resolution, it does so under the broader framework of pronoun resolution, unbound by the categorical constraints of Western linguistic theory.[1]

3.2 Litrature Review

Significant research has been conducted in the Western world, resulting in the development of established modules and programs for anaphora resolution for Western as well as Indian languages.

3.2.1 Western Theories

Early work in anaphora resolution focused on heuristic models, which rely on a set of linguistic rules to identify the most likely antecedent. A foundational contribution in this area was made by Hobbs (1978), whose pronoun resolution algorithm remains widely used today. Hobbs' model is based on a set of heuristic rules, including proximity, grammatical constraints, and governing theories. These rules help resolve pronouns by identifying the most plausible antecedent based on the structure and meaning of the surrounding discourse.

One of the most influential theoretical frameworks in anaphora resolution is Centering Theory (Grosz et al., 1995), which focuses on the notion of the "Center" in discourse. Centering Theory proposes that each sentence in a discourse has a central element, referred to as the "Center," which serves as the focus of the sentence. The theory categorizes transitions between these Centers into three types: continue (where the Center remains the same across sentences), retain (where a new Center is introduced but retains the previous one), and shift (where a completely new Center is

[1]A more detailed exposition of this shift in perspective, along with the resulting classification and limitations, is provided in Section 3.2 of this chapter.

introduced). By analyzing the transitions between these Centers, Centering Theory provides insights into how anaphoric expressions are resolved in discourse, particularly how pronouns relate to their antecedents in terms of focus and prominence.

In recent years, computational approaches to anaphora resolution have evolved, leveraging machine learning and deep learning techniques such as the Mention Ranking Model by Lee et al. (2017), which employs neural-network-based methods to rank potential antecedents for a given mention. This model learns to detect mentions and rank their antecedents using span representations and neural scoring, offering a more sophisticated approach than traditional heuristic models. The integration of contextual factors into mention ranking has significantly enhanced the performance of anaphora resolution systems.

Another notable contribution is the Ptr-Net model by Song et al. (2018), which utilizes a neural pointer network architecture combined with contextualized representations. The model employs a novel pointer mechanism to directly select antecedents for each mention, enabling more accurate coreference resolution by effectively capturing the broader discourse context. Due to its strong performance in modeling complex referential relationships, Ptr-Net has been influential in advancing state-of-the-art systems for coreference and anaphora resolution.

The Gazetteer method by Neville (1999) uses predefined lists of proper nouns, locations, and other entities to assist in identifying mentions in text. By recognizing these entities, the method supports anaphora resolution by providing reliable antecedent candidates, particularly when pronouns refer to specific named entities, thereby improving resolution accuracy.

GuiTAR (Graph-based Universal Information and Textual Anaphora Resolution) by Senapati and Garain (2013) is a graph-based model that resolves anaphora by linking co-referring expressions across discourse. It leverages contextual and syntactic features to disambiguate references, thereby enhancing the accuracy of anaphora resolution.

Conditional Random Fields (CRFs) by Lafferty et al. (2001) are used in anaphora resolution to model dependencies between words, enabling accurate prediction of a pronoun's antecedent. By incorporating features such as syntactic cues and proximity, CRFs improve resolution performance, especially in complex syntactic structures.

3.2.2 Efforts in Anaphora Resolution for Indian Languages

Bangla Senapati and Garain (2012) proposed an approach to pronominal resolution that utilizes global discourse knowledge in combination with traditional features. This method significantly improved the accuracy of identifying antecedents by considering the entire discourse context. Similarly, Senapati and Garain (2013) adapted the GuiTAR system to Bangla, modifying the language-specific preprocessing modules and its anaphora resolution component to suit the syntactic characteristics of Bangla. Sikdar et al. (2013) adapts the BART system for Bengali anaphora resolution, achieving a 50.80% F-measure and addressing the challenges of domain adaptation in resource-poor languages.

Hindi Prasad and Strube (2000) examined discourse salience in Hindi pronoun resolution through the lens of centering theory. They found that grammatical function (e.g., subjecthood) outweighed word order in determining referential prominence. Further advances in machine learning-based systems include the work by Vaswani et al. (2015), who developed models for Hindi to resolve anaphora, focusing on identifying pronouns and their antecedents in discourse. Sikdar et al. (2016) developed a generalized framework for anaphora resolution in resource-scarce Indian languages, including Hindi, integrating language-specific syntactic and semantic features with strong performance across co-reference metrics.

Konkani Navelker and Pawar (2024) propose a rule-based anaphora resolution system that emphasizes gender agreement to handle the challenges of a low-resource language.

Malayalam The VASISTH system, developed by Sobha and Patnaik (2000), represents one of the earliest efforts in anaphora resolution for Indian languages. The initial version employed a rule-based approach grounded in lightweight parsing and morphological analysis to resolve pronominals, reflexives, and ellipses. A more refined version was introduced in Sobha and Patnaik (2014), extending coverage to intra- and inter-sentential anaphora with improved rules and linguistic resources. Deepthi and Prabhakar (2014) further developed an anaphora resolution system tailored to Malayalam that enhances the modeling of discourse coherence by tracking referential continuity across narrative texts.

Marathi Khandale and Mahender (2019) presented a rule-based approach grounded in Marathi transformational grammar and binding theory. The system incorporates gender and number agreement, as well as animistic knowledge, enhancing the accuracy of anaphora resolution in Marathi texts.

Punjabi Kaur et al. (2020) develop an anaphora resolution framework achieving 47% accuracy and highlighting the difficulties in processing Punjabi text.

Telugu Mohan et al. (2010) developed a rule-based system focusing on pronominal reference resolution using syntactic and semantic cues. Their approach addressed challenges specific to Telugu, such as its morphology and syntactic flexibility.

Tamil Devi and Rani (2012) explores pronominal resolution using Conditional Random Fields (CRFs). The paper introduces a CRF-based approach for resolving pronouns, laying foundational work on handling Tamil's complex syntax. Building on this, Ram and Devi (2013) present an enhanced method using Tree CRFs incorporating syntactic tree structures.Sikdar et al. (2016) also evaluated their generalized framework on Tamil, achieving strong performance through language-specific adaptations.

multilingual efforts Annan et al. (2021) address the challenge of anaphora resolution within dialogue systems tailored for South Asian languages. Their work highlights the complexities arising from linguistic diversity and conversational context, proposing methods that improve pronoun and reference resolution in multilingual settings.

3.2.3 In Sanskrit

Anaphora resolution in Sanskrit presents unique computational challenges due to the language's complex morphology, flexible word order, and reliance on context-sensitive semantics. One of the foundational contributions in this domain is by Jha (2008), who provide a comprehensive linguistic classification of Sanskrit anaphors. Their work identifies various types of anaphoric expressions, such as pronominal, reflexive, and elliptical forms, and contextual dependencies that affect resolution. They emphasize the influence of rich case marking and free constituent ordering, which complicates the direct application of standard syntactic heuristics from resource-rich languages. This study is notable not only for its descriptive insights but also for establishing an empirical foundation upon which future computational models could be built.

Building upon these linguistic insights, Pralayankar and Sobha (2010) introduce a rule-based algorithm tailored to Sanskrit's morpho-syntactic structure. Their system integrates syntactic and semantic rules derived from *Pāṇinian* grammar to resolve intra-sentential anaphora. By leveraging clause boundaries, verb agreement, and case relationships, the model attempts to identify likely antecedents for pronouns and reflexives. Their experiments demonstrate that rules grounded in classical Sanskrit grammar, particularly those accounting for kāraka roles, yield promising results in constrained domains such as legal and literary texts. The rule-based nature of their approach makes it both interpretable and adaptable to domain-specific corpora.

Gopal (2012) presents a rule-based approach to anaphora resolution in Sanskrit using narrative texts from the Panchatantra. The work proposes SARS (Sanskrit Anaphora Resolution System),

which identifies lexical anaphors by leveraging syntactic roles and case markers within a structured rule framework. This study stands out as one of the earliest practical applications of computational linguistics to classical Sanskrit literature. Gopal and Jha (2014) propose an enhanced version of the Sanskrit Anaphora Resolution System (SARS) that combines rule-based methods with POS-tagged and morphologically enriched corpora. The system addresses both anaphora and cataphora, demonstrating improved resolution accuracy in Sanskrit texts.[2]

Collectively, these studies provide a computational foundation for Sanskrit anaphora resolution; however, they notably lack a comprehensive engagement with traditional or theoretical linguistic frameworks, highlighting an area for future research.

3.3 Sarvanāma : Pronouns

Pronouns are linguistic forms or expressions functioning as substitutes for nouns and noun phrases (Lyons, 1977). According to *Pāṇini*, the grammatical tradition identifies a technical category called Sarvanāma (pronoun/nominal) that comprises 49 elements listed across ten aphorisms in the Aṣṭādhyāyī (1.1.27–36), which are known to behave morphologically similarly. We are not considering this extensive, conservative set; instead, this study focuses on only nine words that qualify as pronouns, semantically categorized into Personal (*yuṣmad, asmad, bhavatu*), Relative (*yad*), and Demonstrative (*tyad, tad, etad, idam, adas*) categories due to their clear anaphoric behavior.

3.4 Power of reference

When we state that pronouns are used in place of nouns or noun phrases, we essentially mean that pronouns rely on nouns, which serve as their referents, for their meaning. Pronouns alone

[2]This work appears in the lecture-note series- Lecture Notes in Artificial Intelligence (LNAI), volume 8387. It is not clear whether it has been published elsewhere or in any other form.

do not convey a complete sense. Their significance is contingent upon contextual references. This relationship has vast potential, and without a defined context, meaning cannot be effectively communicated.

To explore the foundational aspects of this referring power inherent in a pronoun *(Sarvanāma)*, we refer the statement from *Vaiyākaraṇa Siddhānta Mañjūṣā* by Nagesh (1989, p. 1163)-
Sanskrit: *vaktṛśrotṛbuddhisthatvameva hi prakaraṇam. sarvanāmnāṃ buddhisthaparāmarśakatvamiti.*
English: The referent of a pronoun is always bound to the cognitive status. Context means one that is there in the speaker's and the hearer's cognition.

This statement provides various insights:

- All *Sarvanāma pada* possesses the power of *parāmarśa*, meaning referring. Thus, they are anaphoric in nature.
- This *parāmarśa* or reference is always *buddhistha* or cognitive, and sometimes *pūrva* or preceding. *Śaktivādaḥ* describes it as *'buddhi-stha-arthasya upalākṣaṇikaṃ smaraṇam'*.[3] The referent must already be imprinted in the minds of both the speaker and the hearer, rather than being unknown. And such a reference is recalled at the required time via delimitations on the referent.
- Such referencing endows pronouns with their meaning.
- This power of endowment is not unlimited; it is context-bound (*prākaraṇika*). In simpler terms, while *tad* (that) can potentially refer to anything and everything in the world, in reality, it pertains only to those things or concepts that are contextually relevant.

Suspected polysemous nature of pronouns: According to *Śaktivāda* ((Bhatta, 1994)), *sarvanāma*s are linguistic forms that can potentially substitute for all nouns or expressions. This raises

[3]The remembrance of mental presence of an object

the question of whether pronouns should be regarded as inherently polysemous. Unlike genuinely polysemous lexical items such as *hari*, which conveys distinct senses (e.g., *Viṣṇu*, monkey, lion), pronouns like *tad* denote varied referents such as *devadatta*, *ghaṭa*, or *paṭa* through a single unified signification. The possibility of pronouns being polysemous is, however, refuted on the following grounds:

1. A pronoun can denote only one entity at a time.
2. A pronoun can refer only to entities already known to both speaker and hearer.
3. In reference, a pronoun does not denote the external entity directly (e.g., *ghaṭa* as delimited by *ghaṭatva*), but rather the mental image of that entity (delimited by *buddhi-viṣayatva*) as present in the cognition of the speaker. Such mental presence constitutes the context, or *prakaraṇa*.

This expressive power is elaborated in *Śaktivāda* by Bhatta (1994). as:

> *Vaktṛbuddhisthā upalakṣitā viṣayatāvacchedakāvacchinne sarvanāmnā pada-nirūpitā śaktiḥ*

Meaning, the semantic flexibility is enabled by the mechanism of delimitation (*avacchinnatā*), whereby the pronoun's referential power (*śaktiḥ*) is understood to reside in its capacity to signified entities (*padārtha*s), such as *ghaṭa*, that are characterized by qualifying attributes (*ghāṭatva*, etc.) externally associated (*upalakṣita*) with the cognitive presence of the speaker's referent.

From this understanding, three critical aspects of pronominal reference emerge:

1. Pronouns function as stand-ins for discrete referents (*padārtha*s) such as *devadatta, ghaṭa*, or *paṭa*.
2. The reference of a pronoun is constrained by the specific delimiting property inherent in the object (*tad-tad-padārtha-avacchinnatā*).

3. Pronouns do not refer to the object as they are in the speaker's mind in general, but pronouns evoke the memory *(smaraṇam)* of the association with the delimiting property. Thus, such delimiting properties, such as *ghaṭatva, paṃatva*, etc., are adventitious *(upalakshaṇam)*.

3.5 Referents

Essentially, all Pronouns refer to an object (pot, cloth, etc.) within the knowledge domain of the speaker and hearer[4] (Pāṇḍuraṅgī Vīranārāyaṇa, 2020, p. 28). In addition to its cognitive property, *śaktivādaḥ* constrains the set of referents associated with each enumerated pronoun as follows:

3.5.1 Personal Pronouns

The grammatical category of personal pronouns depends upon the notion of participant-roles (i.e., whether they refer to the speaker, the addressee, or a third party) in a given language (Lyons, 1977). Sanskrit recognizes three persons, of which the first two are classified as personal pronouns.

Gadādhara, a writer of the text Śaktivāda, was not keen on generalizing the referents of the first-person (asmad) and second-person (yuṣmad)[5] pronouns simply as the speaker and the addressee, respectively (As seen in 1a and 2a). Rather, he preferred identifying their referents contextually in each use case.

Referents of *Asmad*

According to Gadādhara, determining the referent of the first-person pronoun *asmad* requires consideration of syntactic independence, embedding, communicative intention, and speaker roles. He proposes three discourse situations in which the referential power of *asmad* is established.

[4] *sarvanāmapadena buddhi-sthāḥ ghaṭapaṭādayaḥ pratyānyante.*

[5] When the pronoun yuṣmad is mentioned, it should be considered applicable to all the pronouns in the domain of second person. Such as: bhavatuṃ

1. **Situation 1: Independent utterance (*svatantroccāritatva*)** The referent of *asmad* is fixed when the pronoun occurs in an independently uttered sentence. *Svatantroccāra* or the independent utterance is the basis for the attribution of referential qualification through the basis of usage of the particular word (*pravṛttinimitta*) in this case *Asmad*. It refers to an utterance consisting solely of the first-person pronoun by the original speaker, not dependent on any intention to convey the meaning of a larger phrase where the pronoun would serve as the *karma* of a speech verb such as *uvāca*.[6] The essential idea is that the utterance stands independently and is not syntactically or semantically embedded.

2. **Situation 2: Embedded utterance** The *kartā* of an activity, whose *karma* is the meaning of the sentence consisting of the first person pronoun *Asmad*.

   svaghaṭita-vākyārtha-karmaka-vākyāntarastha-kriyākartari

   When the sentence containing *asmad* is embedded within another sentence, the referent is the *kartā* of the matrix sentence whose *karma* is the subordinate sentence.

3. **Situation 3: Speaker-based optional reference** Even in embedded contexts, the actual speaker of the utterance (*kevalam-uccārayitṛ*) may serve as the referent if the communicative intention requires it (*tātparya-sattve*). This situation is optional and is invoked when the contextual intention makes the speaker the most coherent referent.

Together, these three situations form the theoretical foundation for the referential behaviour of *asmad*. We now turn to the examples that illustrate the application of each situation.

Examples

1a: Sanskrit: *ahaṃ sundaraḥ asmi.*

Gloss: I[nom,sg] Beautiful[m,nom,sg] to-be[present,1p,sg].

[6]Translation by V. P. Bhatt (Bhatta, 1994, p. 184-185) of
vākyāntarastha-kriyākarmatāpanna-svaghaṭita-vākyārtha-pratyayanechchhā-anadhīnatvam

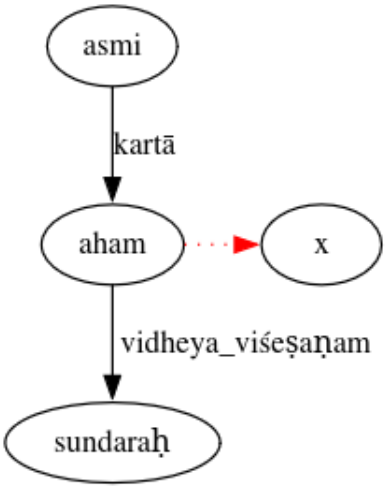

Figure 3.1: Example 1a - Direct statement with pronoun Asmad

English: I am Beautiful.

In this independent, non-embedded utterance, the speaker say Chaitra, expresses a self-referential statement. There is no higher verb governing the sentence and no embedding. Therefore, Situation 1 applies, and the referent of *asmad* is the speaker himself. See figure 3.1

1b. Sanskrit: *ahaṃ sundaraḥ asmi. iti caitraḥ vadati*

Gloss: I[nom,sg] beautiful[m,nom,sg] to-be[present,1p,sg]. so[indcl] Caitra[m,nom,sg] say[present,3p,sg].

English: "I am beautiful," says Chaitra.

Here, the utterance "*ahaṃ sundaraḥ*" is embedded within a larger sentence governed by the verb "says." Thus, the embedded sentence becomes the *karma* of "says," and according to Situation 2, the referent of *asmad* is Chaitra, the *kartā* of the matrix sentence. See figure

In addition, if Maitra is the speaker reporting the statement, Situation 3 allows that, if the context conveys such intention, the referent could optionally be the actual speaker (*kevalam-uccārayitṛ*). Thus, both Chaitra (Situation 2) and Maitra (Situation 3, if intention exists) are theoretically eligible referents. See figure 3.2

1c. Sanskrit: *Devī uvāca "mayā tvayi hate atraiva garjiṣyanti āśu devatāḥ "iti Medhāmaharṣī Suratham prati uvāca.*

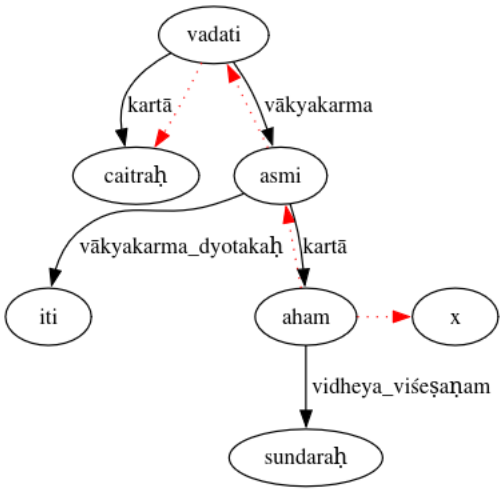

Figure 3.2: Example 1b - Embedded statement with pronoun Asmad

Gloss: Devi[f,nom,sg] say[avo,3p,sg]. I[instr,sg] you[loc,sg] kill[past.part] here[indcl] emph roar[fut,3p,pl] immediate[indcl] God[m,nom,pl]. So[indcl] Medhamarshi[m,nom,sg] Suratha[m,acc,sg] Say[avo,3p,sg].

English: Devi said, "When you are killed by me, the gods will immediately roar here." Thus was this said to King Suratha by sage Medhamaharshi.

The original utterance is embedded first within Devi's speech, and then both are embedded within Medhamaharshi's report. Because the subordinate sentence ("mayā tvayi hate…") becomes a *karma* within this multilayered structure, Situation 2 applies. The referent of *mayā* can be attributed to the *kartā*s of the matrix sentences (Devi, and hierarchically also Medhamaharshi). As explained in Chapter 2, the subordinate sentence *avāntaravākya*'s meaning merges into the *mahāvākya*, so Situation 2 governs such complex embeddings. In addition, as per situation 3, if another speaker utters the sentence, he/she is also eligible to be the referent. See fig 3.3

1d. Sanskrit: *ahaṃ paṇḍitaḥ asmi. iti ayaṃ devadattaḥ jānāti*

Gloss: I[nom,sg] scholar[m,nom,sg] to-be[present,1p,sg]. so[indcl] he[m,nom,sg] devadatta[m,nom,sg] know[present,3p,sg].

English: This Devadatta knows that "I am a scholar."

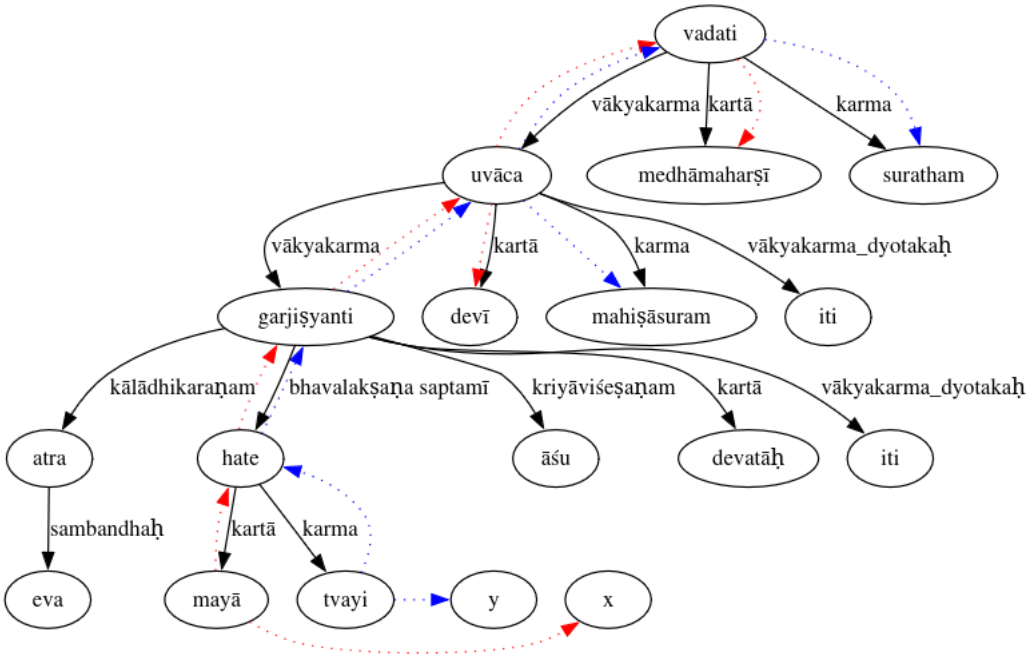

Figure 3.3: Example 1c - Reported, Embedded statement with pronoun Asmad

As previously noted for examples 1b and 1c, when the matrix clause contains predicates other than knowledge predicates, the *kartā* is primarily interpreted as the referent. In contrast, in example 1d, where the matrix clause involves a knowledge predicate, it would be pragmatically odd to interpret the *kartā* as reporting knowledge to themselves - a reading incompatible with Situation 2. Therefore, the referent is instead taken to be the speaker, as in Situation 3. However, if the strong intention at play is one of self-assertion (e.g., the speaker clarifying or reaffirming their own claim), then Situation 2 may apply, and the referent can optionally be the speaker if the context supports such a reading, rather than the *kartā*, here *Devadatta*. See Figure 3.4.

Taken together, these examples clarify how the three situations differ in principle: Situation 1 involves independent, self-referential expressions; Situation 2 demonstrates cases in which reference is fixed by the *karma* role within an embedded construction; and Situation 3 shows contexts where the speaker's intention allows for optional reference. Viewed collectively, they outline a coherent framework for understanding how *Asmad* determines its referent across a range of syntactic and pragmatic settings.

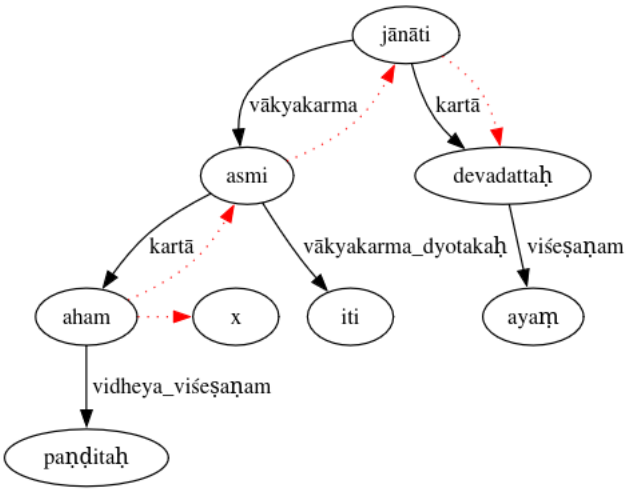

Figure 3.4: Example 1d - Embedded statement with knowledge predicate with pronoun Asmad

Referents of *yuṣmad*

The attribution of the referent of the second-person pronoun *yuṣmad* ("you") follows the same theoretical trajectory as that of *asmad*, but with addresseehood as its central semantic property. Gadādhara outlines three situations in which the referential power of *yuṣmad* is determined.

1. **Situation 1: Addressee of the independent utterance (*svasambodhyatva*)** *Svasambodhyatva* refers to the basis for attributing the referential power of a word according to its usage, in this case the second-person pronoun *yuṣmad* (*padapravṛttinimitta*). It is further defined as the speaker's intention for the other person to be the locus or object (*viṣaya*) of the cognition of the intended meaning of a sentence containing a second-person pronoun. Importantly, this utterance does not depend on any other sentence or context beyond the one containing the second-person pronoun.[7] In simpler words, the referent of *Yuṣmad* is the receiver of the intended information conveyed by an independent utterance.

[7] *Vākyāntarastha–kriyākarmatvena svaghaṭita-vākyārthānavagāhinī svajanya-pratīti–āśrayatvena icchā–viṣayatvam* simplified paraphrasing of the translation - (Svāsaṃbodhyatva)means that the referent of the second person pronoun *Yuṣmad*, should be perceived as being the object (qualified) of the desire (intention) of the speaker, that the other person should be the abode of the cognition produced from the sentence consisting of the second person pronoun. However, here such an addresseeness must be held to be not dependent on the speaker's intention that the meaning of the phrase consisting of a second person pronoun is the object of the action of saying expressed by the verb in the other principle (indirect) sentence (speech). - by V. P. Bhatt in (Bhatta, 1994, p. 187).

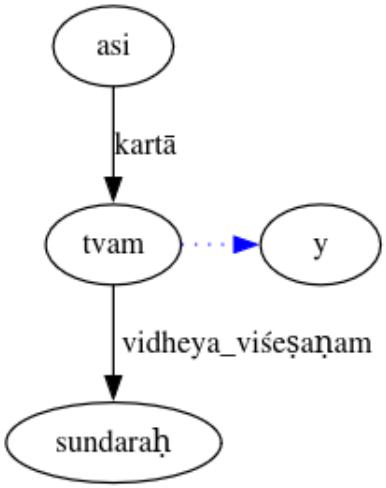

Figure 3.5: Example 2a - Direct statement with pronoun Yuṣmad

2. **Situation 2: Embedding-based determination** The *karma* of an activity, whose *karma* is the meaning of the sentence consisting of the second person pronoun *Yuṣmad*.

   svaghaṭita-vākyārtha-karmaka vākyāntarastha-kriyā-karmaṇi

   This situation prioritizes syntactic embedding over speaker intention.

3. **Situation 3: Direct-addressee** (*kevalaṃ saṃbodhye śaktiḥ*) When the context guides, the referential power lies in the actual addressee of the utterance, even in subordinate sentences. Such an addressee can be explicitly or implicitly indicated by the use of vocative case marking.

We now examine examples that demonstrate the application of these three situations.

2a: Sanskrit: *tvam sundaraḥ asi*
Gloss: you[nom,sg] beautiful[m,nom,sg] to-be[present,2p,sg]
English: You are beautiful.

Here, the speaker attributes beauty to the addressee (e.g., Maitra) with reference to situation 1. The sentence is not embedded, and Maitra, who is the intended locus of the speaker's intention to convey the prescribed meaning, becomes the referent. See fig 3.5

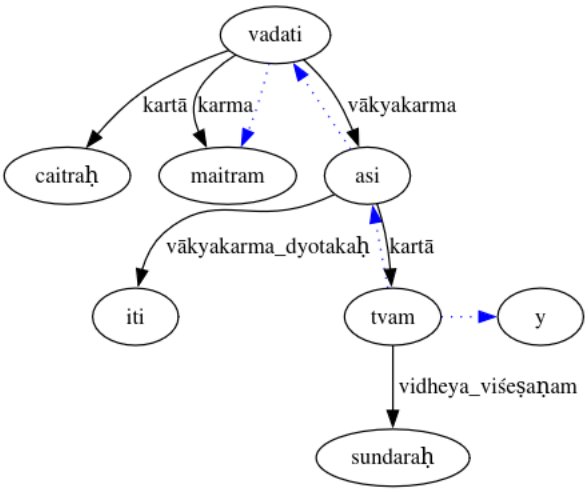

Figure 3.6: Example 2b - Embedded statement with pronoun Yuṣmad

2b. Sanskrit: *tvam sundaraḥ asi. iti caitraḥ maitram vadati*

Gloss: you[nom,sg] beautiful[m,nom,sg] to-be[present,2p,sg]. so[indcl] Caitra[m,nom,sg] maitra[m,acc,sg] say[present,3p,sg].

English: "You are beautiful," says Chaitra to Maitra.

Here, the utterance "*tvam sundaraḥ*" is embedded within a larger sentence governed by the verb "says." Thus, the embedded sentence becomes the *karma* of "says", and according to Situation 2, the referent of *yuṣmad* is Maitra, the *karma* of the matrix sentence.

Additionally, Situation 3 allows that, if the context conveys such intention, the referent could optionally be the actual addressee of the utterance, say *Devadatta* (*kevalam-sambodhya*). Thus, both Maitra (Situation 2) and Devadatta (Situation 3, if intention exists) are theoretically eligible referents. See figure 3.6

In Example 1c (from above), the sentence *mayā tvayi hate . . .* is embedded under *devī uvāca* as its *karma*, is again embedded in a sentence *Medhāmaharṣī Suratham prati uvāca*. Due to the multiple embeddings, the *karma* of both matrix sentences, i.e. *Mahiṣāsura* and *Suratha*, are eligible to be the referents of *yuṣmad* pronoun. Additionally, if context suggests, the actual receiver/addressee is also eligible to be the referent as per situation 3. See figure 3.3

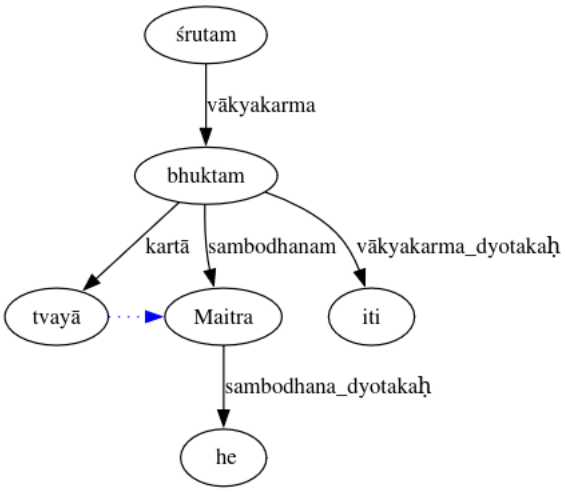

Figure 3.7: Example 2c - Embedded statement with pronoun Yuṣmad in vocative case

2c: Sanskrit: he Maitra! tvayā bhuktam iti śrutam

Gloss: Oh[indcl] maitra[m,voc, sg], you[m,instr,sg] eat[past.part] so[indcl] heard[past.part]

English: O Chaitra! I have heard that you have eaten.

The embedded sentence *tvayā bhuktam* lies within a reporting construction *iti śrutam*. However, in this case, there is no explicit *karma* to the matrix sentence. So, as per the 3rd situation, the mere addressee, i.e., Maitra, indicated by the vocative case, becomes the referent. See figure 3.7

To conclude, the referential mechanics of *yuṣmad* operate within a hierarchical framework: Situation 1 applies when structural independence and speaker intention align; Situation 2 applies when structural embedding governs reference; Situation 3 applies when direct address alone determines the referent. Together, these situations explain how the Sanskrit semantic theory systematically accounts for the behaviour of second-person pronouns across varied syntactic and pragmatic contexts.

3.5.2 Relative and Co-relative Pronouns

According to *Gadādhara*, the referring power of the pronoun *yad* (what) resides in the entity qualified or delimited by a property (*dharmāvacchinna*) with which it resides in the speaker's mind

(*vaktṛ-buddhi-viṣayatā-vacchedakatva-upalakṣita*). Such an entity is expressed by the pronoun *tad*[8] (that) as well (*tatpratipādya*). And the pronoun *tad*'s referring power (*tatpadaśaktiḥ*) resides in an entity which is expressed by the pronoun *yad* (*yatpadaupasthāpya*).

Therefore, both *yad* and *tad* possess expressive power and exhibit syntactic as well as semantic expectancy for each other, allowing them to co-occur coherently in discourse, even though they differ in function and direction of reference and are not mutually dependent.

Gadādhara distinguishes between relative (*yad*) and demonstrative or correlative (*tad*) pronouns based on their distinct referential roles in meaning comprehension. The relative pronoun *yad* is anticipatory and introduces an entity yet to be fully identified, typically by qualifying it with a property present in the speaker's cognition. For example, in (3a.):

3a. Sanskrit: *yaḥ iha asti, taṃ ānaya*

Gloss: who[m,nom,sg] here to-be[present,3p,sg] that[m,acc,sg] bring[imp,2p,sg]

English: Bring the one who is here.

yaḥ refers to an entity by specifying its location (*iha asti*), thereby expressing meaning through qualification.

In contrast, *tad* is retrospective and refers back to an entity already introduced or assumed in the discourse, as in (3b.):

3b. Sanskrit: *ghaṭaḥ tatra asti. taṃ ānaya*

Gloss: pot[m,nom,sg] there to-be[present,3p,sg] that[m,acc,sg] bring[imp,2p,sg]

English: The pot is there. Bring it.

In examples (3a.) and (3b.), *taṃ* (the masculine accusative singular form of *tad*) merely recalls *ghaṭaḥ* without contributing any additional descriptive content. Thus, while *yad* carries expressive force by qualifying the referent, tad serves a primarily referential function, relying upon prior mention or memory. The difference between (3a.) and (3b.) is both syntactic and semantic.

[8]Forms of the pronoun *tad* etc. are the same but function as both co-relative and demonstrative.

Syntactically, in (3a.), the presence of tad is dependent upon *yad*, which introduces the entity being referred to. In (3b.), however, tad occurs independently of *yad*, functioning autonomously. Semantically, one may observe that in (3a.), where *tad* and *yad* co-occur, the meaning is situational, context-dependent, and somewhat abstract, whereas in (3b.) it is factual and demonstrative.

According to Nāgeśa, in the case of *prakramyamāna-parāmarśaka* usage - where the referent is expected to appear in the subsequent portion of the discourse - the pronouns *yad* and tad exhibit a constant expectancy (*niyamena apekṣā*). By contrast, in the case of *prakrānta-parāmarśaka* usage, the expectancy of yad and tad (together with the conditions of *prasiddha* 'already known' and *anubhūta* 'already experienced' in the case of tad) is eventual rather than constant. Nāgeśa further proposes that in contexts where *yad* and tad occur together, their functions should be delimited: *yad* by *uddeśyatva* (subject-hood) and tad by *vidheyatva* (predicate-hood). Additionally, he specifies other delimitors that apply to both pronouns, namely *parokṣatva* (non-immediacy to perception, i.e., not being in front of the eyes) and *vaktṛ-buddhi-viṣayatva* (being present in the speaker's cognition).

Even though there exists an expectancy between paired markers, there is nevertheless an optionality regarding using a single marker. For the sake of ease in processing, we adopt the following convention: when a marker derived from any inflection of the *yad* stem is used alone, without the explicit presence of its counterpart derived from the *tad* stem, it shall be interpreted as an instance of a relative-correlative construction. In such cases, the presence of the corresponding *tad*-based marker is implicated. However, when a marker derived from any inflection of the *tad* stem occurs independently, the interpretation becomes categorically ambiguous between treating it as a correlative marker or as a simple demonstrative. In the absence of sufficient evidential cues, we adopt the convention of treating such lone *tad*-based markers as demonstratives.

3.5.3 Demonstrative Pronouns

Third-person pronouns *(prathama puruṣaḥ) - tyad, tad, yad, etad, idam, adas* (that generally demonstrate object, time, place, etc.)

The group of third-person pronouns can refer to a wide range of entities as intended by the speaker, excluding the addressee and the speaker. However, at any given time, a third-person pronoun refers to only one specific entity. Thus, it is accurate to state that a pronoun operates on a multi-referential basis in terms of potential reference but is mono-referential in its actual application, depending on the speaker's intentions and the specific referent they wish to indicate with the chosen pronoun.

According to the Śaktivāda (Pāṇḍuraṅgī Vīranārāyaṇa, 2020), demonstrative pronouns are classified in two distinct ways.

1. *idam* and *etad* fall under *pratyakṣa-buddhi-viṣaya*, wherein the referent is immediately available either through direct perception or through the mental presence of the object.

2. *adas* and *tad* are categorized as *parokṣa-buddhi-viṣaya*, in which the referent denoted by the pronoun is not perceptually available to the speaker but is recalled or inferred through mental representation.[9]

Nagesh (1989, p. 1163)[10] explores four fold classification:

4a. *idam* is used when referent is close and visible to eyes *(sannikṛṣṭe)*.

For example:

Sanskri : idam vātāyanam.

Gloss : This[n,nom,sg] window[n,nom,sg]

English: This is a window.

[9] *Hariṇātha*, in his commentary, further clarifies this distinction: *adas* refers exclusively to entities that are obstructed or intervened (*vyavahita*) and thus not visible, whereas *tad* designates other non-visible referents.

[10] *idamastu sannikṛṣṭe samīpataravarti caitado rūpam. adasastu viprakṛṣṭe taditi parokṣe vijānīyāt.*.

4b. *etad* is used when referent is very close *(samīpataravarti)*.

For example:

Sanskri : etad upanetram.

Gloss : This[n,nom,sg] spectacles[n,nom,sg]

English: These are spectacles.

4c. *adas* is used when the referent is far enough but visible to the eyes *(viprakṛṣṭe)*.

For example:

Sanskri : asau parvataḥ.

Gloss : That[m,nom,sg] mountain[m,nom,sg]

English: That is a mountain.

4d. *tad* (also, *tyad*)[11] is used when the referent is not directly in front of the eyes *(parokṣe)*.

For example:

Sanskrit: saḥ sītāpatiḥ rāmaḥ.

Gloss: He[m,nom,sg] sita-husband[m,nom,sg] Ram[m,nom,sg]

English: He is Sita's husband, Ram.

3.6 Direction of reference

According to *śaktivādaḥ*, the pronouns (including *'tad, asmad, yusmad* etc.') can be classified into two categories, determined by the direction of reference: *prakramyamāṇaparāmarśakāni*, that is referent being succeeding and *prakrāntaparāmarśakāni*, that is referent being preceding[12] as mentioned by Pāṇḍuraṅgī Vīranārāyaṇa (2020, .p 28).

[11](www.ashtadhyai.co/sutraani/1/1/27)

[12]*tat ityādini padāni prakrāmyaparāmarśakāni, prakrāntaparāmarśakāni iti dvividhāni*

3.6.1 Prakramyamāṇa : Anaphora

If the referent noun is positioned to the left of the referring expression, this scenario pertains to the *prakrāmyaparāmarśaka* pronoun (meaning, the one that has occurred before). This construction is analogous to anaphora, as anaphora denotes references to entities or concepts that have been previously mentioned in the discourse.

For example (5a.):

Sanskrit: śrībhagavān uvāca. aham ādiḥ hi devānām.

Gloss: God[m,nom,sg] say[past,3p,sg]. I[nom,sg] beginning emphatic God[m,gen,pl]

English: The God said – I am indeed the beginning of the gods

In this example the referent *śrībhagavān* of the Pronoun *aham* (*asmad*, nominative form) is antecedent to the sentence.

In relative–correlative constructions, particularly in *prakramyamāna-parāmarśaka* cases, the expectancy between the pronouns *yad* and *tad* is constant (*niyamena āpekṣā*), and not rather eventual and context-dependent.

3.6.2 Prakrānta : Cataphora

If the referent noun is positioned to the right of the referring expression, this scenario pertains to the *prakrāntaparāmarśaka* pronoun (meaning, the one whose reference is upcoming). This construction resembles cataphora, as cataphora anticipates and precedes the information that will be articulated later.

For example (5b.):

Sanskrit: kim tava nāmadheyam iti ācāryā chātram pṛcchati

Gloss: What[m,nom,sg] you[gen,sg] name[n,nom,sg]. So teacher[f,nom,sg] student[m,acc,sg] ask[present,3p,sg]

English: The teacher asks the student, “What is your name? ”

In this example, the referent - student of the pronoun - *tava* (*yuṣmad* genitive form) is in the precedence.

Special cases: This course applies to all pronouns. However, the pair of *yad-tad* (Relational-Corelational) pronouns follows a different trajectory in the context of the *prakrāmyamāṇa-parāmarśaka* case. In such instances, *yad-tad* indicates a constant mutual expectancy. The pronoun *yad* always expects the pronoun *tad* to be present. Such is the case when the referent of the pronoun *tad* is in precedence. In case of the referent of the pronoun being in precedence, pronoun *tad* invariably expects *yad* to be present[13] as seen in the following example as mentioned by Pāṇḍuraṅgī Vīranārāyaṇa (2020, p. 125).

For example:

Sanskrit: taṃ ānaya yaḥ iha asti

Gloss: that[n,acc,sg] bring[Imp,3p,sg] which[m,nom,sg] here to-be[present,3p,sg]

English: Bring him who is here.

These types of sentences exemplify co-referring constructions, characterized by a pair of sentences: one containing a form of the *yad* pronoun and its complementary counterpart featuring a form of *tad*. The forms of *yad* or *tad* may be either implicit or explicit, yet they maintain a relationship of constant expectancy, termed *nitya-sambandhaḥ*. This relationship underscores the interconnectedness of the sentences, wherein the reference in one sentence anticipates and relies upon the reference in the other.

For example:

Sanskrit: ye ca api akṣaram avyaktam. teṣāṃ ke yoga-vittamaḥ.

Gloss: who[m, nom, pl] and also imperishable[n, nom, sg] unmanifested[n, nom, sg]. of those [m, loc, pl] who[m, nom, pl] yoga-knower[m, nom, pl]

[13] *yatpadena arthapratyāyane niyamataḥ tatpadāpekṣā. iyam ca vyatpattiḥ prakramyamāṇaparāmarśaka yatśabdasya. prakramyamāṇaparāmarśakatat śabdenāpi niyamataḥ yat padam apekṣate*

English: And those who are also unmanifested and imperishable ones, among them, who are the best yogis?

In the example above, *ye* (plural nominative form of *yad*) co-refers with *teṣāṃ* (plural genitive form of *tad*) and they exhibit mutual expectancy.

3.7 Role of Padaikavakyata

The challenge of pronoun resolution and anaphora resolution more broadly can be addressed by applying the principle of *padaikavākyatā*. This principle offers a structured framework to resolve referential dependencies in discourse by (i) facilitating the joining of those fragments into coherent wholes and (ii) enabling comprehension of fragmented expressions.

According to *padaikavākyatā*, referent elements may undergo logical relocation across different syntactic or sentential positions to complete a discourse unit. This process can be subdivided into two main stages:

ia. Identification of the unit(s) to be relocated.

ib. Determination of the position to which the relocation is to occur.

The unit(s) to be relocated typically correspond to the referent(s): the antecedent in the case of anaphora, or the consequent in the case of cataphora. The key factor in determining the referent is morphological agreement between the anaphoric expression and the potential referent. Once the referent has been identified, it is not simply replaced in place of the anaphoric expression; rather, it complements it to form a semantically complete structure.

The number of relocations required is contingent upon both the number of referents and the number of anaphoric expressions. While a one-to-one correspondence often holds, it is not universally applicable. For example, when the anaphoric expression is plural, two scenarios arise:

1. If the referent itself is in plural form, then a single relocation suffices.

2. If the referent denotes a group composed of multiple members, each member must be individually relocated, resulting in multiple movements.

For example: (6a.)

Sanskrit: *rāmaḥ sītayā saha vanaṃ gacchati. saḥ tatra kuṭīṃ karoti.*

Gloss: Rāma[m,nom,sg] Sītā[f,inst,sg] forest[n,acc,sg] go[pres,3p,sg]. he[m,nom,sg] there hut[f,acc,sg] build[pres,3p,sg]

English: Rāma goes to the forest with Sītā. There, he builds a hut.

In this example, the pronominal expression *saḥ* ('he') is an anaphoric expression whose meaning remains incomplete without its referent. Through the principle of *padaikavākyatā*, the referent *Rāmaḥ* is identified based on its morphological properties and relocated just after the expression *saḥ*, producing the unit:

saḥ rāmaḥ tatra kuṭīṃ karoti ("He, meaning Rāma, builds a hut there").

This reconstructed sentence forms a semantically coherent and syntactically complete unit. Thus, *padaikavākyatā* functions as the foundational principle underlying such referential shifts.

The newly formed unit - "he, meaning Rāma, built the hut" - is internally comprehensible. Here, *saḥ* and *rāmaḥ* exhibit a *viśeṣaṇa-viśeṣya-sambandhaḥ* (modifier-modified relationship), which conforms to the dependency grammar framework. Without this shift, the second sentence lacks the necessary referential prerequisites (such as *Ākāṅṣā*) and thus remains incomplete. By contrast, with the application of *padaikavākyatā*, the sentence is rendered interpretable and semantically unified.

In *Paribhāṣenduśekhara*, *tasmīniti nirdiṣṭa pūrvasya* illustrates how a locative expression refers back to a preceding element, aiding the interpretation of rules like *iko yaṇaci*. The use of locative case triggers *Ekavākyatā*, linking referring terms such as *tasmīn* and *aci* to yield an integrated meaning. This mechanism serves as a use case (jñāpaka) for Pronoun Resolution, which is further elaborated in Chapter 2 (Vyākaraṇa section).

3.8 Representation

In this section, mainly, we ponder how this principle interacts with other parts of the sentence within the framework of a dependency structure. In this approach, we have adopted a dependency graph-based representation within the *Pāṇinian* framework, as it allows for expanding the predefined boundaries of the sentence using the same parameters employed in the dependency structure. We intend to extend this framework to include anaphora and discourse relations as well. In the following examples, we can see three kinds of relations embedded within a single graph.

1. Intra-sentential that is kāraka relation
2. Inter-sentential (sentence-level) discourse relation
3. Inter-sentential (word-level) anaphora relation

Example A :

Sanskrit: Yadā rāmaḥ mithilām gacchati. Tadā saḥ sītāṃ varati.

Gloss : When Ram[m,nom,sg] Mithila[f,acc,sg] go[present,3p,sg]. Then he[m,nom,sg] Sita[f,acc,sg] choose[present,3p,sg]

English: When Rama goes to Mithila, he chooses Sita.

Example B :

Sanskrit: yaḥ kāryaṃ karoti. saḥ phalaṃ labhate.

Gloss : who[m,non,sg] work[n,acc,sg] do[present,3p,sg]. He[m,nom,sg] fruit[f,acc,sg] get[present,3p,sg]

English: Who does the work, gets the fruit.

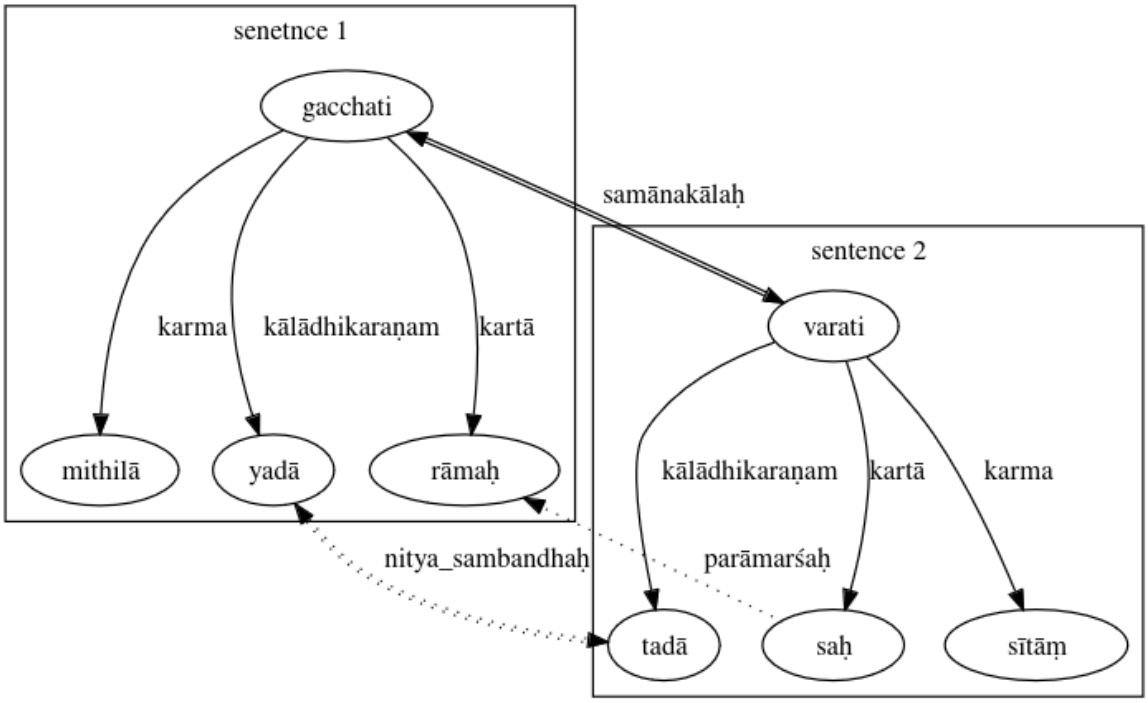

Figure 3.8: Pronominal Anaphora with Discourse Relation

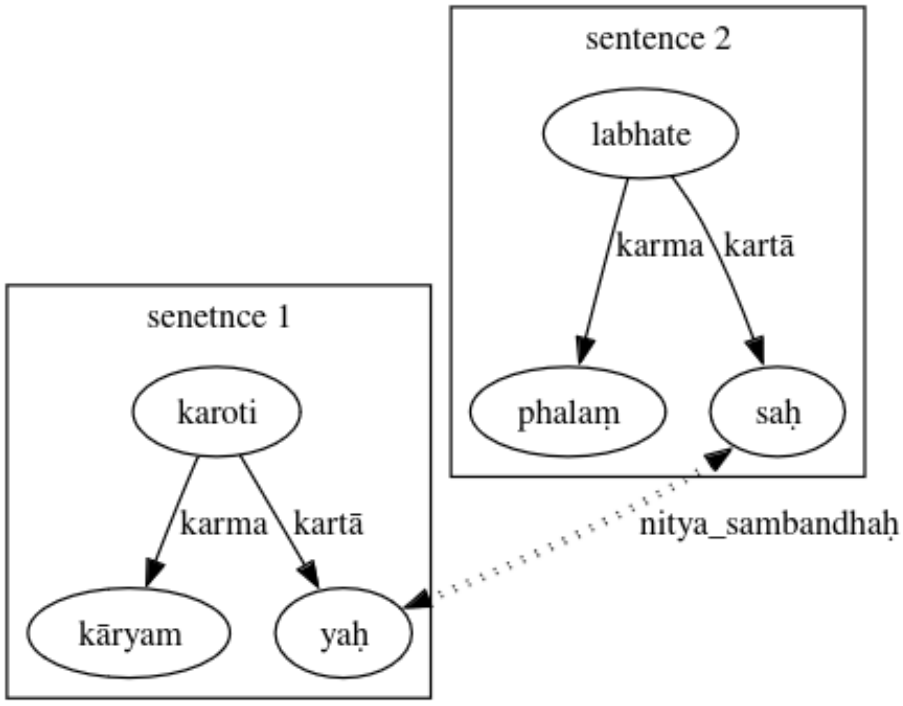

Figure 3.9: Coreferring Pronominal Anaphora

We should consider the following points :

1. *kāraka* Relations: Within a sentence, *kāraka* relations such as *kartā, karma* are represented by a single directed line, indicating the relationship between the head (finite verb) and its dependents.

2. Extension of Dependency Graphs: To extend the sentence graph, we incorporate discourse and anaphora relations using the same parameters, schemes, structures, and graphical representations.

3. Discourse Relations: In Figure 3.8, discourse relations such as *samānakālaḥ* are depicted using double-lined bidirectional arrows, indicating connections between two linearly arranged verbs. The relationship of constant expectancy is illustrated by double-dotted bidirectional arrows between the discourse markers of each sentence. Chapter 4 discusses these relations in detail.

4. Anaphoric Relations: In Figure 3.8, the anaphoric relation exists between the referring expression *saḥ* from sentence 2 and the referent *rāma* from sentence 1. This relation is represented by a single dotted directed line, termed *parāmarśaḥ*, meaning 'referring '. It is directed from the referring expression towards its referent. Such a relationship is illustrated at the word level; due to the concept of *padāikavākyatā*, there is no expected relation between verbs resulting from this anaphoric connection.

5. Anaphoric Relations: Figure 3.9 presents a pair of anaphoric expressions, each expecting the presence of the other constantly. This relationship, marked by double dotted bi-directional lines and termed *nitya_sambandhaḥ*, operates at the lexical level rather than the verb level (unlike discourse relations), governed by the principle of *padaikavākyatā*.

3.9 Corpus's Statistics

In the pilot survey of Pronominal Anaphora Relations, we analyzed a text from the *Saṅṣeparāmāyaṇam* consisting of 100 shlokas. The following statistical findings summarize the data collected:

3.9.1 Total Found Entries

- Total Entries: 84
  - Indeclinable Pronominal (PN) Entries: 28 (33.33%)
  - Inflectional PN Entries: 56 (66.66%)

3.9.2 Analysis of Indeclinable PN Entries

- Total Indeclinable PN Entries: 28
  - Total Discourse Relations: 14 (50%)
    - 12 instances of the marker *tataḥ* indicating *anantarakālaḥ* relation.
    - *Yathā* (1) and *tathā* (1) exhibiting *sādṛśya* relation.
  - Total Anaphoric PN Relations: 14 (50%)
    - *Tatra* (5 entries) representing location anaphora.
    - *Tadā* (9 entries) representing time anaphora.

3.9.3 Analysis of Inflectional PN Entries

- Total Inflectional PN Entries: 56
  - PN Inflections as *viśeṣaṇa*: 29 (51.78%)
  - PN Inflections as Anaphora: 27 (48.21%)

* *Asmad (1st person)*: 3 (11.11%)
    - *Aham*: 2 entries
    - *me*: 1 entry
* *Yuṣmad (2nd person)*: 3 (11.11%)
    - *Tvam*: 2 entries
    - *Tvayā*: 1 entry
* *Tat (3rd person - that)*: 16 (59.25%)
    - *Saḥ*: 6 entries
    - *Tam*: 2 entries
    - *Tasya*: 2 entries
    - *Tena*: 2 entries
    - *Te*: 1 entry
    - *Tasmīn*: 1 entry
    - *Tat*: 1 entry
    - *Tām*: 1 entry
* *Etat (3rd person)*: 2 (7.4%)
    - *Etat*: 1 entry
    - *Enam*: 1 entry
* *Idam (3rd person - this)*: 2 (7.4%)
    - *Asya*: 2 entries
* *Yat (3rd person - correlative)*: 1 (3.7%)
    - *Yaḥ*: 1 entry

3.9.4 Positioning of Anaphoric Entries

- Anaphora (Referent towards left): 32 (78%)[14]
  - Indeclinable Anaphora: 11
  - Inflectional Anaphora: 21
- Cataphora (Referent towards the right): 7 (17%)
  - Indeclinable Cataphora: 2
  - Inflectional Cataphora: 5
- Undecided[15] Entries: 2

This analysis elucidates the distribution and characteristics of pronominal anaphora relations within the *Saṅṣeparāmāyaṇam*, highlighting the significant role of both indeclinable and inflectional pronominal forms in the text. However, there was no example of co-referring, that is, a pair of pronouns in the text.

3.10 Algorithms on Pronominal Anaphora Resolution

Pronominal anaphora resolution is crucial for interpreting referents across sentences in Sanskrit discourse. Three main types of algorithms have been designed for this purpose: (1) for **Personal Pronouns (PPN)**, and (2) for **Referring-Co-referring Structure (RC)**.[16] Each operates with

[14]% is of total 41 anaphoric entries. That is 14 indeclinable and 27 inflectional.

[15]The referent is ambiguous and yet to be resolved, so the positioning of the referent remains undecided.

[16]We have excluded the category of demonstrative pronouns for constructing algorithms because finding the reference is highly contextual and subjective

distinct linguistic cues and resolution rules, but all converge toward identifying the referential linkage between pronouns and their antecedents.[17]

3.10.1 Personal Pronouns

Taking theoretical reference from Gadādhara's *Śaktivāda*, we have devised a primary algorithm to locate the referents of the words *asmad* ("I") and *yuṣmad* ("you"). The apparent complexity does not stem from the theory but from the highly technical language of Navya-Nyāya in which it is expressed. When this terminology is unpacked, the underlying algorithms are revealed to be straightforward and direct. This enables their expression as a compact and computationally viable set of rules. It is important to consider the dependency graphs of the concerned sentences generated by existing syntactic parsers. These graphs are useful both for comprehension and for traversing the path from the pronoun to its referent.

The heuristic rules are as follows:

1. ***Vākyāntarastha-kriyākartari:asmad*** and ***Vākyāntarastha-kriyākarmaṇi:yuṣmad*** The *kartā* and *karma* of the matrix sentence are attributed with the reference of the pronouns asmad(I) and yuṣmad(you) respectively.

2. If there is **more than one matrix sentence**, the *kartā*-s and *karma*-s of all the hierarchical sentences are eligible to be referents of the pronouns asmad(I) and yuṣmad (you) respectively. In such cases, priority is determined hierarchically according to the depth of embedding. If an matrix sentence does not satisfy the case expectations, such as kartā for asmad and karma for yuṣmad, then the path must move to the higher-level matrix sentence.

3. If the matrix sentence contains a **verb with a meaning similar to *jānāti* ("to know")**, then

[17]These algorithms should be regarded as a preliminary attempt. While considerable effort and brainstorming were invested, time constraints limited the scope of exploration. Consequently, we present here only a few draft rules that capture the most explicit cases. We recognize, however, that the phenomenon is considerably richer and warrants further refinement and expansion in future work.

the second, i.e., next to immediate *kartā* in the hierarchy, is attributed with the reference of the pronouns asmad(I) and yuṣmad(you).

4. If there is only a single sentence and **no embedding structure**, check whether there is a proper name (identified through named entity recognition); such a name from the same sentence should be attributed as the referent of the pronoun asmad(I) yuṣmad (you).

5. If there is a single sentence without embedding, check for **any other clues** to identify the speaker and addressee (such as a character name in a drama or a speaker label in texts like the *Bhagavadgītā*) for attributing the reference of the pronouns asmad(I) and yuṣmad(you).

6. In case of the pronoun yuṣmad(you), the referent is commonly marked by the **vocative case**, within the same sentence or matrix sentences.

3.10.2 Referring-Coreferring Pronouns

The resolution of referring-coreferring (Ref-Coref) pronouns plays an important role in Sanskrit discourse interpretation, especially in inter-sentential contexts. Such pronouns typically involve the relative pronoun *yad-* in one sentence (S1) and a demonstrative pronoun such as *tad-* (or other members of the demonstrative set like *idam, adas, etad*) in the following sentence (S2). The relation between the two forms establishes a *nitya-sambandhaḥ* (sustained referential relation) that contributes to textual coherence.

The algorithm formalized here aims to automatically detect and establish such relations by relying on **grammatical agreement** and **dependency parsing**.

Objective: To identify coreferential relations between inflections of the *relative pronoun base yad-* in S1 and those of *tad-* (or other demonstratives) in S2, by exploiting agreement in gender, number, and case.

Algorithm for Ref-Coref Pronoun Resolution:

Step 1 : **Identification of Candidates**

(a) Search for any inflection of *yad-* in Sentence 1 (S1).

(b) Search for any inflection of *tad-* in Sentence 2 (S2).

(c) Extract **Gender (G)**, **Number (N)**, and **Case (C)** information for both instances.

Step 2 : **Co-reference Resolution (Priority-Based)**

(a) **Priority 1 (P1):** If *yad-* and *tad-* match in G, N, and C:
then, Connect them with relation *nitya-sambandhaḥ* using a bi-directional dotted link.

(b) **Priority 2 (P2):** If P1 fails, but *yad-* and *tad-* match in G and N (but not in C):
then, Connect with relation *nitya-sambandhaḥ* using a bi-directional dotted link.

Step 3 : **Special Cases and Exceptions**

- If neither P1 nor P2 applies, check for partial matches such as:
  (a) Gender and Case match (but Number differs), or
  (b) The coreferent in S2 is not *tad-*, but another demonstrative (*idam, adas, etad*).
- In such cases:
  (a) Attempt a partial match guided by syntactic plausibility and discourse coherence.
  (b) If plausible, connect with *nitya-sambandhaḥ*, but flag for manual review.
- In highly exceptional cases (no grammatical match, unresolved ambiguity):
  (a) Do not assign automatic coreference.
  (b) Mark as “unresolved” or “ambiguous” for manual inspection.

Empirical Basis

- This rule set successfully accounts for over 90% of *yad-tad* examples in the *Bhagavad Gītā*.

- Around 8% of cases involve mismatches in number or demonstratives other than *tad-*.
- Fewer than 2% remain unresolved and require contextual interpretation.

Significance: The Ref-Coref algorithm illustrates how Sanskrit's *relative-correlative (yad-tad)* construction, a hallmark of classical prose and poetry, can be modeled computationally. By prioritizing grammatical agreement and permitting controlled relaxation of constraints, the algorithm balances **precision** (through strict agreement rules) with **coverage** (through partial matching).

Thus, the algorithm not only formalizes a traditional syntactic phenomenon but also provides a foundation for broader discourse analysis, where relative-correlative pronouns function as key anchors of coherence across sentences.

In sum, the algorithms on PPN and RC demonstrate complementary approaches to pronominal anaphora resolution: verb-driven identification of participants (PPN), and structural chaining of referents across sentences (RC), together forming a layered model of coherence in Sanskrit texts.

3.11 Summary

This chapter has examined the problem of **pronominal anaphora resolution** in Sanskrit by combining traditional grammatical insights with preliminary algorithmic modeling. It began with a discussion of the nature of reference in pronouns, their polysemous potential, and the major classes of referents: personal, relative-correlative, and demonstrative pronouns. Directions of reference, including anaphora and cataphora, were also considered, together with the conceptual framework of *padaikavākyatā*. Corpus statistics were presented to highlight the distribution and positioning of pronominal entries, thereby motivating the design of computational procedures.

On this basis, algorithms were proposed for **Personal Pronouns (PPN)**, drawing on verb semantics and syntactic roles and for **Referring-Coreferring Pronouns (RC)**, formalizing the *yad-tad*

construction through agreement priorities. While the algorithms remain initial in scope, their implementation demonstrates the possibility of systematically modeling pronominal resolution in Sanskrit, providing a basis for further refinement and for exploring larger questions of discourse coherence.

CHAPTER 4

Discourse Markers

4.1 Introduction

The initial efforts in the field of computational linguistics were primarily focused on linking two sentences. However, discourse analysis extends beyond the scope of sentence boundaries, considering the text as a whole. Chapter 2 discusses in detail the definition of a sentence and the necessity of integrating sentences to achieve coherent comprehension. One method of connecting such dependent and fragmented sentences is through lexical markers, typically in the form of indeclinables or particles (*Avyaya*), such as “if-then ”, “because ”, “so-that ”, “but ”, “or ”, etc. While these sentence fragments are processed naturally by the human brain, a computer requires deliberate, structured efforts to handle them.

In Western linguistic theories, such sentences and constructions are categorized into two types: 1) Complex (also known as subordinates), which follow a process of integration, and 2) Compound (also known as coordinates), which simply involve joining. However, adhering to the Indian theoretical perspective, we distinguish only between an individual sentence (an independent one) and a group, where all dependent sentences form a single macro-sentence.[1] A group encompasses all coordinate constructions and certain subordinate constructions, particularly where more than one finite verb is employed.

This chapter aims to examine such Discourse Markers: how to identify them, how to join them using the principle of *Ekavākyatā*, their meanings, and how to represent them within the existing

[1] See chapter 2 for a more detailed explanation.

dependency framework.

4.2 Literature Review

Extensive work has been carried out in Western academia, leading to the creation of well-established frameworks and tools for analyzing discourse relations, applicable to both Western and Indian languages.

4.2.1 Western Theories

Discourse analysis, central to linguistics and NLP, examines how language structures extend beyond single sentences to ensure coherence and comprehension. This section outlines major theoretical insights and recent discourse, and anaphora resolution progress.

Discourse Analysis and Cohesion: One of the earliest and most influential contributions to discourse analysis was made by Halliday (1976) in their seminal work on cohesion. Their framework emphasizes the importance of cohesion devices such as conjunctions, references, substitution, and ellipses in creating links between sentences within a text. According to Halliday and Hasan, cohesion ensures that a text is not a random collection of sentences but a unified entity, with different linguistic elements contributing to its overall meaning. Their work laid the foundation for later studies in discourse coherence, highlighting how cohesive ties help readers understand the flow of ideas across sentences.

The development of Rhetorical Structure Theory (RST) further advanced discourse analysis by focusing on the hierarchical relationships between text segments. RST, proposed by Mann and Thompson (1988), conceptualizes discourse as a series of interconnected units, each linked by a discourse relation. These relations, such as causality, contrast, and elaboration, help shape the interpretive structure of a text. RST allows for a graphical representation of discourse structure,

providing a framework for analyzing how discourse relations function across different levels of text. The RST framework has been instrumental in understanding how text units interact to create a coherent narrative or argument.

Discourse Treebanks: Developing discourse treebanks has been a major advancement in computational discourse analysis. The Penn Discourse Treebank (PDTB) by Prasad et al. (2006) (Version 1), Prasad et al. (2008) (Version 2), Prasad et al. (2017) (Version 3) have played a key role in modeling discourse relations. By focusing on the structure of arguments and the role of connectives in establishing discourse relations, the PDTB provides an extensive resource for studying how connectives link ideas within a text. The treebank is based on large corpora, such as the Wall Street Journal, and analyzes discourse relations at a granular level. Similarly, other discourse databases, such as the Linguistic Discourse Model (Polanyi, 2008) and Discourse GraphBank (Wolf and Gibson, 2005), have contributed to the development of robust computational models for discourse processing.

In addition to treebanks, computational models such as the Discourse-Lexicalized Tree Adjoining Grammar (DLTAG) (Webber and Joshi, 1998) have been used to capture discourse structures. DLTAG integrates lexical information with syntactic structures, providing a deeper understanding of how lexical choices impact discourse relations. These models have advanced our understanding of the complexities of discourse and have been crucial for developing automated systems that process and analyze texts.

Research on non-Indian languages such as Chinese and Turkish has laid foundational groundwork for discourse analysis through resources like the Jiang et al. (2018) based on the RST framework and Zeyrek et al. (2010), leveraging PDTB to model discourse relations. These efforts have provided methodologies and tools that have significantly influenced discourse parsing in Indian languages. A single illustrative example has been presented for each module to convey the general idea; however, extensive research exists across a wide range of major languages. The following section will

concentrate specifically on Indian languages, which form the central focus of this study.

The novel effort of integrating Western and Indian theories was executed by Subalalitha and Parthasarathi (2012) in an innovative approach to discourse parsing by combining traditional Indian linguistic concepts such as *Saṃgati* with Rhetorical Structure Theory (RST). The authors propose a hybrid framework where the identification of discourse relations draws on both indigenous notions of textual coherence and formal structures from RST. Their study demonstrates how *Saṃgati* markers, which denote semantic connectivity between units, can be systematically mapped onto RST relations for parsing purposes. This work is significant as it bridges classical Sanskrit theories of discourse with contemporary computational models, particularly in the context of Indian languages. It offers a culturally grounded yet technically rigorous methodology for discourse parsing.

4.2.2 Existing modules for Indian Languages

Discourse relation analysis in Indian languages is challenged by complex syntax, rich inflectional morphology, and word order variability across language families. Initial efforts to address these issues primarily involved the development of linguistic resources and computational models for anaphora resolution, which were largely grounded in Western theoretical frameworks, particularly those based on the Penn Discourse Treebank (PDTB)and Rhetorical Structure Theory (RST).

Hindi: Oza et al. (2009) developed the Hindi Discourse Relation Bank (HDRB), an extensive resource that adapts the Penn Discourse TreeBank (PDTB) framework for annotating discourse relations in Hindi. This work provides critical infrastructure for advancing computational models in tasks such as machine translation, summarization, and question answering, particularly in low-resource settings. Building on the need for systematic discourse representation, Sweta Khandelwal (2013) proposed a discourse tagging scheme for Hindi that integrates Discourse Markers and structural indicators, facilitating improved syntactic-semantic alignment within text. Complementing these efforts, Jain et al. (2016) focused on the task of explicit argument identification in Hindi

discourse parsing, introducing methods to automatically identify and align discourse arguments, which are essential for downstream applications in discourse-aware NLP systems.

Bengali: Corpus Creation and Annotation for Discourse Relations in Bangla (Kar et al., 2013) describes the development of a shallow discourse-annotated corpus for Bangla, following the Penn Discourse Treebank (PDTB) framework. The annotation primarily targets explicit discourse connectives and the corresponding argument spans, without establishing a full hierarchical discourse structure. In contrast, the Bangla RST Discourse Treebank (Das and Stede, 2018) adopts the Rhetorical Structure Theory (RST) approach to construct a hierarchical discourse treebank for Bangla, wherein texts are segmented into Elementary Discourse Units (EDUs), and rhetorical relations are systematically annotated between units. Collectively, these two resources provide complementary perspectives on discourse analysis in Bangla - one emphasizing shallow connective-based annotation and the other offering a deeper structural representation of discourse.

Telugu: Efforts toward developing syntactic resources for Telugu have laid the foundational work for future discourse-level analysis. Vijayalakshmi and Sharma (2007) presents one of the earliest computational approaches to Telugu discourse analysis by segmenting text into discourse units and identifying Discourse Markers as cues for inter-sentential relations. Nallani et al. (2020) expanded the existing Paninian-style dependency treebank by introducing intra-chunk dependencies, creating a fully expanded dependency treebank comprising 3,220 manually annotated sentences. Their work not only improved syntactic coverage but also updated POS tags to the Bureau of Indian Standards (BIS) format, enhancing downstream NLP applicability. Earlier, Vempaty et al. (2010) addressed challenges in building a Telugu treebank using Paninian grammar, focusing on sentence segmentation and inter-chunk relations. Their corpus of 1,487 annotated sentences demonstrated how structural complexity in Telugu demands specialized annotation frameworks. Although these resources focus on syntax rather than discourse, they provide essential groundwork for future development of Telugu discourse relation treebanks.

Tamil: Rachakonda and Sharma (2011) undertook the development of a discourse tagging scheme specifically for Tamil, recognizing the language's unique syntactic and discourse characteristics. Their work proposed an annotation framework that captures discourse relations within Tamil texts, taking into account the agglutinative nature of the language and the role of Discourse Markers. This contribution is significant as it laid early groundwork for systematic discourse parsing efforts in Tamil, addressing challenges specific to South Indian languages.

Malayalam: Kumari and Devi (2016) proposed a discourse annotation scheme for Malayalam, focusing on the systematic identification and annotation of discourse connectives in a language characterized by rich morphology and complex syntactic constructions. Their work highlights the challenges posed by Malayalam's free word order and intricate clause structures, offering methodological insights for discourse analysis in other morphologically rich Indian languages.

Surveys and Multilingual efforts: Rajan et al. (2015) presents a comprehensive survey of discourse tagging efforts in Indian languages, critically evaluating existing annotation schemes and resources, while highlighting the necessity of language-specific adaptations due to the syntactic and semantic diversity across Indian languages. Building on these foundations, Lalitha Devi et al. (2014) reports empirical efforts within the Indian Language Discourse Project, proposing multilingual discourse tagging schemes for Hindi, Tamil, and Malayalam that address challenges such as explicit and implicit discourse connectives. In a foundational contribution, Bharati et al. (2013) discusses the development of standardized discourse annotation frameworks tailored for Indian languages, emphasizing the complexities of their syntactic and semantic structures and advocating for a unified approach to resource creation. Further advancing this field, Paul and Bhattacharyya (2019) provides a detailed overview of discourse relation annotation frameworks specifically designed for Indian languages, covering theoretical underpinnings, annotation guidelines, and computational applications. Collectively, these works underscore the challenges involved and the importance of developing adaptable, multilingual discourse resources for diverse Indian languages.

4.2.3 In Sanskrit

Dr. Monali Das has played a pivotal role in advancing computational methodologies for the analysis of Sanskrit texts, particularly within the scope of discourse analysis. Her doctoral thesis (Das, 2016), titled Discourse Analysis of Sanskrit Texts: First Attempt towards Computational Processing, investigates the inherent complexities of Sanskrit literature, emphasizing its oral heritage and intricate textual architecture. The primary objective of her research is to develop computational frameworks capable of systematically interpreting these features, thereby integrating classical Sanskrit studies with modern computational paradigms.

The study presented in Kulkarni and Das (2012) extends the discourse analysis framework to Sanskrit by evaluating its theoretical and practical applicability. It introduces techniques for automated tagging and syntactic parsing at the discourse level and juxtaposes Indian traditional discourse frameworks with Western computational models. This work serves as a foundational step toward enabling comprehensive computational discourse processing for Sanskrit.

Basis for this work: The tag-set outlined in Kulkarni and Ramakrishnamacharyulu (2016) includes relations across sentence boundaries, particularly those signaled by specific indeclinable connectives. Some of these connectives function in pairs and are instrumental in articulating inter-sentential relationships. A tagging methodology tailored to such connectives was earlier proposed by Kulkarni and Das (2012). Each connective operates over two textual arguments, identified as *anuyogika*[2] and *pratiyogī*, adopting a terminology consistent with traditional logical analysis. When a connective C links two clauses or sentences, denoted S1 and S2, their general configuration is illustrated in Figure 4.1.

When a pair of connectives, C1 and C2, is used to associate S1 and S2, the resultant structure is depicted in Figure 4.2.

[2]S2 is the anuyogi. Therefore, if the arrowhead is pointing toward S2, the relation is named *anuyogi*. In this diagram, since the arrowhead is directed toward C, the relation is the inverse of *anuyogi*, i.e., *anuyogika*, or "combining."

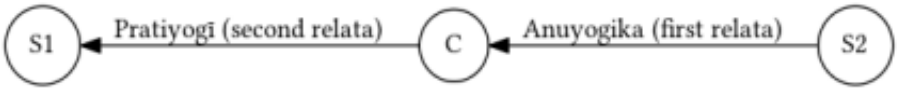

Figure 4.1: Discourse structure with single connective

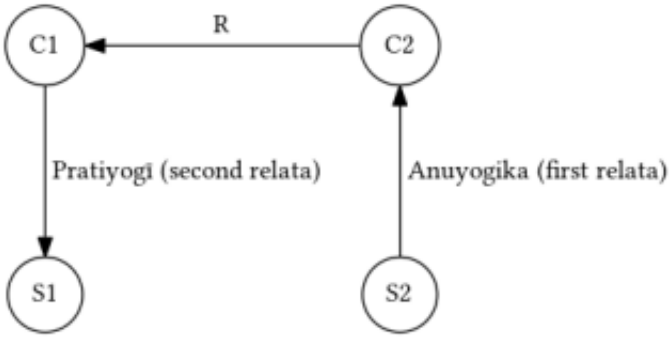

Figure 4.2: Discourse structure with paired connectives

In this paired configuration, the relation R serves to bind the two connectives, C1 and C2. These connectives relate to the respective sentences primarily through their governing verbs, with the sentences themselves represented as dependency structures. In practical application, the usage of paired connectives allows for the presence of either one or both within a sentence. If only a single connective is employed, the structure corresponds to that of Figure 4.1, effectively reducing the complexity shown in Figure 4.2.

Nevertheless, two main limitations can be identified in this approach. Firstly, the designations *anuyogika* and *pratiyogī* are rather broad and do not adequately encapsulate the nuanced semantic relations that may exist between two connected clauses. Secondly, the study does not consider several other indeclinable elements that also serve as Inter-Sentential Markers. Consequently, the following section aims to refine the classification of such sentence relations, focusing on cases marked by explicit linguistic indicators. Furthermore, it seeks to provide a comprehensive inventory of all indeclinables that function as relational markers within Sanskrit discourse.

Relation	Markers
Succeeding (anantarakālaḥ)	tataḥ, atha, anantaram
Simultaneity (samānakālaḥ)	yadā-tadā
Co-location (samānādhikaraṇaḥ)	yatra-tatra
Similarity (sādṛśya)	yathā-tathā
Cause-Effect (kāraṇa-kārya)	yataḥ-tataḥ, ataḥ, yasmāt-tasṃāt, yena-tena, hi
Possiblity (āvaśyakatā-pariṇāma)	yadi-tarhi, cet
Hindrance in cause-effect (vyabhicāra)	yadyapi-tathapi, cedapi, athāpi, tarhyapi, sannapi
Antithesis (virodhaḥ)	parantu, kintu
Conjunction (samucccayaḥ)	ca, api, tataḥ, cāpi, athaca, athāpi, evañca
Disjunction (anyataraḥ)	vā, uta, yadvā, athavā, utāpi, utasvit

Table 4.1: List of discourse relations.

4.3 Lexical Markers

Below, we present a curated list of Discourse Markers along with the relation(s) they express. This comprehensive list has been compiled from various authentic sources, including the *Avyayakośa* by Shukla (2008), *Vyākaraṇacandrikā* by Pandeya (2006), and Speyer (1886). After categorizing the markers, we further refined the meanings associated with each through a detailed analysis. The foundational reference for the creation of this list has been developed by Das (2016) in her thesis, which we have expanded and adapted for the present study. (See 4.1)

Paired vs. Single Markers: The list 4.1 reveals two primary types of Discourse Markers: single and paired. They vary in numbers, expectancy, and representation.[3]

1. Single Markers: Single markers such as *tataḥ*, *anantaram*, *atha* (anantarakālaḥ), *ataḥ*, *hi* (kāraṇa-kārya), *iti*, *cet* (āvaśyakatā-pariṇāma), *parantu*, *kintu* (virodhaḥ), *ca*, *api* (samuccayaḥ), *vā*, and *uta* (anyataraḥ) occur primarily in the pratiyogi sentence, that is, the consequent sentence. As their name suggests, these markers independently signal a discourse relation without requiring a corresponding marker. However, they presuppose the presence of another sentence (the anuyogi or antecedent sentence) to fulfill the syntactic and semantic expectations

[3]See section 4.4 of this chapter.

raised in the embedding sentence.

2. Paired Markers: Paired markers, on the other hand, consist of two markers that are interdependent and conventionally occur together across two sentences. Examples include *yadā-tadā* (samānakālaḥ), *yatra-tatra* (samānādhikaraṇaḥ), *yathā-tathā* (sādṛśya), *yataḥ-tataḥ*, *yasmāt-tasmāt*, *yena-tena* (kāraṇa-kārya), *yadi-tarhi* (āvaśyakatā-pariṇāma), and *yadyapi-tathāpi* (vyabhicāra). In each case, the marker derived from the pronoun *yad* (i.e., typically beginning with 'ya') is associated with the anuyogi (antecedent sentence), while the marker derived from *tad* (beginning with 'ta') is associated with the pratiyogi (consequent sentence). Although these markers conventionally expect their counterpart to be explicitly present, it is not unusual for one of them to be implied rather than overtly stated.

Local Word Grouping (LWG): A third category identified in the list consists of grouped markers, such as *cedapi*, *athāpi*, *tarhyapi* (vyabhicāra),[4] *cāpi*, *athaca*, *athāpi*, *evañca* (samuccayaḥ), and *yadvā*, *athavā*, *utāpi*, *utasvit* (anyataraḥ). These constructions are amalgamations of multiple indeclinables where one component typically dominates in determining the semantic interpretation. Grouped markers are treated separately from single and paired markers for two primary reasons: (1) there exists considerable variation in the formation of such groups based on the speaker's or writer's stylistic choices, making it infeasible to list all possible combinations exhaustively.[5] (2) In textual material, these groupings may appear either as a single concatenated unit or as separately written words, depending on the scribe's or annotator's choice. Consequently, such grouped markers are treated as single semantic units during analysis and representation.

[4]Most of the vyabhicāra-denoting markers are derived by attaching *api* to āvaśyakatā-pariṇāma markers. Another method of marking vyabhicāra is via the use of an infinite verb with a *-satṛ* suffix combined with *api*. However, *yadyapi-tathāpi* is treated differently due to its expectancy relation across sentences.

[5]The list presented here includes the most prominent and widely used groupings.

4.4 Role of Ekavākyatā

Discourse Markers can be analyzed through the principle of *ekavākyatā*. The concept of *ekavākyatā* aids in the syntactic and semantic combination of otherwise independent sentences, facilitating the comprehension of a joint macro-sentence (*mahāvākya*).

Within the domain of *ekavākyatā*, the sub-type *vākyaikavākyatā* effectively accounts for the phenomenon of combining sentences that exhibit Discourse Markers. Here, two *avāntaravākya*-s (independent sentences) are conjoined to form a larger or macro unit (*mahāvākya*), in which each constituent sentence retains its original structure and identity. The connection is established through lexical markers that signal a semantic relationship, though the core syntactic integrity of the sentences remains unaltered.

The semantic relation in such constructions is typically mediated by the finite verbs, which are treated as the syntactic heads in the dependency structure. This is consistent with the Pāṇinian Dependency Framework, into which the discourse relations are being integrated. The lexical markers themselves function merely as triggers or indicators of the relation and do not bear any independent semantic contribution to the relation or the structural integration of the sentences.

The *vākyaikavākyatā* relation between two sentences marked by Discourse markers is governed by the *śeṣa–śeṣi-bhāva* (beneficiary relation) where each element contribute for the completion (in this case, meaning comprehension) of the other element. This governing principle does not imply syntactic subordination alone; rather, it also accommodates coordination within a unified framework and allows for a common representational scheme.

For Example:

Sanskrit S1: *yadā odanaṃ pacati*.

S2: *tadā bhojanaṃ kuru*.

Gloss S1: when rice [n, nom, sg] cook [present, 3p, sg].

S2: then (you) meal [n, acc, sg] do [imp, 3p, sg].

English S1: When rice cooks.

S2: Then you eat a meal.

The sentence pair above features the paired Discourse Markers *yadā–tadā*, which signal a Discourse Markers of type *samānakālaḥ* (simultaneity). This relation is established specifically between the two verbs *pacati* and *kuru*, as the corresponding actions are intended to occur concurrently. Although the sentences remain structurally independent, their semantic integration requires the application of the principle of *ekavākyatā* of type *vākyaikavākyatā*. The two sentences complement each other in comprehension of the whole meaning and thereby fulfilling the *śeṣa–śeṣi-bhāva*, making them eligible for unified interpretation, consequently forming a larger composite sentence - *Maāvākya*:

"yadā odanaṃ pacati, tadā bhojanaṃ kuru."

Such a newly formed *mahāvākya* (macro-sentence) is complete in itself as it conveys a unified meaning (*arthāikatva*) and satisfies all syntactic and semantic expectations (*nirākāṅkṣatva*).

Thus, Discourse Markers can be effectively interpreted, combined, and represented under the governance of the principle of *vākyaikavākyatā*.

4.5 Representation

This section outlines a representation schema for ten distinct discourse relations, modeled within the framework of the Paninian dependency grammar. For each relation, we analyze its corresponding lexical markers and offer semantic interpretations that underline their functional roles in sentence linkage.

1. **Succeeding (anantarakālaḥ) :**

   This relation captures the temporal sequencing of events, where the action described in the current sentence follows that of the previous one. Indeclinable markers such as ‘atha’ and ‘tataḥ’ typically signal this transition. The verb in the succeeding sentence is temporally posterior to that of the preceding sentence. Refer to Figure 4.3.

   Sanskrit: ahaṃ śṛṇomi atha likhāmi.
   Gloss : I[nom,sg] listen[present,1p,sg] atha[indcl] write[present,1p,sg]
   English: I will listen, then I will write.

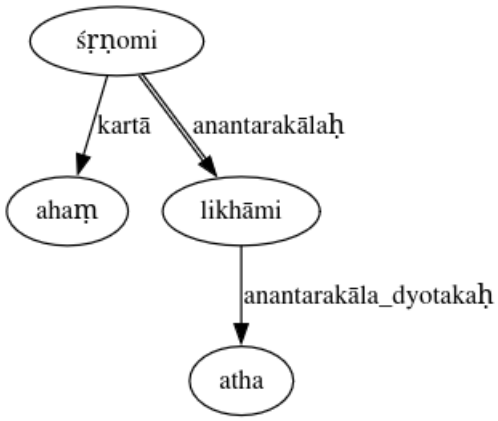

Figure 4.3: anantarakālaḥ relation with marker atha

Before examining the next type of relation, we highlight key characteristics of the discourse trees that illustrate inter-sentential dependencies.

- Discourse relations connect the ‘head’ nodes (mukhya viśeṣya) of the two sentences involved.
- The arrowhead on the link points to the node fulfilling the semantic condition indicated by the edge label.

- Unlike intra-sentential dependencies, arrow direction here does not indicate syntactic dependency.
- Discourse-level dependencies are visually distinguished by a double-lined edge.
- The verbs are assigned the same visual rank when they are semantically co-dependent or when no main–subordinate relation is present. (If, however, a directional or hierarchical dependency exists, the verbs are ranked differently, with the verb that governs the relation placed at the highest rank.)

2. **Simultaneity (samānakālaḥ) :**

   This type of relation expresses the simultaneity of events conveyed across two sentences. Sanskrit employs two primary mechanisms for this: (a) present participial constructions using śatṛ and śānac suffixes, and (b) indeclinable correlatives such as yadā-tadā. In both instances, the verbs in finite forms are semantically linked to reflect co-occurring actions. The correlatives yadā and tadā establish a temporal (kāla-adhikaraṇa) relation. Refer to Figure 4.4.

   Sanskrit : yadā bharataḥ mārge gacchati tadā saḥ devālayam paśyati.
   Gloss when[indcl] bharata[nom,sg,m] path[loc,sg,n] go[present,3p,sg] then[indcl] he[nom,sg,m] temple[m,acc,sg] see[present,3p,sg]
   English: When Bharata goes on the road, he sees a temple.

3. **Co-location (samānādhikaraṇaḥ) :**

   This relation indicates that the events described in two successive sentences occur at the same physical or spatial location. The indeclinable pair yatra-tatra signals such spatial alignment, though at times only one of them may be explicitly present. The relation corresponds to deśa-adhikaraṇa (locative of place). See Figure 4.5.

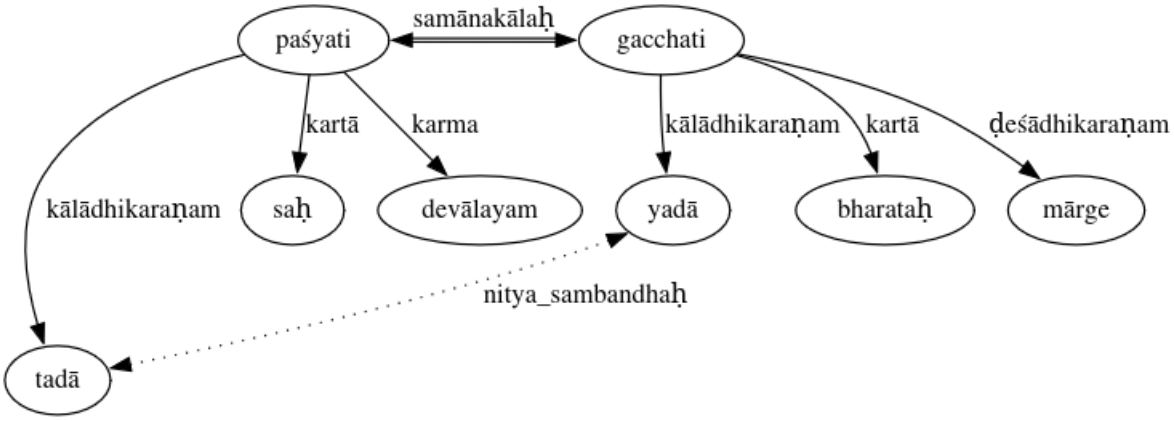

Figure 4.4: samānakālaḥ relation with markers yadā-tadā

Sanskrit: yatra nāryaḥ tu pūjyante tatra devatāḥ ramante.

Gloss : where[indcl] women[f,nom,pl] tu[emph] worship[present,3p,pl] there[indcl] God[m,nom,pl] reside[present,3p,pl]

English: Where women are worshiped, there reside the Gods.

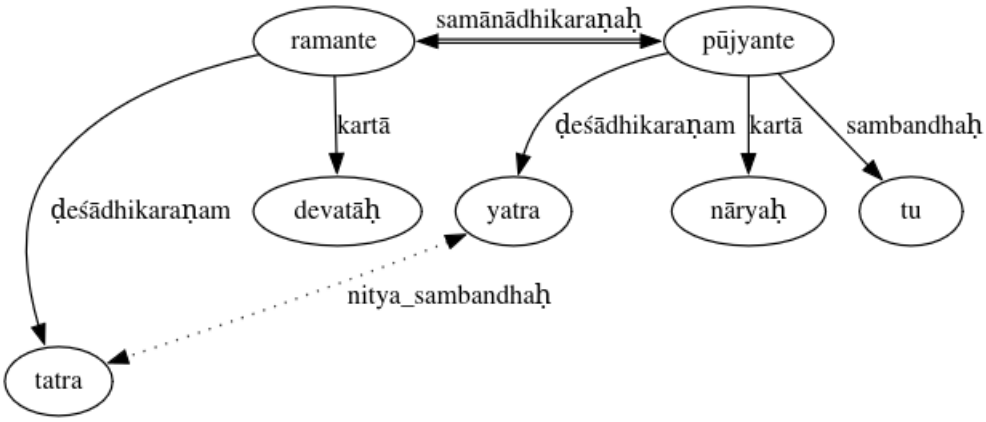

Figure 4.5: samānādhikaraṇaḥ relation with markers yatra-tatra

4. **Similarity (sādṛśya) :**

This discourse relation reflects similarity in action or process across two consecutive sentences. The indeclinables yathā and tathā often serve as markers indicating parallelism. See Figure 4.6.

Sanskrit : yathā janāḥ karmāṇi kurvanti. tathā te phalaṃ prāpnuvanti.

Gloss : yathā[indcl] janāḥ[m,nom,pl] karmāṇi[m,acc,pl] kurvanti[present,3p,pl] tathā[indcl] te[m,nom,pl] phalaṃ[n,acc,sg] prāpnuvanti[present,3p,pl]

English: As people perform actions, they obtain the result.

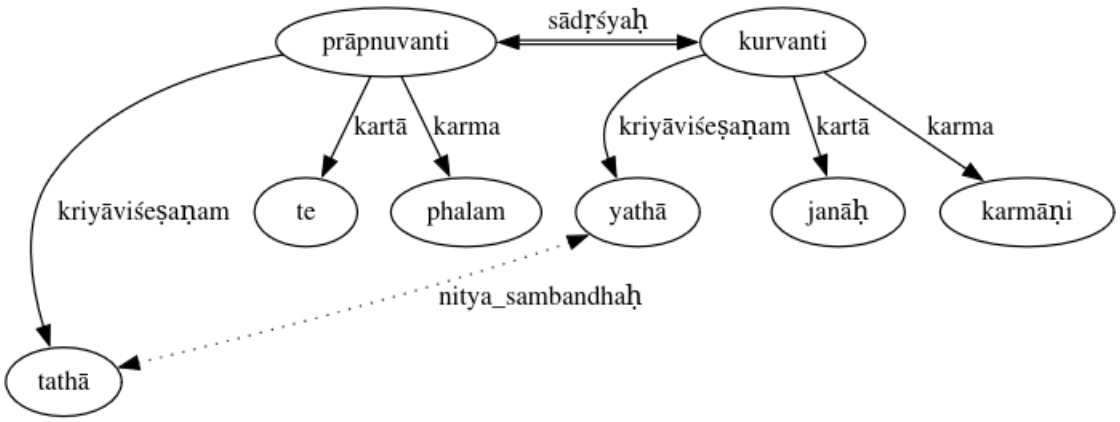

Figure 4.6: sādṛśya relation with markers yathā-tathā

5. **Cause-Effect (kāraṇa-kārya) :**

   This category describes deterministic causal relationships. When the cause is either certain or has already occurred, the result is anticipated with high confidence. Indeclinables like yataḥ (cause indicator) and tataḥ (effect indicator) are employed to express this causal linkage. Either marker may be used independently, though both together provide more clarity. See Figure 4.7.

   Sanskrit: yataḥ avarṣat tataḥ mayūraḥ nṛtyati.

   Gloss : because([indcl] rain[past,3p] therefore[indcl] peacock[m,nom,sg] dance[present,3p,sg]

   English: Because it has rained, the peacock is dancing.

6. **Possibility (āvaśyakatā-pariṇāma) :**

   This structure addresses conditional relations where the cause is not definite but possible. It indicates a potential outcome based on a hypothetical cause. The indeclinables yadi (conditional) and tarhi (consequent) signify this relationship. These may appear together or independently. See Figure 4.8.

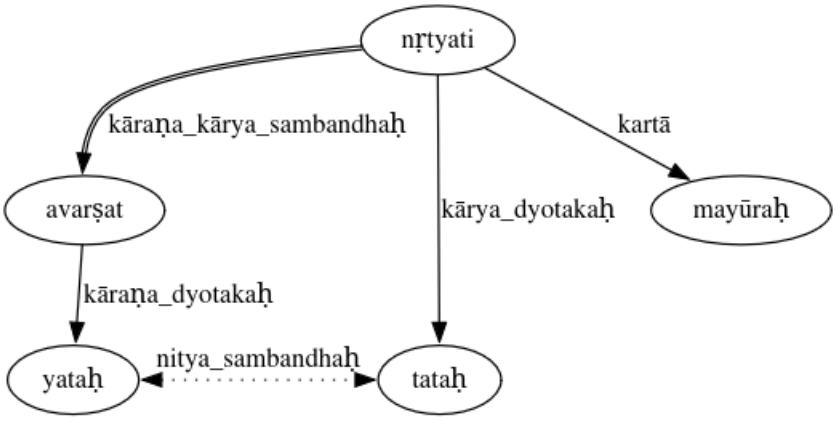

Figure 4.7: kāraṇa-kārya relation with markers yataḥ-tataḥ

Sanskrit: yadi paṭhasi tarhi uttīrṇaḥ bhaviṣyasi.

Gloss : if [indcl] read[present,2p,sg] then[indcl] pass[m,nom,sg] to be[fut,2p,sg]

English: If you study, you will pass.

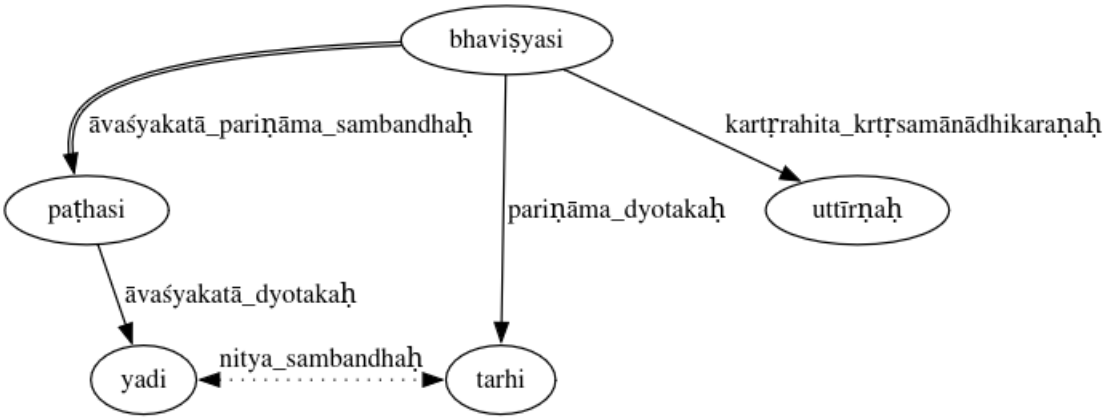

Figure 4.8: āvaśyakatā-pariṇāma relation with yadi-tarhi marker

7. **Anomaly (vyabhicāra) :**

   This relation expresses exceptions to typical cause-and-effect sequences. The presence of yadyapi (even though) and tathāpi (even then) introduces a breach in expected outcomes. The relation is annotated between the finite verbs denoting actions. This anomaly may occur in two ways: the presence of a cause without its expected effect, or the occurrence of an effect

without its conventional cause. See Figures 4.9 and 4.10 respectively.

Sanskrit : *yadyapi varṣā asti tathāpi mayūraḥ na nṛtyati*.

Gloss: even-if rain[f,nom,sg] be[present,3p,sg] even-then peacock[m,nom,sg] neg dance[present,3p,sg]

English: Even if it doesn't rain, even then the peacock dances.

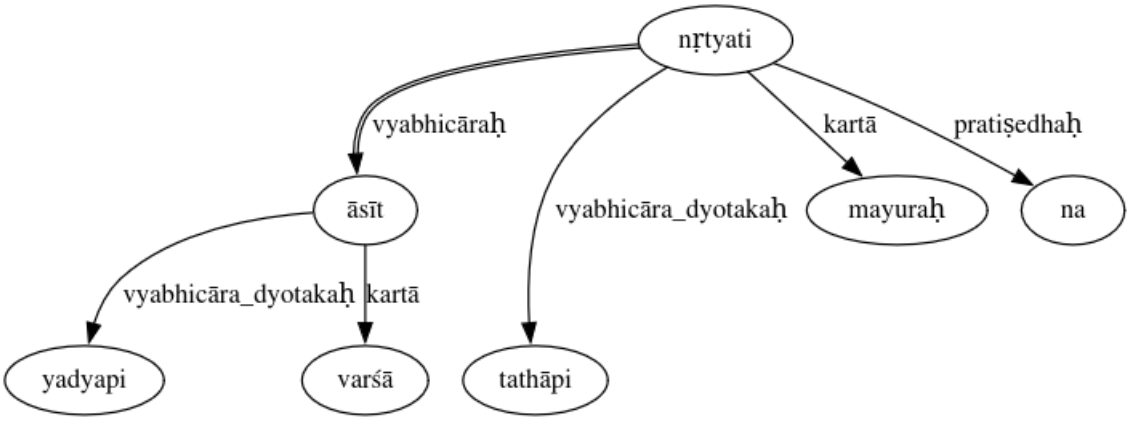

Figure 4.9: anvaya vyabhicāra relation with markers yadyapi-tathāpi

Sanskrit: *yadyapi saḥ vaidyaḥ na asti tathāpi saḥ cikītsāṃ jānāti*.

Gloss: Even-if he[m,nom,sg] doctor[m,nom,sg] neg be[present,3p,sg] even-then he[m,nom,sg] cure[f,acc,sg] know[present,3p,sg]

Eng: Even if he is not the doctor, even theneven-then he knows the cure.

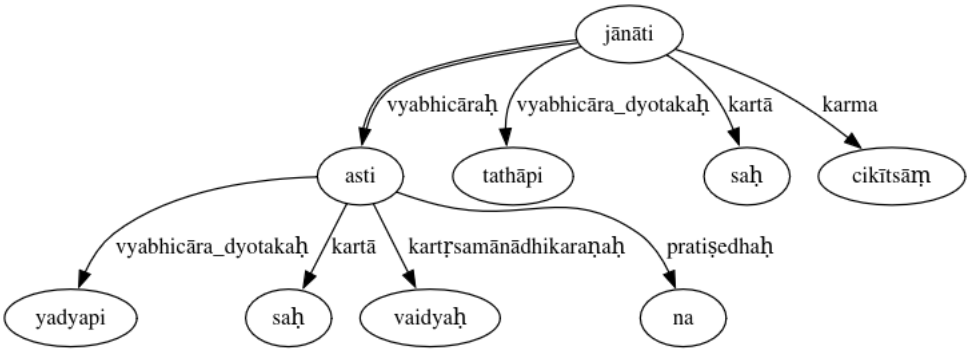

Figure 4.10: vyatireka vyabhicāra relation with markers yadyapi-tathāpi

8. **Antithesis (virodhaḥ) :**

This discourse relation captures semantic opposition or contradiction between two statements. It is typically marked by indeclinables such as *parantu* and *kintu*. These conjunctions indicate a contrastive or adversative relationship, where the second sentence undermines or opposes the proposition of the first. Refer to Figure 4.11.

Sanskrit: gajendraḥ tīvram prayatnam akarot parantu nakra-grahāt na muktaḥ.

Gloss : Elephant[m,nom,sg] hard[adj] effort[m,acc,sg] do[3,sg,past] but[part] crocodile-grip[abl,m,sg] no[part] free[ppp,3p,sg]

English: The Elephant tried hard but couldn't escape from the crocodile's grip.

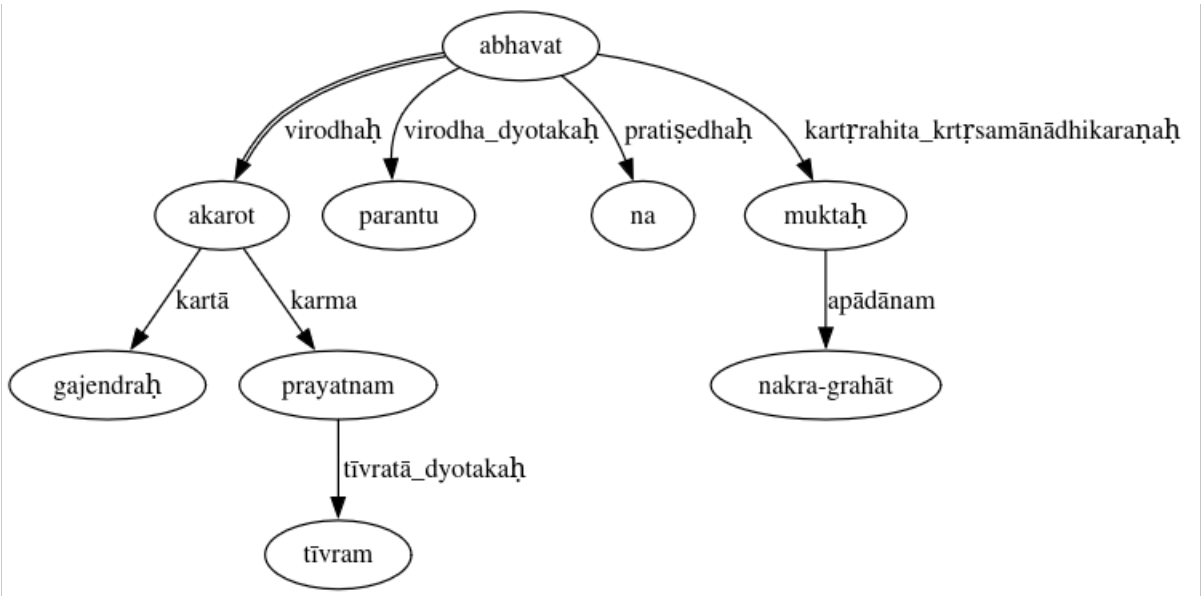

Figure 4.11: virodhaḥ relation with parantu marker

9. **Conjunction (samuccayaḥ) :**

   This relation denotes an additive or cumulative connection between two or more actions or propositions. Indeclinables like *api-ca* function as conjunction markers, indicating that multiple commands or descriptions are grouped without hierarchical distinction. See Figure 4.12.

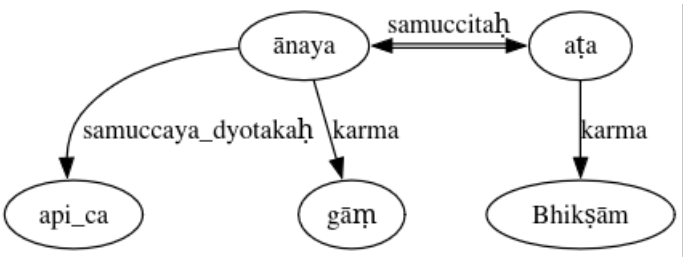

Figure 4.12: samuccayaḥ relation with api-ca marker

Sanskrit : Bhikṣām aṭa api-ca gāṃ ānaya.

Gloss : alms[m,acc,sg] roam[ipm,2p,sg] also[part] and[part] cow[f,acc,sg] bring[imp,2p,sg]

English: Roam around for alms and bring the cow.

10. **Disjunction (anyataraḥ) :**

This final relation type models alternatives or options presented in discourse. The indeclinable *athavā* is used to indicate a disjunctive connection, signifying that one of the possible actions may occur, but not necessarily both. This form captures optionality and choice in actions or states. See Figure 4.13.

Sanskrit : sītā śvaḥ kāryakrame gāsyati athavā nartsyati.

Gloss : Sita[f,nom,sg] tomorrow[part] program[m,loc,sg] sing[fut,3p,sg] or[part] dance[fut,3p,sg]

English: Sita will sing in tomorrow's program or dance.

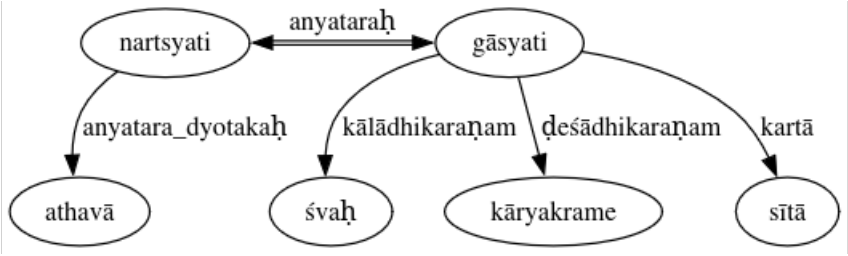

Figure 4.13: anyataraḥ relation with athavā marker

4.6 Arguments Structure and Positioning

The Elementary Discourse Unit (EDU) for Discourse Markers is defined at the level of the individual sentence, where the qualifying definition of a sentence is derived from the traditional rule: *ekatiṅ vākyam* - a sentence is a syntactic unit characterized by the presence of one and only one finite verb.

In such discourse constructions, one (in the case of single markers) or both (in the case of paired markers) of the sentences are embedded with lexical markers that explicitly signal discourse relations. Unlike frameworks such as the Penn Discourse Treebank (PDTB), which rely on argument structure and argument positioning for relation interpretation, the present framework does not assume such structural dependencies and argument positioning; rather, it is the marker that guides the relation. This divergence stems from the flexible syntactic nature of Sanskrit, which allows for considerable word and clause order variation. Consequently, Discourse Markers constructions exhibit this syntactic flexibility as well.

Freedom of argument order across sentences: The ordering of arguments is not fixed. In such cases, largely, the semantic interpretation remains intact despite positional variation. This flexibility, however, varies depending on the specific relation involved and is governed more by the semantics than by a strict syntactic rule. For instance:

> Sanskrit: yadi bhavān icchati, ahaṃ tava gṛhe āgacchāmi.
>
> Gloss: if[indcl] you[m,nom,sg] wish[present,3p,sg], I[nom,sg] your[gen,sg] house[m,loc,sg] come[present,1p,sg]
>
> English: If you wish, I will come to your house.

versus

> Sanskrit: ahaṃ tava gṛhe āgacchāmi, yadi bhavān icchati.
>
> Gloss: I[nom,sg] your[m,gen,sg] house[n,loc,sg] come[present,1p,sg], if[indcl] you[nom,sg]

wish[present,3p,sg]

English: I will come to your house if you wish.

Similarly, in the following pair, it is freer than the previous example:

Sanskrit: yathā tvaṃ karoṣi, tathā ahaṃ karomi.

Gloss: as[indcl] you[m,nom,sg] do[present,2p,sg], so[indcl] I[nom,sg] do[present,1p,sg]

English: As you act, so do I.

versus

Sanskrit: yathā ahaṃ karomi, tathā tvaṃ karoṣi.

Gloss: as[indcl] I[nom,sg] do[present,1p,sg], so[indcl] you[m,nom,sg] do[present,2p,sg]

English: As I act, so do you.

In natural language, the sentence containing a marker can appear in flexible positions (as seen in examples above), and our framework ideally accommodates this variability. From a computational perspective, this variability can be formalized using a set of extracted contextual features for each marker instance, such as its surrounding clause type, presence of negation, and nearby verbal forms. These features serve as predicates in a rule-based system to determine the appropriate discourse relation automatically, allowing the system to handle positional flexibility while maintaining accuracy. However, in any order, or as per their natural inclination, these markers are predictably fit in the following convention.

1. **Occurs in prior-Sentence(S1):** yadā, yatra, yathā, yataḥ, yasmāt, yena, yadi, cet, yadyapi, cedapi, sannapi

2. **Occurs in subsequent-Sentence(S2):** anantaram, tadā, tatra, tathā, hi, tataḥ, ataḥ, tasmāt, tena, tarhi, tathāpi, athāpi, tarhyapi, parantu, kintu, ca, api, cāpi, athaca, evañca, vā, uta, yaxvā, athavā, utāpi, utasvit.

However, the occurrence of these markers may differ. That is one of the indicators for the possibility of non-discourse relations. In such cases, we look for disambiguation clues. This flexibility creates a challenge for automated processing, as the system cannot reliably determine whether the relevant context lies in the preceding or following sentence. To address this, we adopted a fixed positional convention for each marker and its associated sentence. These fixed positional conventions can be directly leveraged in an algorithmic framework: for each marker, the system can reference its expected sentence position (prior or subsequent) as part of the feature set when evaluating candidate discourse relations. Skipping the ambiguous markers, here is the refined list:

1. **Always in prior sentence:** yadi, cet

2. **Always in subsequent sentence:** anantaram, parantu, kintu, tarhi

3. **In both or subsequent:** ca, api, cāpi, athaca, evañca, vā, uta, yaxvā, athavā, utāpi, utasvit.

We emphasize that, within a clause (in the case of a non-finite verb) or a sentence (in the case of a finite verb), the movement of markers is generally unrestricted, similar to other elements within the sentence. Nevertheless, the position of a marker within a sentence may guide the annotator toward a preferred interpretation of its relational function, as will be discussed in Section 4.12 of this chapter.

4.7 Hierarchy in relations

We observed that there are multiple relations possible in the case of non-congruent markers occurring in a set of sentences. To solve this problem of over-analysis, we discover that a few relations are more discourse-oriented than others. So we decided on the hierarchy among the available relations. This hierarchy is neither semantic nor structural, but merely functional (unlike PDTB, where the hierarchy decides the granularity of the meaning of a relation). The hierarchy is as follows:

1. **Layer 1 relations:** *kārya-kāraṇa, āvaśyakatā-pariṇāma, vyabhicāra* and *virodha*

2. **Layer 2 relations:** *anantarakāla, samānakāla, samānādhikaraṇa, sādṛśya*

3. **Layer 3 relations:** *samuccaya* and *anyatara*

In a rule-based algorithm, this hierarchy can be used to resolve conflicts when multiple discourse relations are simultaneously applicable. Higher-layer relations take precedence in annotation, ensuring that the most discourse-oriented interpretation is selected automatically.

4.8 Ambiguity

Disambiguation plays a crucial role in discourse annotation, especially when constructions exhibit dual interpretability as part of either pronoun resolution (PR) or Discourse Markers. The section also discusses karaka, that is, sentence-level and discourse-level relations. It explores if there is any clue for disambiguation as well. See table 4.2 for the ambiguity overview.

In a computational system, disambiguation is performed by evaluating the extracted contextual features against the conditions specified in the rule set for each marker. When multiple relations are possible, a priority scheme based on the functional hierarchy and rule specificity is applied to select the most appropriate discourse relation for annotation.

Such cases illustrate how a rule-based system can leverage contextual information - such as shared person-number-gender-case markers, clause boundaries, and co-occurrence of paired markers - to determine whether a relation is intra-sentential, inter-sentential, or semantically ambiguous.

1. **Ambiguity At structural level:** In such cases, the semantically same relation occurs but at a different level of analysis. One at the intra-sentence, i.e., Kāraka level and another at the inter-sentence, i.e., discourse level. The markers that show such ambiguity are *ca, vā, uta* and all other local word grouping options. Consider the following examples:

   Sanskrit: rAmaḥ lakṣmaṇaḥ ca vanam gacchataḥ.

   Gloss: Ram[m,nom,sg] Lakshman[m,nom,sg] and[indcl] forest[m,acc,sg] go[present,3p,du]

Marker	Discourse Relations	Kāraka Relations
tasmāt	kāraṇa-kārya	apādānam
yasmāt	kāraṇa-kārya	apādānam
tataḥ	anantarakāla, kāraṇa-kārya	apādānam
yataḥ	kāraṇa-kārya	apādānam
ataḥ	kāraṇa-kārya	apādānam
tena	kāraṇa-kārya	karaṇa
yena	kāraṇa-kārya	karaṇa
hi	kāraṇa-kārya	sambandhaḥ
tadā	samānakālaḥ	kālādhikaraṇam
yadā	samānakālaḥ	kālādhikaraṇam
tatra	samānādhikaraṇam	deṣādhikaraṇam
yatra	samānādhikaraṇam	deṣādhikaraṇam
tathā	sādṛśya, samuccayaḥ	kriyāviśeṣaṇam
yathā	sādṛśya	kriyāviśeṣaṇam
atha	anantarakāla, samuccayaḥ	samuccayaḥ, sambandhaḥ
ca	samuccayaḥ	samuccayaḥ, sambandhaḥ
api	samuccayaḥ	samuccayaḥ, sambandhaḥ
vā	anyataraḥ	anyataraḥ, sambandhaḥ
athavā	anyataraḥ	anyataraḥ
uta	anyataraḥ	anyataraḥ, sambandhaḥ

Table 4.2: Markers with Multiple Relations: Discourse and Kāraka Categorization

English: Rāma and Lakṣmaṇa go to the forest.

versus

Sanskrit: rAmaḥ vanam gacchati. lakṣmaṇaḥ ca tam anugacchati.

Gloss: Ram[m,nom,sg] forest[m,acc,sg] go[present,3p,sg]. Lakshman[m,nom,sg] and[indcl] he[m,acc,sg] follow[present,3p,sg]

English: Rāma goes to the forest, and Lakṣmaṇa follows him.

2. **Ambiguity at semantic level:** In such cases, the ambiguity arises at the semantic level, insofar as the same marker denotes two distinct relations while still operating at the same discourse level. The markers that show such ambiguity are *tathā* marking *sādṛśya* (similarity) and *samuccaya* (conjunction), *atha* marking *samuccaya* (conjunction) and *anantarakāla* (Succeeding), and *tataḥ* marking *anantarakāla* (Succeeding) and *kārya-kāraṇa* (cause-effect). See the following examples:

Sanskrit: saḥ ṣṛṇoti. atha likhati.

Gloss: He[m,nom,sg] listen[present,3p,sg]. then[indcl] write[present,3p,sg].

English: He listens, then writes.

versus

Sanskrit: kim karavāṇi. atha ca kutra gacchāmi.

Gloss: What[m,acc,sg] do[imp,1p,sg]. and[indcl] where[indcl] go[present,1p,sg].

English: What should I do? And where should I go?

3. **Ambiguity at both structural as well as semantic level:** Most of the markers listed in Table 4.2 fall under this category, except those classified under categories 1 and 2. Ambiguity of this type is particularly difficult to resolve.

   because of the nature of the suffix: Some inflections, such as *pañcami* (ablative), in case of markers *yasmāt-tasmāt, yataḥ-tataḥ, ataḥ* and *tṛtiyā* (instrumental), in case of markers *yena-tena* are ambiguous in Sanskrit language. See the examples below:

   Sanskrit: yasmāt karmibhyaḥ ca adhikaḥ yogī bhavati. tasmāt yogī bhava arjuna.

   Gloss:Becuase[indcl] Worker[m,able,pl] and[indcl] greate-er[m,nom,sg] sage[m,nom,sg] to-be[present,3p,sg]. Therefore[indcl] sage[m,nom,sg] to-be[present,2p,sg] Arjuna[m,voc,sg].

   English: Yogis are greater than dutiful people. Therefore, you must become a Yogi.

   Relation: The markers yasmāt-tasmāt together mark the *kārya-kāraṇa* (cause-effect) relation at a discourse level.

   Sanskrit: yasmāt ganṅgā vahate. tasmāt pradeśāt aham āgacchāmi.

   Gloss:Where[m,able,sg] Ganga(river)[f,nom,sg] flow[present,3p,sg]. That[m,abl,sg] land[m,abl,sg] I[nom,sg] come [present,1p,sg].

   English: I am coming from a place where the Ganga flows.

Relation: The markers *yasmāt-tasmāt* together signal the *apādānam* (ablative) relation with the verbs in their respective sentences.

4.9 Algorithms

Discourse Markers play a crucial role in structuring Sanskrit texts, as they explicitly signal relations between sentences. The following algorithms model how such particles can be annotated and linked, thereby making inter-sentential coherence computationally tractable. Each algorithm operates on the principle of detecting specific lexical markers, identifying their syntactic placement, and establishing a semantic-discourse relation between the verbs (V1, V2) of adjacent sentences. See Figure 4.14:

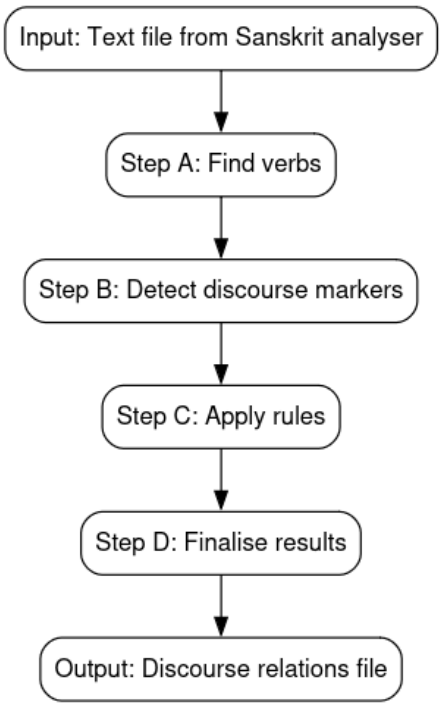

Figure 4.14: Flowchart of the working algorithm on discourse markers

4.9.1 Input

The algorithm takes as input a text file produced by a Sanskrit morphological analyser. Each line of the file represents one word of a sentence and provides detailed linguistic information. For example, each word entry includes:

- **Sentence and word identifiers:** sentence ID (`sid`), word ID (`id`), and clause ID (`cid`),
- **Word form and root:** the surface word (`word`) and its root (`rt`),
- **Morphological features:** for nouns, gender (`liñgam`), case (`vibhaktiḥ`), and number (`vacanam`); for verbs, person (`puruṣaḥ`), number (`vacanam`), tense/mood (`lakāraḥ`), voice (`padī`), and verbal class (`gaṇaḥ`),
- **Prefix and suffix information:** upasarga (prefix), sanādi pratyayaḥ (derivational suffix), kṛt pratyayaḥ (verbal suffix),
- **Syntactic relations:** the basic grammatical relation of the word (`rel_nm`) and the IDs of the related word (`relata_pos_id`, `relata_pos_cid`).

The algorithm pays special attention to two types of words: verbs (including finite verbs and certain participles) and known discourse markers (e.g., *yadi*, *ca*, *api*, *atha*, *kintu*). These serve as the foundation for identifying discourse relations in later steps.

4.9.2 Output

The output is a text file in which each line represents a discourse relation between two verbs. Each line contains:

- the sentence ID and IDs of the two verbs involved,
- the type of discourse relation (e.g., `kārya_dyotakaḥ`, `kārya_kāraṇa_bhāvaḥ`),

- the distance between the verbs in the sentence,
- the rule number that triggered the relation.

In essence, the output specifies which verbal actions are connected and the nature of their discourse relation.

4.9.3 Algorithm Steps

The algorithm proceeds through the following detailed steps:

1. **Identify main verbs:**
    - Scan the input sentence to find all verbs, including finite verbs and certain participles.
    - Mark these verbs as "anchor points" for establishing discourse relations.
2. **Detect discourse markers:**
    - Scan each word in the sentence for known discourse markers (e.g., *yadi*, *yasmāt*, *yathā*, *ca*, *api*, *vā*, *atha*, *kintu*).
    - Determine the scope and direction of the marker:
        - Some markers look forward to the next verb.
        - Some markers look backward to the previous verb.
        - Paired markers (e.g., *yadā–tadā*, *yatra–tatra*, *yathā–tathā*) indicate simultaneous or corresponding events.
3. **Apply discourse rules based on marker type:**
    (a) **Conditional, causal, concessive, and contrast rules:**
        - Interpret markers such as *yadi*, *yasmāt*, *yena*, *yadyapi*, *kintu*.
        - Use grammatical relations (e.g., `karma`, `kāraṇa`, `hetu`) to resolve ambiguity and determine cause, result, concession, or contrast.

(b) **Paired-marker rules:**

- Identify paired markers like *yadā–tadā*, *yathā–tathā*, *yatra–tatra*.
- Mark relations such as simultaneous actions, similar events, or corresponding locations.

(c) **Sequence, conjunction, and alternation rules:**

- Handle temporal sequence markers (*atha*, *anantaram*) to identify order of events.
- Handle conjunction markers (*ca*, *api*, *cāpi*) for simple linking.
- Handle alternation markers (*vā*, *athavā*, *uta*) for optional or alternative actions.

4. **Finalize discourse relations:**

- Remove duplicate relations.
- Sort relations, typically by sentence ID and verb positions.
- Write the final discourse relations to the output file, including:
  - IDs of the two connected verbs.
  - Type of discourse relation (e.g., cause, result, contrast, similarity, sequence, conjunction).
  - Distance between the verbs.
  - Rule number that triggered the relation.

In summary, the algorithm reads the grammatical information of each word, identifies discourse markers, locates the associated verbs, applies the appropriate rules, and outputs the final set of discourse relations.[6]

[6]Detailed input and output samples, along with all algorithm rules, are provided in the appendix.

4.10 Corpus

For the development and evaluation of our rule-based system, we employed the *Śrīmad-bhagavadgītā* as the primary corpus. The text comprises 700 verses, where each verse may encapsulate multiple sentences, or conversely, a single sentence may span multiple verses. In delineating sentence boundaries, we adhered to the criteria: 'eka ṭiṇ vākyam'. Accordingly, a sentence is defined as a set of words containing precisely one finite verb, with every constituent word semantically or syntactically connected to at least one other word within the same set. Previous work by Hellwig (2016) has addressed the challenge of sentence boundary detection. However, it was noted that the algorithm's accuracy significantly decreases when dealing with shorter sentences, particularly those lacking an explicit copula. Given our objective of assessing the performance of the Inter-Sentential Discourse Analyser, and to prevent error propagation from sentence segmentation, we opted for manual annotation of sentence boundaries.

All sentences that contained explicit Inter-Sentential Discourse Markers were identified and extracted from the *Śrīmadbhagavadgītā*. The frequency and distribution of these markers, along with the types of relations they signal- both across sentence boundaries (discourse-level) and within sentences (kāraka and upapada relations) are presented in Table 4.3.

Table 4.3: Marker-wise Relation and Frequency

Markers	**Relation**	**Frequency**
tasmāt	kāraṇa-kārya	14
tasmāt	apādānam	9
yasmāt	kāraṇa-kārya	1
yasmāt	apādānam	1
tataḥ	kāraṇa-kārya	5
tataḥ	apādānam	13

Continued on next page

Table 4.3 - continued from previous page

Markers	Relation	Frequency
tataḥ	anantarakāla	6
yataḥ	kāraṇa-kārya	3
yataḥ	apādānam	1
ataḥ	kāraṇa-kārya	3
ataḥ	apādānam	2
tena	kāraṇa-kārya	0
tena	karaṇa-kartā	8
yena	kāraṇa-kārya	2
yen	karaṇa	7
hi	kāraṇa-kārya	49
hi	sambandhaḥ	17
tadā	samānakālaḥ	9
tadā	kālādhikaraṇam	3
yadā	samānakālaḥ	10
yadā	kālādhikaraṇam	0
tatra	samānādhikaraṇaḥ	1
tatra	deśādhikaraṇam	12
yatra	samānādhikaraṇaḥ	4
yatra	deśādhikaraṇam	1
tathā	sādṛśya	12
tathā	samuccayaḥ	22
tathā	kriyāviśeṣaṇām	6
tathā	anantarakālaḥ	0
yathā	sādṛśya	12

Continued on next page

Table 4.3 - continued from previous page

Markers	Relation	Frequency
yathā	kriyāviśeṣaṇām	4
yadi	āvaśyakatā-pariṇāmaḥ	5
tarhi	āvaśyakatā-pariṇāmaḥ	0
cet	āvaśyakatā-pariṇāmaḥ	6
ṭathāpi	vyabhicāraḥ	1
yadyapi	vyabhicāraḥ	1
atha	anantarakāla	1
atha	samuccayaḥ(inter)	4
atha	sambandhaḥ	5
cha	samuccayaḥ(inter-s)	40
cha	samuccayaḥ(intra-s)	173
cha	sambandhaḥ	37
api	samuccayaḥ(inter-s)	6
api	samuccayaḥ(intra-s)	18
api	sambandhaḥ	70
vā	anyataraḥ (inter-s)	3
vā	anyataraḥ (intra-s)	9
vā	sambandhaḥ	5
uta	anyataraḥ	0
uta	sambandhaḥ	0

4.11 Testing and Results

4.11.1 Testing Methodology

Testing was conducted in two stages to evaluate the performance of the discourse relation identification algorithm.

Stage 1: Controlled Testing To estimate the algorithm's success and coverage, we curated 110 simple sentences covering all possible meanings of each discourse marker. This ensured that all parts of the algorithm could be evaluated, even those corresponding to rare or unusual marker usages. The test instances were categorized as follows:

1. **Correct:** Sentences where the discourse relation was fully correct. Minor parsing errors, such as compounding or morphological analysis errors, were ignored.

2. **Partially correct:** Sentences where at least one element of the relation was correctly identified, including cases with multiple relations between verbs or relations assigned to incorrect entities.

3. **Incorrect:** Sentences that were fully wrong, non-generated, or overly generated.

Stage 2: Corpus Testing The algorithm was evaluated on a corpus of approximately 700 *shlokas* from the *Bhagavadgītā*, focusing on actual linguistic data. Approximately 260 shlokas were tested. For most markers, all instances were evaluated, except for two high-frequency markers[7], where a subset of occurrences was tested.

Sentences were simplified only when strictly necessary. Simplifications included segmentation at appropriate full stops and addition of implicit verbs when finite or infinite forms were missing. Minor

[7]For 'api' and 'ca', only a representative subset of instances was tested (32/94 and 50/250, respectively). This was done as a practical compromise and does not affect the overall evaluation. The marker *anantaram* was also omitted from the analysis as there were only two instances backed by some other *anantarakāla* marker.

sentence-level inconsistencies, such as compounding or morphological errors, were disregarded, since the primary focus was on discourse relations.[8] The evaluation considered:

1. Whether the discourse relation was correctly identified.

2. Whether the identified discourse relation was correctly represented.

4.11.2 Evaluation Metrics

Stage 1: Controlled Sentences

Out of 110 test instances, 81 were fully correct, 21 partially correct, and 8 incorrect. Using **strict scoring**, the accuracy was **approximately 73.6%** (81 out of 110).

Weighted accuracy, where partial matches receive 0.5 credit, was **approximately 83.2%**, calculated as $(81 + 0.5 \times 21)$ out of 110.

The overall error rate was 7.3%. These results confirm that the model is capable of handling a wide range of marker usages in controlled examples, providing a strong foundation for corpus-level evaluation.

Stage 2: *Bhagavadgītā* Corpus

Evaluation metrics were computed using the standard classification counts of true positives (TP): discourse relations correctly identified as discourse; false positives (FP): discourse relations incorrectly marked as non-discourse or intra-sentential; false negatives (FN): non-discourse or intra-sentential relations incorrectly marked as discourse; and true negatives (TN): non-discourse or intra-sentential relations correctly identified as such. For the Stage 2 corpus, the counts were: **TP = 104, FP = 17, FN = 5, and TN = 134**.

The following performance metrics were obtained for the Stage 2 corpus:

[8] Ambiguities within non-discourse (intra-sentential) relations were not accounted for, as these lie outside the scope of this thesis.

- **Precision** = $\frac{TP}{TP+FP} = \frac{104}{104+17} \approx 0.86$
- **Recall** = $\frac{TP}{TP+FN} = \frac{104}{104+5} \approx 0.95$
- **F1-score** = $2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} = 2 \times \frac{0.86 \times 0.95}{0.86+0.95} \approx 0.90$
- **Overall Accuracy** = $\frac{TP+TN}{TP+TN+FP+FN} = \frac{104+134}{104+134+17+5} \approx 0.91$
- **Specificity** = $\frac{TN}{TN+FP} = \frac{134}{134+17} \approx 0.89$
- **Balanced Accuracy** = $\frac{\text{Recall} + \text{Specificity}}{2} = \frac{0.95+0.89}{2} \approx 0.92$

Overall, the model achieved a **precision of 0.86, recall of 0.95, F1-score of 0.90, overall accuracy of 0.91, specificity of 0.89, and balanced accuracy of 0.92.**

Among the tested corpus, there were 104 true positives, 84 instances were fully correct, 16 partially correct, and 4 incorrect. Using partial credit of 0.5, the **weighted accuracy was approximately 88.5%,** calculated as (84 + 0.5 × 16) out of 104.

While **strict accuracy was 80.8%** (84 out of 104). Errors were primarily due to the ambiguous *tasmat* marker or complex relational structures.

These metrics demonstrate that the algorithm reliably identifies discourse relations and effectively discriminates between discourse and non-discourse instances, even in a complex corpus setting.

Summary The evaluation shows that the system achieves high weighted accuracy and robust performance across multiple metrics, validating its effectiveness in both controlled and natural corpus data.

4.12 Detailed Case Study: The Marker *hi*

This section investigates the indeclinable particle *hi*, a commonly occurring lexical item in Sanskrit literature, noted for its significant role in the structuring of discourse. The term *hi* has been attributed with diverse semantic functions across lexicographic and grammatical traditions.

In the Sanskrit-Hindi dictionary authored by Apte,[9] *hi* is associated with the following four senses:

1. *isaliye ki, kyoṃki* - conveying a causal sense ('because')
2. *niḥsandeha, niścaya hi* - expressing emphasis or certainty ('certainly', 'indeed')
3. *udāharaṇasvarūpa* - used to indicate illustration ('for example')
4. *kevala, akelā* - indicating exclusivity ('only', 'alone')

Correspondingly, the Monier-Williams Sanskrit-English dictionary[10] identifies three principal meanings:

1. 'for', 'because', 'on account of.'
2. 'just', 'pray', 'do' (as an intensifier in imperatives)
3. 'indeed', 'assuredly', 'surely', 'of course', 'certainly.y'

Speyer (1886), in his grammatical exposition, offers further interpretative insight. In section §429, he describes *hi* as originating as a weak emphatic particle ('indeed'), though its usage has evolved to adopt a predominantly causal function, particularly in prose texts. Furthermore, in section §443, he states that the particle frequently serves to introduce clauses expressing motives, reasons, causes, or illustrations relative to prior statements.

[9] `sanskrit.uohyd.ac.in/cgi-bin/scl/MT/dict_options.cgi?word=&outencoding=DEV`
[10] `sanskrit.uohyd.ac.in/cgi-bin/scl/MT/dict_options.cgi?word=&outencoding=DEV`

In the annotation framework employed in this study, we do not disambiguate the emphatic or exclusive uses of *hi* (e.g., senses 2 and 4 in Apte). These are collectively categorized under the label *sambandhaḥ*, as opposed to those instances where *hi* explicitly signals causality. The rationale behind this classification choice is the contextual nature of such distinctions, which often require discourse-level or extra-linguistic knowledge beyond the confines of the sentence.

Within the corpus examined, 66 instances of *hi* were found. It is important to note that the commentary of *Śaṅkarācārya* begins only from the 10th verse of Chapter 2, resulting in the first five occurrences of *hi* remaining unannotated. Among the remaining 61 instances, Śaṅkara interprets 46 as causal in nature and 15 as non-causal (*sambandhaḥ*).

Two annotators independently categorized all 66 instances of *hi* without consulting the Śaṅkarabhāṣya. The outcomes of their annotations are summarized in Table 4.4.

	kārya-kāraṇāḥ	**sambandaḥ**	Total
Annotator 1	33	33	66
Annotator 2	49	17	66

Table 4.4: Inter-annotator classification of *hi*

The pairwise confusion matrices - comparing the two annotators with each other and with Śaṅkarabhāṣya - are displayed in Tables 4.5, 4.6, and 4.7, respectively.

		Annotator 2		
		kārya-kāraṇāḥ	**sambandaḥ**	
Annotator 1	**kārya-kāraṇāḥ**	27	6	33
	sambandaḥ	22	11	33
		49	17	66

Table 4.5: Confusion matrix: Annotator 1 vs. Annotator 2

		Śaṅkarabhāṣya		
		kārya-kāraṇāḥ	**sambandaḥ**	
Annotator 1	**kārya-kāraṇāḥ**	27	6	33
	sambandaḥ	22	11	33
		49	17	66

Table 4.6: Confusion matrix: Annotator 1 vs. Śaṅkarabhāṣya

		Śaṅkarabhāṣya		
		kārya-kāraṇāḥ	**sambandaḥ**	
Annotator 2	**kārya-kāraṇāḥ**	42	7	49
	sambandaḥ	7	10	17
		49	17	66

Table 4.7: Confusion matrix: Annotator 2 vs. Śaṅkarabhāṣya

The results in the confusion matrices indicate that Annotator 1's classification diverges from Śaṅkarabhāṣya in approximately 50% of the cases, whereas Annotator 2 exhibits around 30% disagreement. Notably, the degree of inter-annotator mismatch is also substantial. Much of this variance can be attributed to pragmatic ambiguity, where emphatic or exclusive uses of *hi* may be interpreted as causal, depending on the discourse context.

For example, in the commentary on the expression *vyākhyānato viśeṣe pratipattiḥ na hi sandehāt alakṣaṇam*, traditional exegetes interpret *hi* as emphatic. In contrast, Nāgeśa classifies it as a *kāraṇa-dyotakaḥ* - a marker of causal relation.

These patterns underscore the limitation of sentence-level analysis in reliably identifying the function of *hi*, as broader contextual cues are often necessary for accurate disambiguation.

Based on this empirical evidence, a set of rule-based heuristics was developed for the automatic categorization of *hi*. These rules take into account positional factors, nearby pronouns, and syntactic cues to determine whether *hi* functions as causal or non-causal. Table 4.8 presents the comparative performance of this heuristic method against the manually annotated gold standard.

Gold Data ↓	Machine Results →		
	kāraṇa-kāryam	**sambandhaḥ**	Total
kāraṇa-kāryam	41	5	46
sambandhaḥ	5	10	15
Total	46	15	61

Table 4.8: Automatic classification of *hi* compared with gold standard

Algorithm for Annotating the Particle *hi*

In light of the distributional patterns and inter-annotator agreements, the following heuristic algorithm is proposed for annotation:

1. **When *hi* occurs in the second position of the second sentence (S2):**
    - Establish a ***kāraṇa-kārya-sambandhaḥ*** between V1 (verb of S1) and V2 (verb of S2) with a **double-lined edge**.
    - Annotate a ***kāraṇa-dyotakaḥ*** relation between *hi* and V2 in S2 with a **single-lined edge**.
2. **When a pronoun (*sarvanāma*) appears immediately before *hi* at the beginning of S2:**
    - Annotate a ***sambandhaḥ*** relation between *hi* and V2 in S2.
3. **In all other positions:**
    - Annotate a ***sambandhaḥ*** relation between *hi* and V2 in S2.

This algorithm operationalizes the classification of *hi* within a structured annotation framework. It not only assists in resolving the semantic function of this marker but also provides a consistent method for integrating such relations into larger graph-based models of textual coherence.

4.13 Comparison with other frameworks:

We now turn to a comparative analysis of the framework developed in the preceding sections, which we shall refer to as the *Indian Grammatical Tradition* (IGT) framework, owing to its foundation in insights derived from classical Indian linguistic traditions. This framework is evaluated alongside widely adopted discourse annotation frameworks such as the *Penn Discourse Treebank* (PDTB), *Rhetorical Structure Theory* (RST), and the ISO-standardized approach to discourse relations, namely *ISO-DR*.

4.13.1 Penn Discourse Treebank (PDTB) version 3

The Penn Discourse Treebank 3.0 (Prasad et al., 2017) is a richly annotated corpus of English texts that provides a lexically grounded framework for discourse annotation. It builds upon earlier versions ((Prasad et al., 2006) and (Prasad et al., 2008)) and introduces several refinements in the treatment of discourse relations. At its core, PDTB adopts a lexically-driven, predicate-argument view of discourse, treating discourse connectives (e.g., *because*, *although*) as predicates that relate two abstract objects (typically clauses or sentences).

1. Types of Discourse Relations PDTB 3 classifies discourse relations into the following types:

    - **Explicit Relations:** Signaled overtly by discourse connectives (e.g., *since*, *but*, *so*). Connectives act as lexical predicates taking two arguments:
        - *Arg1:* Typically, the first clause or sentence.
        - *Arg2:* The second clause/sentence, often syntactically attached to the connective.

- **Implicit Relations:** Inferred between adjacent sentences with no explicit connective. Annotators insert a suitable connective from the inventory.
- **AltLex (Alternative Lexicalizations):** Relations signaled by expressions other than standard connectives (e.g., *this suggests that*).
- **EntRel and NoRel:** Tags used when coherence arises from entity continuity (EntRel) or when no discourse relation is present (NoRel).

2. Relation Hierarchy PDTB 3 organizes relations in a three-level semantic hierarchy, each level dealing with the finer semantics than the previous level:

Discourse Relation Inventory in PDTB 3.0

(a) Temporal

- **Synchronous:** Events happen simultaneously.

  Example: He was watching TV while she cooked dinner.
- **Asynchronous**
  - **Precedence:** One event occurs before the other.

    Example: She left before he arrived.
  - **Succession:** One event follows the other.

    Example: After the meeting, they had lunch.

(b) Contingency

- **Cause**
  - **Reason:** Arg2 is the reason for Arg1.

    Example: She cried because she was hurt.
  - **Result:** Arg2 is the result of Arg1.

    Example: He missed the bus, so he was late.
  - **Negative result :** Action leads to failure or a negated outcome.

    Example: He tried hard, but didn't win.

- **Condition**
  - **Arg1-as-cond:** Arg1 sets condition for Arg2.

    Example: If it rains, we'll stay in.
  - **Arg2-as-cond:** Arg2 sets condition for Arg1.

    Example: We'll stay in if it rains.
- **Negative Condition**
  - **Arg1-as-negcond:** Arg1 is a negated condition.

    Example: Unless you leave now, you'll be late.
  - **Arg2-as-negcond:** Arg2 is a negated condition.

    Example: You'll be late unless you leave now.
- **Purpose**
  - **Arg1-as-goal:** Arg1 expresses the goal of Arg2.

    Example: To win the race, he trained daily.
  - **Arg2-as-goal:** Arg2 expresses the goal of Arg1.

    Example: He trained daily to win the race.
  - **Arg2-as-negGoal:** Arg2 is a goal that wasn't achieved.

    Example: He trained daily, but failed to win.

(c) Comparison

- **Contrast:** Highlights difference.

  Example: She likes jazz, but he prefers rock.
- **Similarity:** Highlights similarity.

  Example: He speaks Spanish, and so does she.
- **Concession**
  - **Arg1-as-denier:** Arg1 denies expectation from Arg2.

    Example: Even though she was sick, she went to work.

- **Arg2-as-denier:** Arg2 denies expectation from Arg1.

  Example: She was sick, but she went to work.

(d) Expansion

- **Conjunction:** Adds information.

  Example: He cooked dinner and set the table.

- **Disjunction:** Presents alternatives.

  Example: You can take the train or the bus.

- **Equivalence:** Rephrases content.

  Example: He's a pediatrician - in other words, a child doctor.

- **Instantiation**
  - **Arg1-as-instance:** Arg1 is an instance of Arg2.

    Example: Some mammals are endangered. For instance, the panda.

  - **Arg2-as-instance:** Arg2 is an instance of Arg1.

    Example: The panda is endangered. It's one example of mammals at risk.

- **Level-of-detail**
  - **Arg1-as-detail:** Arg1 elaborates Arg2.

    Example: She's very creative. She paints, writes, and sculpts.

  - **Arg2-as-detail:** Arg2 elaborates Arg1.

    Example: She paints, writes, and sculpts. She's very creative.

- **Substitution**
  - **Arg1-as-subst:** Arg1 substitutes Arg2.

    Example: She wanted coffee. Instead, she took tea.

  - **Arg2-as-subst:** Arg2 substitutes Arg1.

    Example: She took tea instead of coffee.

- **Exception**

- **Arg1-as-excpt:** Arg1 is exception to Arg2.

  Example: Everyone went, except John.

- **Arg2-as-excpt:** Arg2 is exception to Arg1.

  Example: Except John, everyone went.

- **Manner**

- **Arg1-as-manner:** Arg1 describes how Arg2 happened.

  Example: He spoke in a whisper, as if afraid.

- **Arg2-as-manner:** Arg2 describes how Arg1 happened.

  Example: He whispered as if afraid.

3. Lexically Grounded Annotation Philosophy PDTB emphasizes lexical grounding - it annotates relations only when anchored in lexical signals, even if implicit. This approach contrasts with theories like Rhetorical Structure Theory (RST), which impose intentional hierarchies.

4. Applications and Impact PDTB 3 supports discourse parsing, machine translation, pronoun resolution, and question answering. Its fine-grained relation types and coverage of implicit structure make it foundational in discourse-aware NLP.

4.13.2 Rhetorical Structure Theory (RST)

Rhetorical Structure Theory (RST), developed by Mann and Thompson (1988), is a theory of text organization that models the coherence and structure of discourse. It is based on the idea that texts are hierarchically structured and that parts of a text are related to each other through rhetorical (functional) relations that serve communicative purposes.

1. Basic Concepts of RST

   - **Discourse units:** RST divides a text into *Elementary Discourse Units* (EDUs), which are the smallest segments of analysis, usually clauses or short sentences.
   - **Nuclearity:** RST distinguishes between:

  - *Nucleus:* The core or central unit of information in a relation.
  - *Satellite:* The supporting or dependent unit that elaborates, justifies, or otherwise supplements the nucleus.
- **Structure:** RST organizes EDUs into a tree structure, where rhetorical relations recursively connect spans of text.
- **Coherence and intentionality:** The theory assumes that discourse is coherent because of underlying speaker intentions, expressed via rhetorical relations.

2. Features of RST Annotation

   - RST annotations result in a hierarchical, tree-like structure with a single root.
   - Relations can span contiguous or non-contiguous spans of text.
   - Only one nucleus per relation (in most standard relations), but some symmetric relations like *List* treat all units as nuclei.
   - Annotations reflect the writer's intent - why one part of the text is included in another.

3. Common Rhetorical Relations Here is a list of frequently used RST relations, grouped functionally:

   - **Elaboration** - Provides additional detail about the nucleus.
     Example: She bought a car. It's a red sedan.
   - **Explanation** - Justifies or clarifies the nucleus.
     Example: He left early because he had a meeting.
   - **Background** - Supplies background context to understand the nucleus.
     Example: Before he became famous, he lived in a small town.
   - **Circumstance** - Indicates conditions or context for the nucleus.
     Example: While she was shopping, he cleaned the house.

- **Condition** - The nucleus holds if the satellite's condition is satisfied.

  Example: If you study, you'll pass the exam.

- **Contrast** - Highlights differences between two units.

  Example: She likes tea, but he prefers coffee.

- **Concession** - Denies expectation from the satellite.

  Example: Although he is wealthy, he is very humble.

- **Antithesis** - A special form of contrast with opposition.

  Example: He worked hard, but others were lazy.

- **Enablement** - The satellite helps enable or support the nucleus action.

  Example: To fix the car, he bought a new battery.

- **Purpose** - Indicates purpose of the nucleus action.

  Example: He exercised to stay healthy.

- **Result** - The satellite is the result of the nucleus.

  Example: She forgot her umbrella, so she got wet.

- **Cause** - The satellite is the reason for the nucleus.

  Example: He missed the bus because he overslept.

- **Evaluation** - The satellite expresses an opinion about the nucleus.

  Example: It's a great idea; it could help us.

- **Evidence** - Supports the truth of the nucleus.

  Example: The test results were positive. This proves the medicine works.

- **Justify** - Justifies performing the nucleus act.

  Example: He should resign. He broke the rules.

- **Summary** - Gives a summary of the nucleus content.

  Example: He traveled, studied, and worked abroad. In short, he had a rich experience.

- **Restatement** - Restates the nucleus in other words.

  Example: She's a cardiologist. In other words, she's a heart doctor.

- **Sequence** - Presents events in temporal order.

  Example: He woke up, brushed his teeth, and went out.

- **Temporal** - Indicates timing relations like before or after.

  Example: He left before she arrived.

- **List** - Lists multiple similar items.

  Example: I bought apples, oranges, and bananas.

- **Means** - The satellite describes how the nucleus is achieved.

  Example: He solved the problem by using a calculator.

- **Comparison** - Highlights similarities between two segments.

  Example: Just as she is tall, so is her sister.

- **Instantiation** - The satellite is an instance of the nucleus.

  Example: Some fruits are tropical. For example, mangoes.

- **Generalization** - The satellite generalizes the nucleus.

  Example: Apples, bananas, and oranges are common fruits.

- **Exception** - Presents an exception to the nucleus.

  Example: Everyone was invited except Sam.

- **Disjunction** - Indicates an exclusive or inclusive choice.

  Example: You can take the train or the bus.

4. Applications of RST

   - Text summarization and simplification.
   - Discourse parsing and segmentation.
   - Argument mining and legal text analysis.
   - Dialogue systems and tutoring systems.
   - Discourse-aware machine translation.

4.13.3 Standardized Discourse relations (ISO 24617-8):

The Standardized version of discourse marking (ISO, 2016) is part of the ISO Semantic Annotation Framework and aims to establish an interoperable, theory-neutral standard for annotating discourse relations. It builds on analyses of multiple discourse theories - such as RST, SDRT, PDTB, Hobbs' theory, and Sanders' taxonomy - and provides a flat set of 20 core discourse relations. These relations can be used to annotate both written and spoken discourse, including dialogues. ISO DR-Core supports both explicit and implicit discourse relations and emphasizes semantic over syntactic distinctions.

1. Core Concepts

    - **Theory-neutral and low-level**: ISO DR-Core allows interoperability by avoiding assumptions about global discourse structures. It focuses on direct annotation of local relations between two abstract objects (e.g., events, facts, conditions).
    - **Flat relational inventory**: It defines a non-hierarchical list of 20 core relations, but can be extended with a taxonomy.
    - **Asymmetry encoded semantically**: Argument roles (e.g., Reason, Result) are specified explicitly in the relation definition.
    - **Abstract and concrete syntax**: Provides formal semantics, XML-based annotation language (DRelML), and integration with dialogue annotations (DiAML from ISO 24617-2).
    - **Multimodal and Dialogue support**: Extends annotation to dialogue acts, incorporating *functional dependence* and *feedback dependence* as discourse relations.

2. ISO Core Discourse Relations with Definitions and Examples

    (a) **Cause:** Arg1 provides a reason for Arg2 to occur.

    Example: He was late because he missed the bus.

(b) **Condition:** Arg1 is a condition that, if realized, leads to Arg2.

Example: If you run, you'll catch the train.

(c) **Negative Condition:** Arg1, if not realized, would result in Arg2.

Example: Unless you hurry, you'll miss it.

(d) **Purpose:** Arg1 enables or sets the goal for Arg2.

Example: To study abroad, she applied for a scholarship.

(e) **Manner:** Arg1 specifies the manner in which Arg2 occurs.

Example: He whispered as if afraid.

(f) **Concession:** An expected result from Arg1 is denied or canceled by Arg2.

Example: Even though it rained, they played.

(g) **Contrast:** Highlights differences between Arg1 and Arg2.

Example: She is outgoing, but he is shy.

(h) **Exception:** Arg2 provides exceptions to a general statement in Arg1.

Example: Everyone went except Sam.

(i) **Similarity:** Highlights similarities between Arg1 and Arg2.

Example: She's a doctor, and so is her brother.

(j) **Substitution:** Arg2 is the preferred or chosen alternative to Arg1.

Example: She wanted tea, but took coffee instead.

(k) **Conjunction:** Arg1 and Arg2 jointly relate to the same situation.

Example: He cooked, and she cleaned.

(l) **Disjunction:** Presents Arg1 and Arg2 as alternatives.

Example: Take the train or fly.

(m) **Exemplification:** Arg2 is a specific example of Arg1.

Example: Many birds migrate - swallows, for example.

(n) **Elaboration:** Arg2 adds more detail to Arg1.

Example: She is artistic. She paints and sculpts.

(o) **Restatement:** Arg2 restates Arg1 from another perspective.

Example: She's a pediatrician - in other words, a children's doctor.

(p) **Synchrony:** Arg1 and Arg2 occur simultaneously.

Example: He whistled while walking.

(q) **Asynchrony:** Arg1 temporally precedes Arg2.

Example: After dinner, they left.

(r) **Expansion:** Arg2 extends or continues the situation in Arg1.

Example: The team won. They celebrated later.

(s) **Functional Dependence:** Arg2 is a dialogue act responding to Arg1.

Example: Q: Are you ready? A: Yes.

(t) **Feedback Dependence:** Arg2 provides feedback about Arg1.

Example: A: I agree. B: Okay, thanks.

3. Additional Features of ISO DR-Core

- **Pragmatic variants**: Recognizes that arguments may involve beliefs or dialogue acts. These are modeled by assigning properties to the arguments, not altering the relation itself.
- **Flat vs. hierarchical organization**: The current standard proposes a flat relation set but is open to taxonomical development via projects like TextLink.
- **Symmetry and argument roles**: Symmetric relations (e.g., Contrast, Conjunction) do not impose ordering, while asymmetric ones (e.g., Cause, Condition) define clear roles like Reason, Result.
- **No strict adjacency or syntactic constraints**: Arguments can be clauses, phrases, or even extended passages, adjacent or non-adjacent.
- **Compatible with dialogue standards**: Integration with ISO 24617-2 allows joint annotation of rhetorical and dialogue coherence.

4. Applications

- **Corpus Annotation**: Suitable for manual and automatic annotation of corpora in both monologic and dialogic data.
- **Discourse Parsing and Summarization**: Supports shallow discourse parsers and systems for textual entailment, summarization, and QA.
- **Interoperability**: Serves as a mapping pivot across major frameworks (PDTB, RST, SDRT), enhancing cross-lingual and cross-framework consistency.

4.13.4 Comparison

Below we present a comparative analysis of four major discourse frameworks: PDTB 3.0, RST, ISO-DR (ISO 24617-8), and the proposed Indian Grammatical Tradition (IGT). Each framework differs significantly in theoretical orientation, with PDTB being lexically grounded, RST relying on rhetorical relations, ISO-DR being dialogue-oriented, and IGT rooted in traditional Sanskrit linguistic theory. The units of annotation, treatment of explicit/implicit relations, and granularity also vary accordingly. While PDTB and ISO-DR emphasize connectives and dialogue structure, respectively, RST focuses on nuclearity, PDTB on hierarchical organization, and IGT integrates lexical Discourse Markers within a dependency grammar framework. See table 4.9 for the detailed comparison. Additionally, we have also demonstrated how these frameworks graphically represent relations differently in Figure 4.15.[11]

4.14 Summary

This chapter examined the function of **Discourse Markers** as structural signals of inter-sentential connectivity in Sanskrit. These particles were shown to go beyond mere lexical fillers, operating

[11]We have excluded the ISO-DR framework because there is no explicit information on the graphical representation.

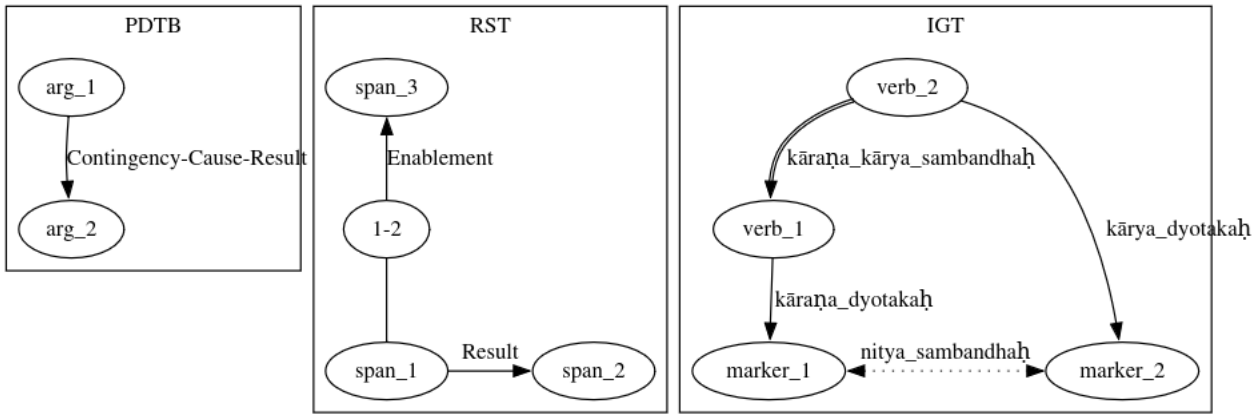

Figure 4.15: Graphical comparison between different frameworks.

instead as explicit indicators of logical, temporal, concessive, and coordinative relations across sentences.

The analysis distinguished **paired and single markers**, and demonstrated how they interact with **local word grouping (LWG)** and map onto arguments and syntactic structures. The principle of *ekavākyatā* was presented as a theoretical basis for treating sentences connected by Discourse Markers as forming a unified interpretive whole. A systematic set of algorithms was then proposed for all relations such as cause–effect, condition–result, concessive contrast, coordination, simultaneity, and opposition. These rules were implemented and tested on corpus data, confirming their practical applicability. Evaluation on the *Bhagavadgītā* corpus shows that the system achieves a strict accuracy of 80.8%, weighted accuracy of 88.5%, precision of 0.86, recall of 0.95, and an F1-score of 0.90, demonstrating reliable identification of discourse relations in both curated and natural text. Finally, a detailed case study of the particle *hi* illustrated both the utility and the challenges of annotation, especially in distinguishing causal from non-causal usages. Taken together, the framework shows how Discourse Markers can be computationally modeled to preserve the semantic and structural coherence of Sanskrit texts.

Parameter	PDTB (3.0)	RST	ISO-DR (ISO 24617-8)	IGT
Theoretical basis	Lexically grounded, predicate-argument model	Rhetorical, intentional, textual cohesion	Dialogue act theory, SDRT-inspired semantics	Inclusion in dependency framework, Traditional insights, Theory-driven
Unit of annotation	Arguments (Arg1, Arg2) tied to connectives	Elementary Discourse Units (EDUs), hierarchical trees	Discourse segments (turns, EDUs, relations)	Sentences (governed by finite verbs)
Relation types	Flat hierarchy: Temporal, Contingency, Comparison, Expansion	Hierarchical: approx 78 nuclearity-based relations	Dialogue structure and semantic relations; 20 core relations	10 lexical relations guided by Ākāṅkṣā (Expectancy), Yogyatā (Semantic compatibility) and Sannidhi (Proximity)
Relation structure	Binary, non-nested, connective-centered	Binary, nested, tree-structured (nucleus-satellite)	Binary or non-binary; supports complex configurations	Binary, Tree-structure
Explicit vs. Implicit	Both annotated distinctly (Implicit requires connective insertion)	Primarily Implicit (rhetorical)	Annotates both (explicit and inferred)	Primarily explicit (optionality of paired markers)
Connective treatment	Central; used to define relation class; lexicalized	Not central; relations inferred from context	Optional; Discourse Markers included where available	Included within sentence, bound by constatnt expectancy
Nuclearity	Not applicable	Yes (Nucleus-Satellite distinction)	Optional (some relations may have nucleus/satellite)	Verbs as nucleus
Granularity	Fine-grained, especially for Implicit	Coarser semantic goals (e.g., JUSTIFY, EVIDENCE)	Medium granularity; includes feedback, control, topic relations	Basic connections
Annotation domain	Written monologue (newswire)	Mostly written; some spoken corpora annotated	Multimodal, written, and spoken dialogue (task-oriented or open)	Textual
Multilingual coverage	English (PDTB), adapted for others (e.g., Hindi, Arabic)	Multiple (RST-DT, Brazilian Portuguese, etc.)	Designed for cross-lingual application	Primarily Sanskrit (with possibility of extension to other languages)
Use in NLP	Widely used in discourse parsing, coherence modeling	Discourse parsing, summarization, coherence modeling	Dialogue systems, multi-agent communication, annotation standardization	Text Analysis, Question-Answer

Table 4.9: Comparison of PDTB, RST, and ISO-DR Discourse Frameworks

CHAPTER 5

Textual Coherence

Textual coherence has gained recent interest in the field of discourse. Coherence refers to the connectedness of text. Connectedness is a quality of good text. Coherence in the text may vary depending on various factors, such as genre and nature of the text, stylistics, speaker and hearer dynamics, environmental settings, and mode of communication. One of the ways to find the connectedness or coherence is through graph, centering, etc, which are among the others. A textual graph can be built using different syntactic, semantic, or discourse relations. And when such is seen to be well-structured, well-connected, then we may assume that the text is coherent. In Indian Textual Traditions, because of its vast and nested nature, there is an inherent coherence in different kinds of texts. In this chapter, we are exploring the possibility of connecting the entire text in a graph form, following the perspective of 'Prabandha-vākya', where the entire text is treated as one homogeneous unit. For such representation, we resort to different possible relations such as Anaphora, Lexically marked discourse relations, verb-expectancies, sentence-groups, event-connections, etc.

5.1 Literature Review

While comprehensive work in the domain of discourse analysis - particularly about establishing connectivity across an entire text - remains limited, a few notable and influential contributions do exist. In the following section, we briefly examine these key efforts.

5.1.1 Western Theories:

Cohesion and Coherence - While Halliday (1976) primarily focused on cohesion, their work significantly informs the broader concept of textual coherence. They argued that cohesion- via reference, conjunction, substitution, and lexical ties- serves as the linguistic mechanism that enables readers to perceive a text as a semantically unified whole. Although they did not provide a direct theory of coherence, their emphasis on how surface-level links contribute to the interpretation of connected meaning laid a crucial foundation for coherence studies. Their model has influenced both theoretical and applied discourse analysis, particularly in identifying how textual elements function together to build global understanding.

RST - Rhetorical Structure Theory (RST), developed by Mann and Thompson (1988), is a theory of text organization that models coherence by positing hierarchical relationships between parts of a text. It treats the entire text as a structured tree composed of connected units called Elementary Discourse Units (EDUs), which are linked via rhetorical relations such as Cause, Elaboration, or Contrast. These relations form a tree structure where one unit (the satellite) supports the interpretation of another (the nucleus), collectively contributing to the communicative goal of the text. This view of coherence emphasizes intentionality and inter-unit dependency, allowing the entire text to be represented as one unified structure. A more detailed discussion of RST, including its taxonomy of relations and applications, is provided in chapter 4.

Text Linguistics - Text Linguistics is a branch of linguistics that investigates the structure and communicative function of texts beyond the sentence level. It views a *text* as more than a random sequence of sentences, emphasizing its status as a structured and meaningful communicative unit. According to de Beaugrande and Dressler (1981), a text qualifies as such only if it satisfies seven *standards of textuality*:

1. **Cohesion**: grammatical and lexical links that hold the text together;

2. **Coherence**: the logical and semantic connectedness of ideas;

3. **Intentionality**: the speaker's or writer's purpose in constructing a meaningful text;

4. **Acceptability**: the reader's or listener's recognition of the text as coherent and appropriate;

5. **Informativity**: the degree to which new or expected information is presented;

6. **Situationality**: the relevance of the text to the context or situation;

7. **Intertextuality**: the relationship of the text to other texts and discourses.

These standards help distinguish a coherent text from a disjointed collection of sentences. Importantly, they underscore the role of both linguistic structure and communicative context in enabling a text to function as a unified whole. From the perspective of discourse analysis, these principles are essential in modeling how readers process entire texts as single units, making text linguistics a foundational framework for understanding and representing textual coherence in both theoretical and computational settings.

Several linguistic and computational models have emerged to treat texts as coherent, unified structures beyond sentence-level analysis. van Dijk and Kintsch (1983) introduced the idea of global coherence by proposing "macrorules" to extract a text's gist, thus modeling how readers construct overarching mental summaries. Veins Theory (Cristea et al., 1998a) builds on discourse trees (like those in RST) to trace referential accessibility paths, allowing long-distance coherence to be formally modeled. Entity-based models, such as Centering Theory and the entity grid (Barzilay and Lapata, 2008), evaluate coherence through continuity in referents across discourse, as seen in co-referent chains in English or topic tracking in Japanese. Text World Theory (Werth, 1999) views text comprehension as a dynamic cognitive process where readers construct and navigate mental representations of narrative worlds. Additionally, knowledge graphs, concept maps, and event schemas also model texts as structured networks, e.g., a news article might be represented as a temporal graph of entities and actions. Graph-based approaches such as Abstract Meaning

Representation (AMR) and Concept Graphs model the semantic structure of entire texts as unified networks, enabling computational systems to capture and process textual coherence holistically. These frameworks- spanning linguistic theory and computational modeling- demonstrate a shared goal: capturing the interrelatedness of parts to reflect a text's unity.

Efforts on Indian Languages - Several studies have advanced the understanding of textual coherence and discourse structure within individual Indian languages. For Hindi, Singh and Sharma (2017) examined textual cohesion, highlighting linguistic devices such as adjective promotion and co-referential noun phrases that help maintain discourse coherence. Verma and Gupta (2019) reviewed topic modeling techniques in Hindi, emphasizing coherence metrics for evaluating thematic consistency. In Gujarati, Joshi et al. (2021) enhanced semantic coherence in topic models by reducing inflectional forms and removing single-letter words, improving thematic clarity. In Bangla, Das et al. (2020) developed the DiMLex-Bangla lexicon and discourse treebank to systematically connect textual segments via discourse relations. For Hindi, Sahoo et al. (2020) proposed an event extraction framework that identifies event triggers and their arguments across sentences, facilitating structured event representation and narrative flow. In Telugu and other languages, Vakada et al. (2023) introduced GAE-ISumm, an unsupervised graph-based summarization model that preserves discourse coherence in summaries.

Cross-lingual efforts have also shown promising progress. Kolluru et al. (2021) proposed Multilingual Fact Linking, introducing the IndicLink dataset spanning English and six Indian languages. Their retrieval-plus-generation model improved fact linking to knowledge graphs, supporting cross-lingual semantic coherence. More recently, Wasi et al. (2024) developed BanglaAutoKG, a system that constructs knowledge graphs from Bengali text using multilingual language models and graph neural networks to enhance semantic accuracy. These graph-based approaches significantly advance the modeling of textual coherence by mapping interconnected concepts across diverse Indian language texts.

5.1.2 In Sanskrit:

Das and Kulkarni (2013) explores textual coherence in Sanskrit through a formal representation of the Mahābhāṣya, emphasizing discourse-level continuity within traditional grammatical theory. Similarly, Subalalitha and Parthasarathi (2012) work on the RST+ framework and saṅgati demonstrates how structural relations can model coherence. These contributions are discussed in greater detail in chapter 4.

Terdalkar and Bhattacharya's work contributes to structured representations of Sanskrit texts that indirectly support coherent interpretation. Hrishikesh Terdalkar (2019)'s study uses knowledge graphs for factoid question answering, demonstrating how relational extraction organizes textual information. Terdalkar and Bhattacharyya (2022)'s work on the *Bhāvaprakāśanighaṇṭu* similarly builds an ontology and query framework to structure domain-specific concepts. Together, these approaches make the texts' internal organization more accessible through graphical representation.

The concept maps available on the *Gītāsupersite* (`https://www.gitasupersite.iitk.ac.in/conceptmaps`) offer a structured and semantic overview of major philosophical themes found in the *Bhagavadgītā*. These maps visually represent key ideas such as *karma*, *jñāna*, and *bhakti*, and connect them across multiple verses and chapters. From the perspective of textual coherence, these maps act as interpretive tools that make explicit the conceptual integration present throughout the text. Despite the dialogic format and diverse content of the *Bhagavadgītā*, the concept maps reveal its cohesive argumentative and philosophical progression. By highlighting inter-dependencies between concepts, the maps provide evidence that the text functions as a coherent whole rather than a mere aggregation of independent teachings. Such representations are particularly valuable for both computational modeling and instructional purposes, as they operationalize coherence through visual and semantic linkage of discourse-level structures.

In his study, Panchukrishnan (2024) presents a compelling analysis of the Vālmīki Rāmāyaṇa through the lens of Tantrayukti, the classical Indian theory of textual construction and reasoning. Focusing on the Kacchit Sarga, where Rāma interrogates Hanumān about the state of Ayodhyā and its people, Panchukrishnan demonstrates how this passage systematically employs specific Tantrayukti devices such as upamāna (analogy), hetu (reasoning), pratyakṣa (direct observation), and nigamana (conclusion). His analysis suggests that the Rāmāyaṇa is not merely a literary narrative but a discourse structured around coherent reasoning and knowledge transmission, following traditional scholastic methodologies. This work not only foregrounds the internal coherence of the text but also proposes a method for interpreting Sanskrit literature as unified, semantically layered wholes. By aligning classical Indian principles with modern textual analysis, the study offers significant contributions to the understanding of textual coherence in Sanskrit, with implications for digital humanities and knowledge modeling frameworks.

Despite the variety of discourse-related efforts across Indian languages, a close examination reveals that these initiatives are relatively few and often narrowly focused. Most existing works tend to address isolated components of discourse processing- such as anaphora resolution, discourse segmentation, or relation identification- without integrating them into a unified framework for holistic discourse understanding. This piecemeal approach, while useful for specific tasks, falls short of capturing the overall structure and coherence of texts as comprehensive semantic units. Given the linguistic richness and diversity of Indian textual traditions, a singular focus on one type of relation or annotation scheme does not adequately represent the full depth of textual meaning. Therefore, this study proposes a more holistic and integrative methodology that systematically searches for and connects multiple coherence cues and discourse strategies within a single framework. By aiming to treat the entire text as a coherent unit, this approach addresses the critical gap evident in the current landscape of discourse analysis in Indian languages. The proposed model aspires to be a one-of-a-kind, unified framework- filling the void left by prior efforts and aligning more closely with the linguistic and philosophical perspectives inherent in classical Indian texts.

5.2 Tools of Coherence

For this pilot study, we select the *Saṃkṣepa-rāmāyaṇam*- a concise poetic composition comprising 100 ślokas- as our representative text. This section outlines a set of coherence tools that may be employed to systematically analyze and represent the text as a unified structure (e.g., a graph). The ultimate aim is to evaluate whether, and to what extent, the text demonstrates internal coherence.

While the tools discussed herein are intended to be broadly applicable across diverse genres of Sanskrit texts, it is important to acknowledge that certain genres may necessitate additional domain-specific strategies. For example, conceptual blocks such as *tapas* or *āsurī sampat* may be essential in interpreting the *Bhagavadgītā*; kinship relations may serve as key structural elements in epic narratives like the *Rāmāyaṇa* and *Mahābhārata*; and definitional or terminological relations may be critical for technical treatises such as the *Nyāya Sūtra* or *Mīmāṃsā*. Nonetheless, the tools presented below are formulated to offer a generalizable framework for assessing coherence across a wide range of texts.

1. **Anaphora Resolution:** Anaphora resolution refers to the process of identifying the antecedent of a pronoun or referring expression within a discourse, thereby enabling coherent interpretation across sentence boundaries. In the present corpus, a total of 84 pronominal entries were identified, comprising 28 indeclinable forms (33.33%) and 56 inflectional forms (66.66%). Within the indeclinable set, 14 instances (50%) functioned anaphorically- most notably through the locative marker *tatra* (5 instances), which referred back to previously mentioned spatial entities, and the temporal marker *tadā* (9 instances), which pointed to earlier or simultaneous temporal contexts. The remaining 14 indeclinable entries were interpreted as indicators of discourse relations.

   Among the 56 inflectional pronominal forms, 27 (48.21%) were used in an anaphoric capacity. These included forms from across the pronominal paradigm: first-person references

such as *aham* and *me* (3 instances), second-person forms like *tvam* and *tvayā* (3 instances), and a broader range of third-person forms (21 instances), including *saḥ*, *tam*, *tasya*, *te*, *tasmīn*, *tat*, and *tām*, among others. The remaining 29 inflectional entries (51.78%) primarily served as modifiers, fulfilling adjectival or referential functions within the sentence.

In terms of referent positioning, out of the 41 total anaphoric pronouns (14 indeclinable and 27 inflectional), the majority- 32 instances (78%)- exhibited a leftward reference (i.e., the antecedent occurred before the pronoun), thus conforming to canonical anaphoric patterns. A smaller proportion, 7 instances (17%), were cataphoric, pointing forward to referents introduced later in the discourse. In 2 cases, the referent could not be conclusively identified due to ambiguity or underspecification. A more granular discussion of these patterns, including classification heuristics and annotation challenges, is presented in Chapter 3, which focuses in depth on anaphora resolution in Sanskrit discourse.

2. **Referring and Co-referring Sentences:** While the present corpus does not exhibit explicit instances of referring or co-referring pronouns used in succession, such constructions are notably observed in other Sanskrit texts, such as the *Bhagavadgītā*. In those texts, co-referring sentences serve as a crucial device for establishing textual coherence by linking adjacent or thematically connected clauses. Their presence enhances continuity and semantic flow, thereby contributing significantly to discourse-level connectivity.

3. **Discourse Markers (DM):** Discourse particles refer to instances in which discourse-level semantic relations between clauses or sentences are signaled through explicit lexical items- typically indeclinable markers- that trigger specific interpretive connections. In the present corpus, out of the 28 identified indeclinable pronouns, 14 instances (50%) were found to function as discourse markers, signaling relations that extend beyond sentence-internal grammar to contribute to the coherence of the surrounding discourse.

Notably, several instances encoded temporally sequential relations, such as *anantarakālaḥ* (immediate succession), which was marked through the use of the indeclinable *tataḥ*. Additionally, semantic similarity or analogy- classified under the relation *sādṛśya-* was observed through the paired usage of *yathā* and *tathā*, each occurring once. These examples demonstrate how Sanskrit employs fixed lexical elements to encode structured semantic relations, facilitating coherent multi-sentence interpretations even in the absence of syntactic subordination.

Such instances are especially significant in languages like Sanskrit, where the syntactic word order is relatively free, and lexical cues play a critical role in guiding discourse processing. A detailed typology and treatment of LMDR, along with criteria for their identification and categorization, are discussed in depth in Chapter 4.

4. **Named and Co-referring Entities:**

Named Entity Recognition (NER) is a fundamental task in Natural Language Processing (NLP) that involves the identification and classification of specific expressions- such as names of persons, places, organizations, etc. - into pre-defined semantic categories. Within the broader framework of discourse analysis, NER plays a crucial role in establishing textual coherence by tracking entities across sentences and larger discourse segments. This enables the maintenance of thematic continuity and referential stability throughout a text.

NER also provides a foundational layer for co-reference resolution, a process that links different expressions that refer to the same real-world entity. These expressions may include alternative names, titles, honorifics, epithets, or descriptive attributes. Together, NER and co-reference resolution significantly enhance our ability to generate a coherent and unified representation of the text's narrative structure.

For example, in the Bhagavad Gītā, the character Arjuna is a named entity who is referred to by multiple alternate names such as Pārtha, Bhārata, Kaunteya, and Guḍākeśa. In addition to these proper nouns, he is also referred to through attributive epithets such as Anagha and Mahābāhu. Though grammatically these may function as viśeṣaṇas (qualifiers), they are pragmatically employed as referential expressions. All such references, which are dispersed across the entire text, point to a single individual. Identifying and linking these occurrences across the discourse is central to understanding the continuity of narrative and argumentation.

The present work combines two key computational modules- Named Entity Recognition and Co-reference Resolution- to address this challenge. This integrated approach has multiple benefits:

(a) Improved Textual Coherence: By tracking entities across verses, we can more accurately capture the thematic flow of the text.

(b) Enhanced Comprehension: Readers can more easily follow the actions, emotions, and transformations of a character, even when different names or descriptions are used.

(c) Enabling Downstream Applications: These methods support further NLP tasks such as automatic summarization, question answering, and semantic analysis.

As part of this research, we manually analyzed the Saṃkṣeparāmāyaṇam and identified a set of named entities along with their co-referential expressions. Each entry in 5.1 includes the named entity, its alternate referential forms[1] and the śloka-s in which these occur.

5. **Kulaka:**

In the Sanskrit literary tradition, the term 'kulaka' denotes a cohesive group of verses that are unified either thematically or contextually. Such clusters typically revolve around a single

[1]Pronouns and adjectives (viśeṣaṇas) that function solely as modifiers and do not serve as independent referential expressions have been excluded from this dataset.

No	Named Entity	Co-referring Expressions
1	Rāmaḥ(8)	Bharatāgrajaḥ(38), anaghaḥ(89), rāghavaḥ(54,57,63,64,69,84,96)
2	Sītā(28)	Maithilī(53,77), vaidehī(74)
3	Bharatḥ(33)	Mahābalaḥ(34)
4	Hanumān(59)	Mahākapiḥ(77)
5	Sugrivaḥ(59)	Vānararājaḥ(61), vānaraḥ(63)
6	Vālī(62)	Hariśvaraḥ(68)

Table 5.1: Named entities and their corresponding co-referring expressions

subject, description, or emotional core. The kulaka structure is particularly employed when a single verse or stanza is insufficient to fully convey the intended expression of the author. In such cases, multiple verses are grouped to form a semantically and syntactically continuous unit.

Importantly, kulakas are not to be interpreted in isolation. Their internal coherence arises from the continuity of sense and structure across verses, which are always consecutive. For a reader familiar with the literary conventions of Sanskrit, the boundaries of a kulaka- where it begins and ends- are usually quite evident. This is why some commentators, such as Mammaṭa, recognize the kulaka as a single analytical unit for descriptive or rhetorical analysis. A more detailed discussion on the theoretical underpinnings of kulaka, along with its contribution to textual unity and coherence, is provided in chapter 2.

While identifying the boundaries of a kulaka may be relatively intuitive for a human reader, it presents a greater challenge for computational systems. However, it is feasible to develop algorithms to detect such groupings by analyzing the presence or absence of finite verbs, both explicit and implicit. In many cases, ślokas within a kulaka are syntactically connected through shared viśeṣaṇas (noun modifiers), infinitival constructions such as 'ktvā, sānaċ, tumun, lyap', and various types of conjunctions. These linguistic features can be leveraged in constructing computational models or algorithms that serve as sentence-boundary identifiers or kulaka detectors.

For this thesis, we have manually identified 3 prominent kulakas in the Saṃkṣepa-rāmāyaṇam. These identifications have been made through close reading and syntactic analysis, in the absence of automated tools.

(a) Kulaka 1: Spanning from śloka-s 2 to 4, this kulaka comprises all the viśeṣaṇas describing Rāma and corresponds to the questions posed by Nārada.

(b) Kulaka 2: Extending from śloka-s 8 to 19, this group also contains descriptive viśeṣaṇas of Rāma but represents Vālmīki's response to Nārada's inquiries.

(c) Kulaka 3: Extending from śloka-s 26 to 28, contains the description of Sītā.

The following figures 5.1, 5.2, 5.3, 5.4, 5.5, 5.6 illustrate the structure, internal cohesion, and linguistic features that unify each kulaka, respectively, providing a visual representation of their formation and analytical significance.

6. **Semantic Expectancies:**

In Sanskrit narrative discourse, verbs play a pivotal role not just in encoding action but in structuring the coherence of the text. A notable phenomenon observed in such texts is semantic expectancy, where certain verbs or actions semantically project or presuppose the occurrence of related actions or events. These expectations operate similarly to Gricean conversational implicatures or pragmatic presuppositions, where the use of one predicate implies the contextual necessity of another. For example, verbs such as *pṛcch* ('to ask') typically imply that a reply will follow; *mṛ* ('to die') presupposes some action such as striking or wounding or kill (*han*); *śodha* ('to search') projects the possibility of discovery (*labh*); *ni+yuj* ('to command') suggests subsequent execution of the task; '*sup* (to sleep) is presupposed by the action of 'waking up' (*ut+stha* and *ā+rabh* ('to begin') evokes the continuation or completion of an action. These pairs exhibit a natural semantic dependency that contributes to discourse-level coherence, even across sentences or verses.

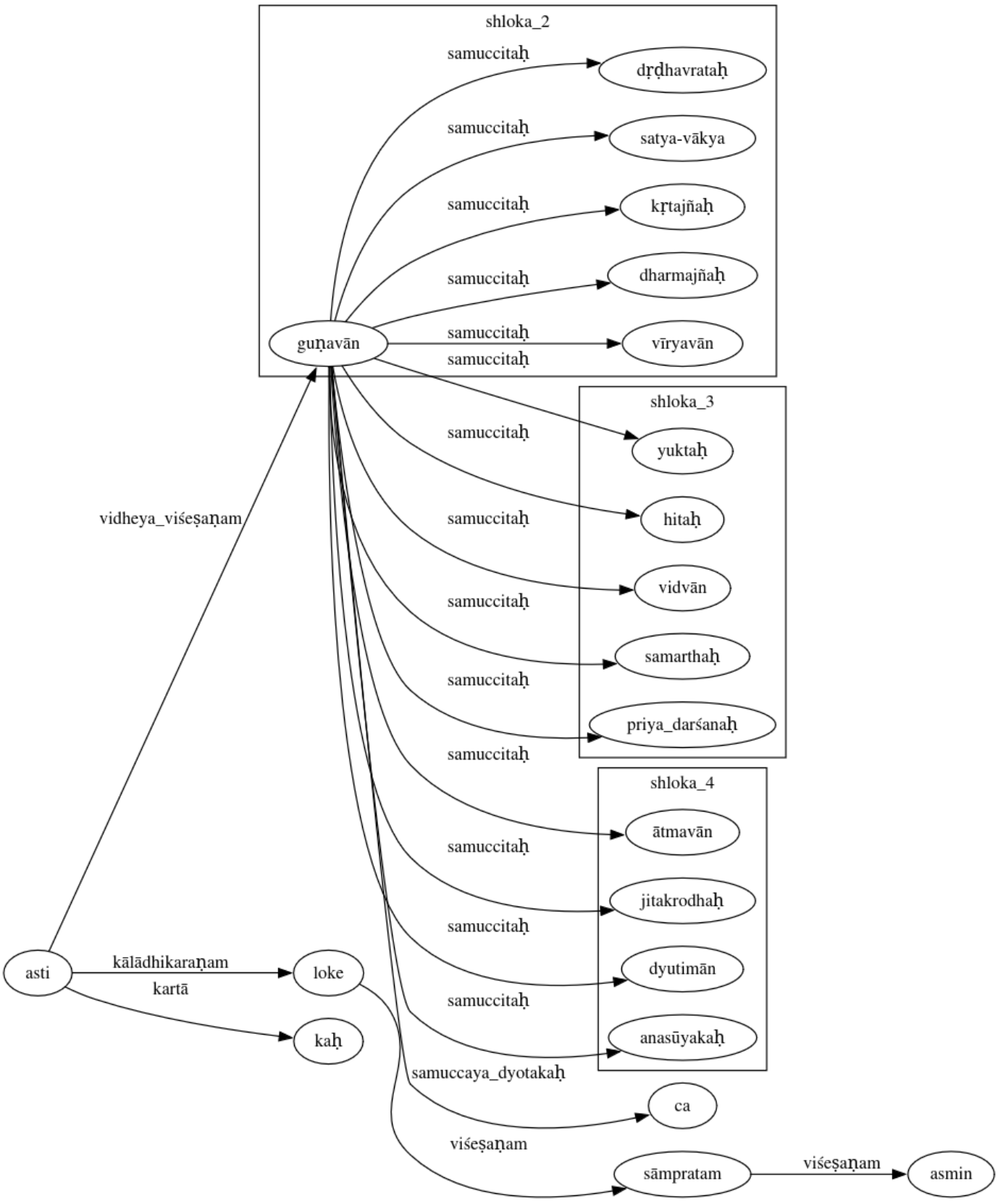

Figure 5.1: Representation of first Kulaka

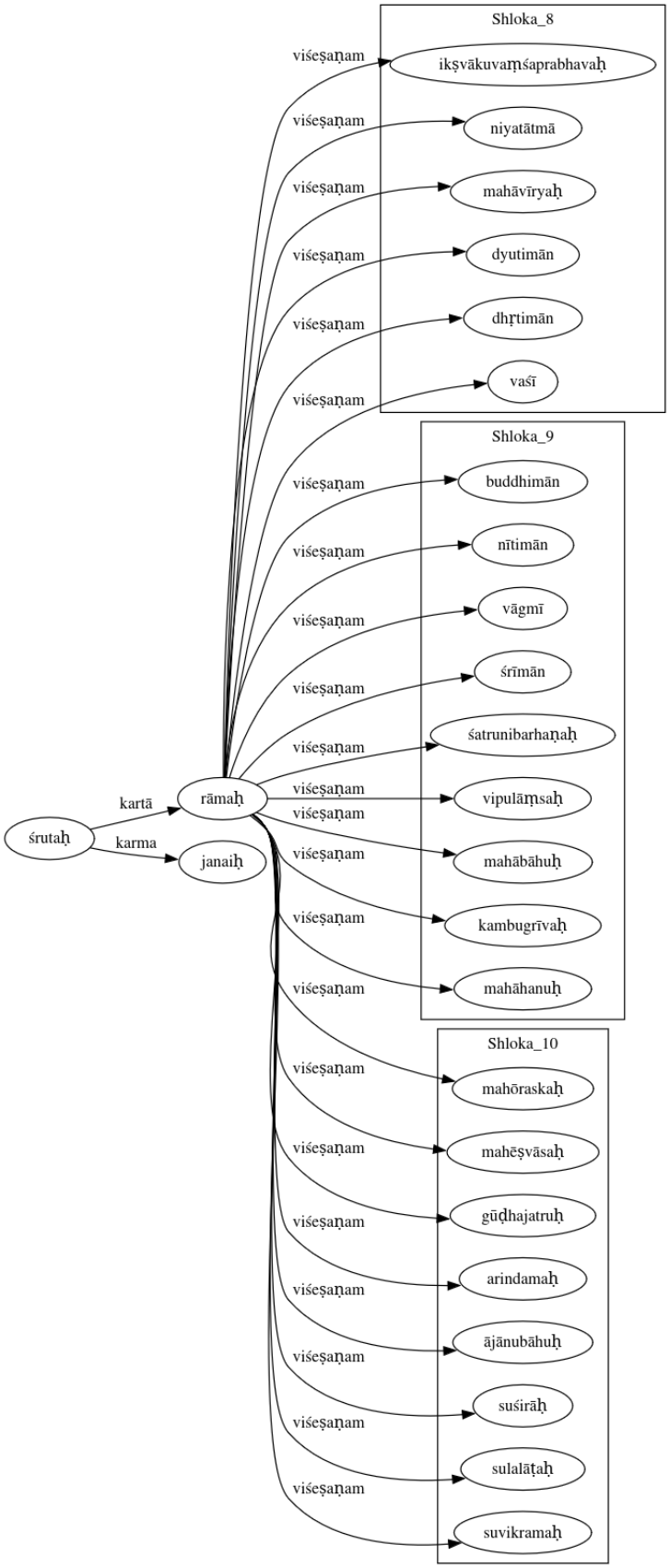

Figure 5.2: Representation of the second Kulaka - part 1

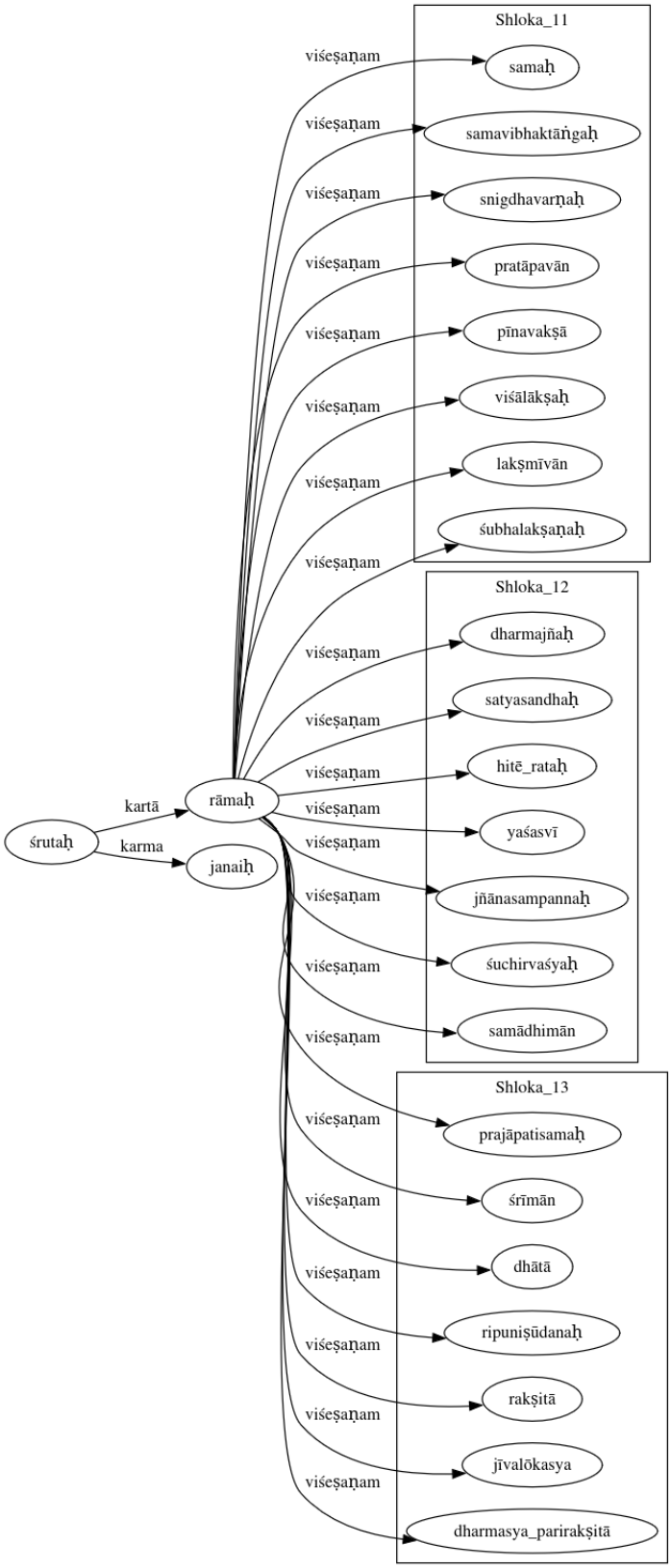

Figure 5.3: Representation of the second Kulaka - part 2

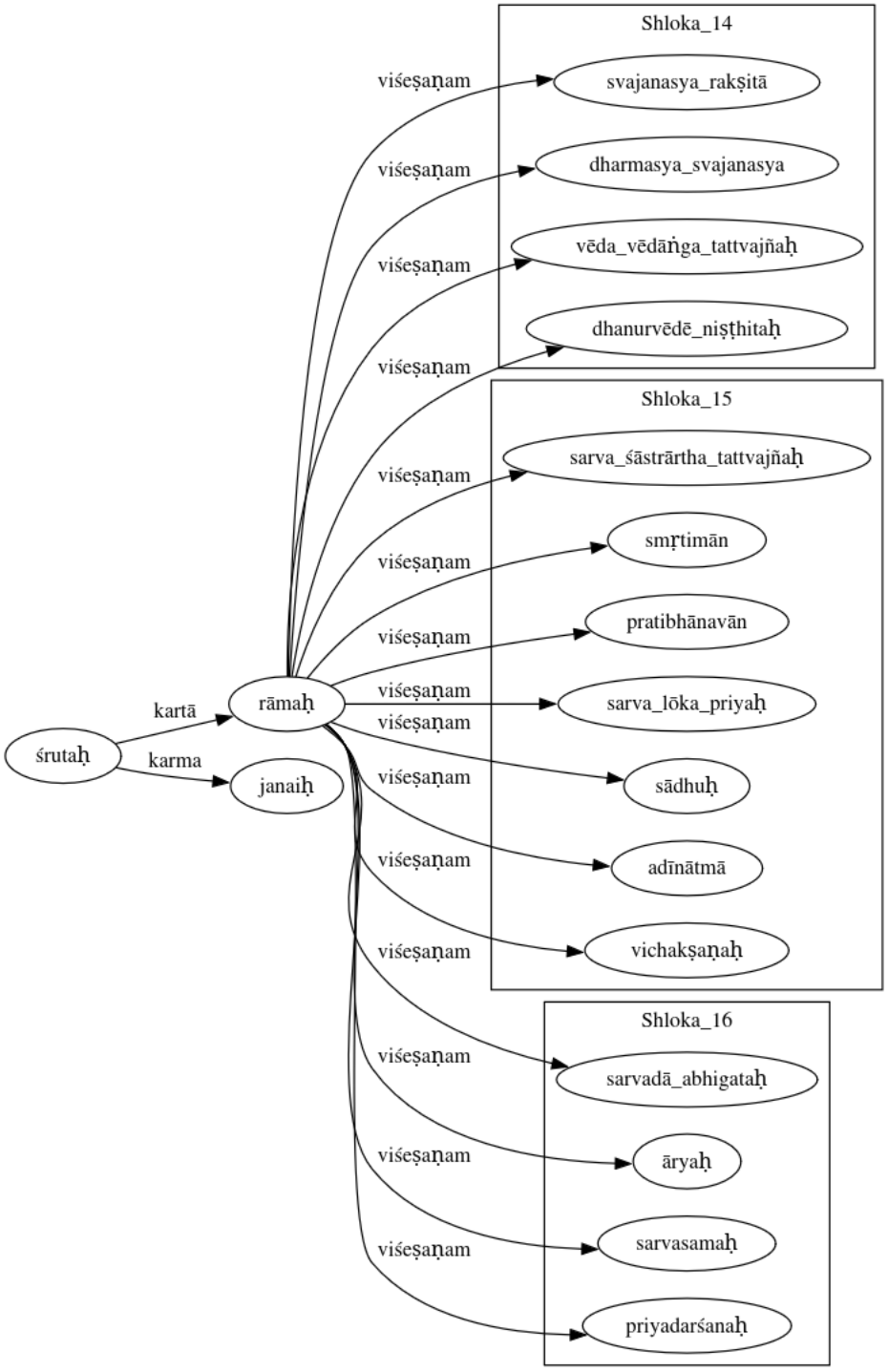

Figure 5.4: Representation of the second Kulaka - part 3

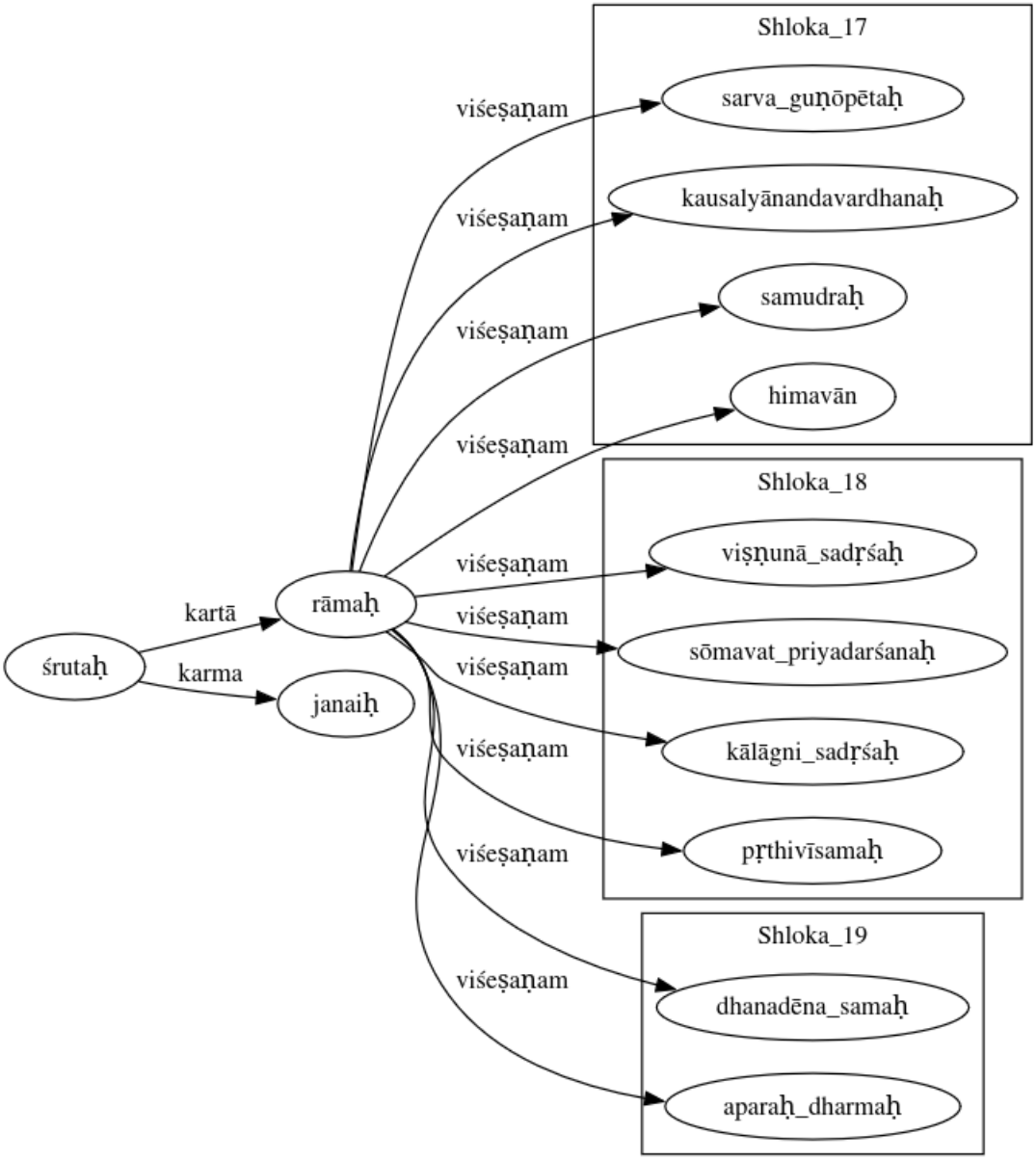

Figure 5.5: Representation of the second Kulaka - part 4

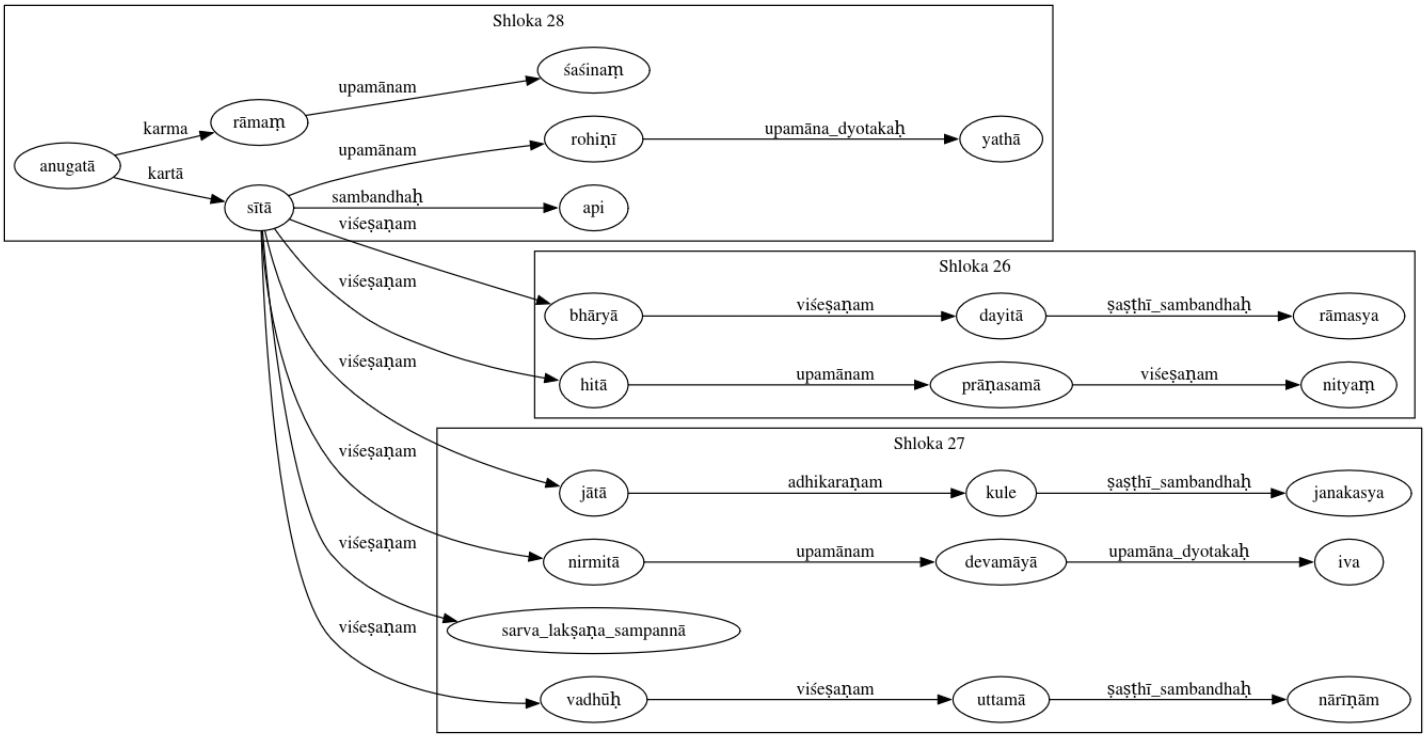

Figure 5.6: Representation of the third Kulaka

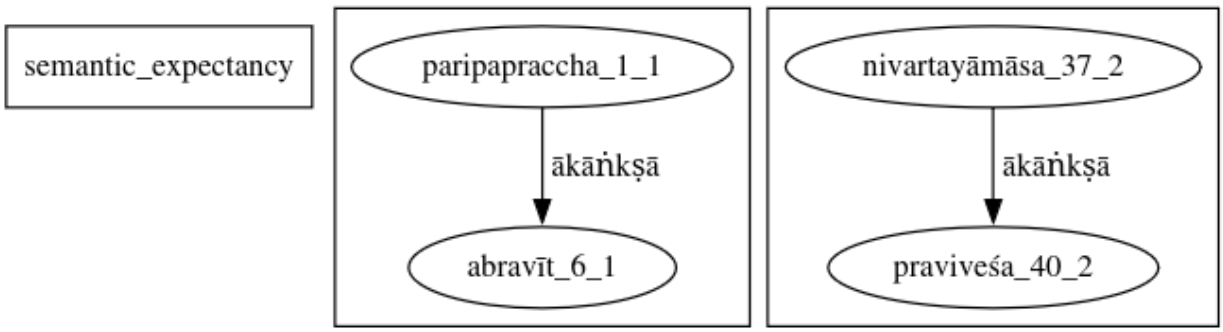

Figure 5.7: Semantic Expectancies in Saṃkṣeparāmāyaṇam

In the corpus of *Saṃkṣepa-rāmāyaṇam*, we find two clear illustrations of such semantic expectancy exerted by an action or event, which support coherence even in the absence of explicit discourse markers. See Figure 5.7. The first example occurs in the verse:

nāradaṃ paripapraccha vālmīkiḥ munipuṅgavam — (1.1)
Gloss: Narada[m,acc,sg] ask[avo,3p,sg] Valmiki[m,nom,sg] foremost-sage[m,acc,sg]
Translation: Vālmīki asked Nārada, the foremost of sages.

śrūyatāmiti ca āmantya prahṛṣṭaḥ vākyam abravīt (6.1)
Gloss: listen[imp,3p,sg,pass] so and say[gerund satisfied[m,nom,sg] sentence[n,acc,sg] say[past-perf,3p,sg]
Translation: Joyfully addressing him and saying "May it be heard," he spoke the speech.

Here, the verb *paripapraccha* ('asked') naturally expects a speech act of response. This expectation is fulfilled by the second verb *abravīt* ('spoke'), even though no overt connective marks the transition. The verbs together form a semantically complete dialogic exchange. The reader infers coherence not from structural linkage but from semantic contingency. The first verb presupposes the second, and thus, their co-occurrence satisfies the principle of *ākāṅkṣā* (expectancy), a key prerequisite in classical Sanskrit sentencehood. This intersentential expectation models coherence as continuity in communicative intent.

The second example is drawn from a later portion of the same text:

nivartayāmāsa tataḥ bharataṃ bharatāgrajaḥ —(37.2)
Gloss: Send[avo,3p,sg,cau] then Bharata[m,acc,sg] Bharata-elder-brother[m,nom,sg]
Translation: Then Rāma, the elder brother of Bharata, caused Bharata to return.

tatra āgamanam-ekāgraḥ daṇḍakān praviveśa ha (40.2)

Gloss: tatra-āgamanam[acc.sg.n] ekāgraḥ[nom.sg.m] daṇḍakān[acc.pl.m] praviveśa[3sg.pf.act] ha[indecl.]

Translation: Focusing on that arrival, he entered the Daṇḍaka forest.

The causative verb *nivartayāmāsa* ('caused to return') implies a conclusion or redirection of an earlier interaction- in this case, sending Bharata back. This verb semantically projects an expectation of what the actor will do next. The subsequent clause fulfills this expectancy with *praviveśa* ('he entered'), indicating Rāma's movement into the forest. Even though these verbs appear across different verses and are not syntactically dependent, the semantic progression they imply- of departure followed by independent action- contributes to the cohesion of the narrative. The coherence here is not established via connectives but through the thematic unfolding of action sequences, preserving narrative unity.

Such expectancy- driven patterns highlight that discourse relations may arise implicitly via semantic dependencies among verbs. In Sanskrit texts like the *Saṃkṣepa-rāmāyaṇam*, where overt discourse markers are often sparse, it is these deep verb-driven structures that uphold textual coherence. They also align with the traditional Mīmāṃsā and Vyākaraṇa principles such as *ākāṅkṣā* and *sannidhi*, which provide a theoretical foundation for treating discourse as semantically interlinked rather than structurally bound. The examples above underscore that verbs are not only syntactic heads but also coherence anchors that semantically connect disparate portions of a text into a meaningful whole.

7. **Event or Episode:**

A text is fundamentally a coherent set of sentences that together convey a unified idea, such as a story, a conversation, or a subject matter. While a text may vary in length, it often lends itself to further segmentation into sub-texts or sub-parts, depending on its complexity and content. This segmentation is not unlike the structure of a human body: just as the body

is composed of organs, which in turn are made up of tissues and cells, a text is composed of sections, paragraphs, sentences, and words. Each of these units performs a specific communicative function, and their mutual interconnectedness contributes to the overall coherence of the text.

In a coherent text, these sub-parts- whether they are paragraphs, conceptual units, episodes, or topics- each play a distinctive role, contributing meaningfully to the whole. The natural flow of a text, therefore, depends on how effectively these units interact. Each unit may be relatively self-contained, conveying a single idea, action, or concept, but it also typically depends on preceding and succeeding units for fuller understanding. This interplay generates a sense of continuity and purpose, which enhances both comprehension and reader engagement.

The nature and boundaries of such sub-units vary across textual genres, media, and communicative goals. For instance, a newspaper article may be organized into paragraphs; a narrative text into episodes or events; an academic or didactic text into topics and subtopics; and an interview into question–answer pairs. One commonality across these formats is that each unit conveys a distinct and comprehensible idea, yet also functions within the broader discursive context. Hence, for analytical purposes, it becomes useful to segment a text not merely by its natural divisions[2]

This process involves both analysis (breaking the text into parts) and synthesis (examining the structural and functional relationships among those parts). The degree of coherence in a text depends on the extent to which its units are systematically arranged and meaningfully interrelated. A highly coherent text exhibits well-integrated sub-units with clear semantic,

[2]Sometimes texts are explicitly segmented - into paragraphs, stanzas, adhyāyas (chapters), kāṇḍas (books), or āhnikas (daily portions) - but not all such divisions reflect conceptual breaks. For example, an āhnika might simply be a portion designed for daily recitation, not necessarily aligned with semantic boundaries. Similarly, an adhyāya may combine multiple, loosely related ideas, but by functional or conceptual units based on thematic coherence and discursive role.

causal, or rhetorical links, thereby facilitating deeper understanding and greater engagement for the reader.

We have divided the text Saṃkṣeparāmāyaṇam into 21 episodic units, each centered around a dominant event. These events are identified primarily through their main verb,[3] typically taken from one of the verses in the unit, following a dependency-based framework where the verb serves as the syntactic and semantic head of the clause. The table 5.2 explains these parts in detail.

The interconnections among the 21 identified episodes are illustrated in Figure 5.8fig:epi, providing a visual representation of the narrative structure and relational dependencies within the text. While the episodes generally form a sequential chain of events, the connections between them are not always strictly linear. Based on our preliminary analysis, we identify three major types of inter-event relations:

(a) Anantarakāla - where one event follows another with only temporal continuity.

(b) Kārya - where one event leads causally to the next, forming a consequence chain.

(c) Arthavāda - where an event or action is supplemented with praise, censure, or commentary.

The process of text segmentation is inherently subjective, shaped by a range of interdependent factors such as genre, discourse structure, interpretive context, and the analyst's perspective. As a result, the task of automating such segmentation remains highly challenging.

[3] In certain instances, a discrepancy may be observed between the designated name of the event and the identified head verb. This arises from the methodological constraint of considering only finite verb forms as heads. While the event name may more intuitively align with an infinitive or non-finite construction, the nearest contextually relevant finite verb has been selected to maintain syntactic consistency within the dependency framework.

Figure 5.8: Episodes in Saṃkṣeparāmāyaṇam

No.	Shloka (from-to)	Name of Event	Head Verb	Things Included
1	1-5	Narada's question	paripapraccha	Praise of Valmiki, Qualities of Rama
2	6-19.1	Valmiki's Answer	abravīt	Qualities of Rama, Rarity of such a human
3	19.2-22	Kaikayi's Demand	ayācat	Preparation of Rama's coronation, Kaikayi asks two wishes
4	23-29	Beginning of Rama's exile	vivāsayāmāsa	Rama is asked for exile, Sita and Lakshman accompany, Qualities of Sita and Lakshmana
5	30-32.1	Rama's entry to the Forest	nyavasan	Guha takes them through Ganga, Activities in Chitrakoot forest
6	32.2-33.1	Dasharatha's demise	jagāma	Dasharatha's sorrow and death
7	33.2-39.1	Bharata's meet	ayācat	Planning for Bharata's coronation, Bharata goes to the forest to convince Rama to return, Rama refuses and sends Bharata with his 'Paduka'
8	39.2-45	Dandakaranya Activities	praviveśa	Kills different demons
9	46-49.1	Deformation of Shurpanakha	virupitā	Deformation of Shurpanakha, Killing of Khara-Dushana demons, Ravana got angry
10	49.2-53.1	Abduction of Sita	jahāra	Marich's help and warning, Sita's abduction, Killing Jatayu
11	53.2-58.1	Rama's grieve	vilalāpa	Rama's grieve for Sita and Jatayu, Search for Sita, Kabandh's kill, Shabari's worship
12	58.2-62	Meet with Hanuman and Sugriv	samāgataḥ	Meets Hanumana, Meets Sugriva, Sharing the tragedy, Friendship, Decision of Vali's death
13	63-70	Vali's kill	nijaghāna	Vali's description about his strength, Rama proves his strength, kills Vali, Sugriva's coronation
14	71-74	Sita's search	prasthāpayāmāsa	Vali orders search, Sampati directs, Hanumana crosses the sea, finds Sita, delivers her the message
15	75-78	Hanumana burns Lanka	āyāt	Hanumana deforms and burns Lanka, Returns and Reports Rama
16	79-80	Creation of bridge	akārayat	Rama takes permission from the sea, Nala creates the bridge
17	81	Obtainment of Sita	upāgamat	Rama goes to Lanka, Kills Ravana, Gets Sita, Becomes happy
18	82-84	Sita's fire-test	viveśa	Sita enters agni as a test, Rama is satisfied
19	85-89	Rama returns Ayodhya	avāpya	Coronation of Vibhishana, Travel by Pushpaka plane, Hanuman's meet with Bharata, Acquires kingdom
20	90-97	Praise of Rama's Kingdom	bhaviṣyati	Health, Fearlessness, Wealth, Happiness, Sacrifice, Donation, Duty-bound-ness in the kingdom, Rama's exit after fulfillment in the future
21	98-100	Praise of Ramayana	paṭhet	Freedom from evil, Achieves heaven and Greatness

Table 5.2: Content and Structure of episodes

5.3 Representation

In the course of our analysis, we have identified and described a set of tools that contribute significantly to textual coherence. These tools serve as structural markers that help delineate and bind various parts of the text into a cohesive whole. While the focus of our study has been the Saṃkṣeparāmāyaṇa, the proposed framework and coherence tools are not limited to this text alone. With appropriate modifications- such as the addition or exclusion of specific features depending on the genre, structure, and thematic content- these tools can be adapted for the analysis of other literary and scriptural compositions as well.

In the visual representation provided in the following figure 5.9 and 5.10,[4] We have integrated all the coherence factors discussed in the previous sections to construct a comprehensive structural map of the Saṃkṣeparāmāyaṇam. This synthesis aims to illustrate how various linguistic, referential, and structural features work in concert to produce a unified and interpretable discourse. The figure 5.9 offers a bird's-eye view of the text's internal architecture and highlights the interplay of the elements that support its narrative continuity. Figure 5.9 represents the very well-connected graph inter-twined with various relations, whereas figure 5.10 represents comparatively isolated or only locally connected ślokas.[5] Even comparatively less connected ślokas display implicit or thematic coherence, as each fulfills a specific function within the larger textual context

5.4 Summary

We have addressed the concept of textual coherence, understood as the semantic and structural connectedness that enables a text to function as a unified communicative whole, in this chapter.

[4]We have used different colours for easily identifying different relations. *red* for Named and coreferring entities, *green* for Kulakas, *black* for semantic expectancies, *orange* for Anaphoric relations, *blue* for episode links, and *purple* for Discourse relations via particles.

[5]For visual clarity, the graph includes only anaphoric expressions and discourse relations that extend across different ślokas; within śloka, sentence-level links have been excluded.

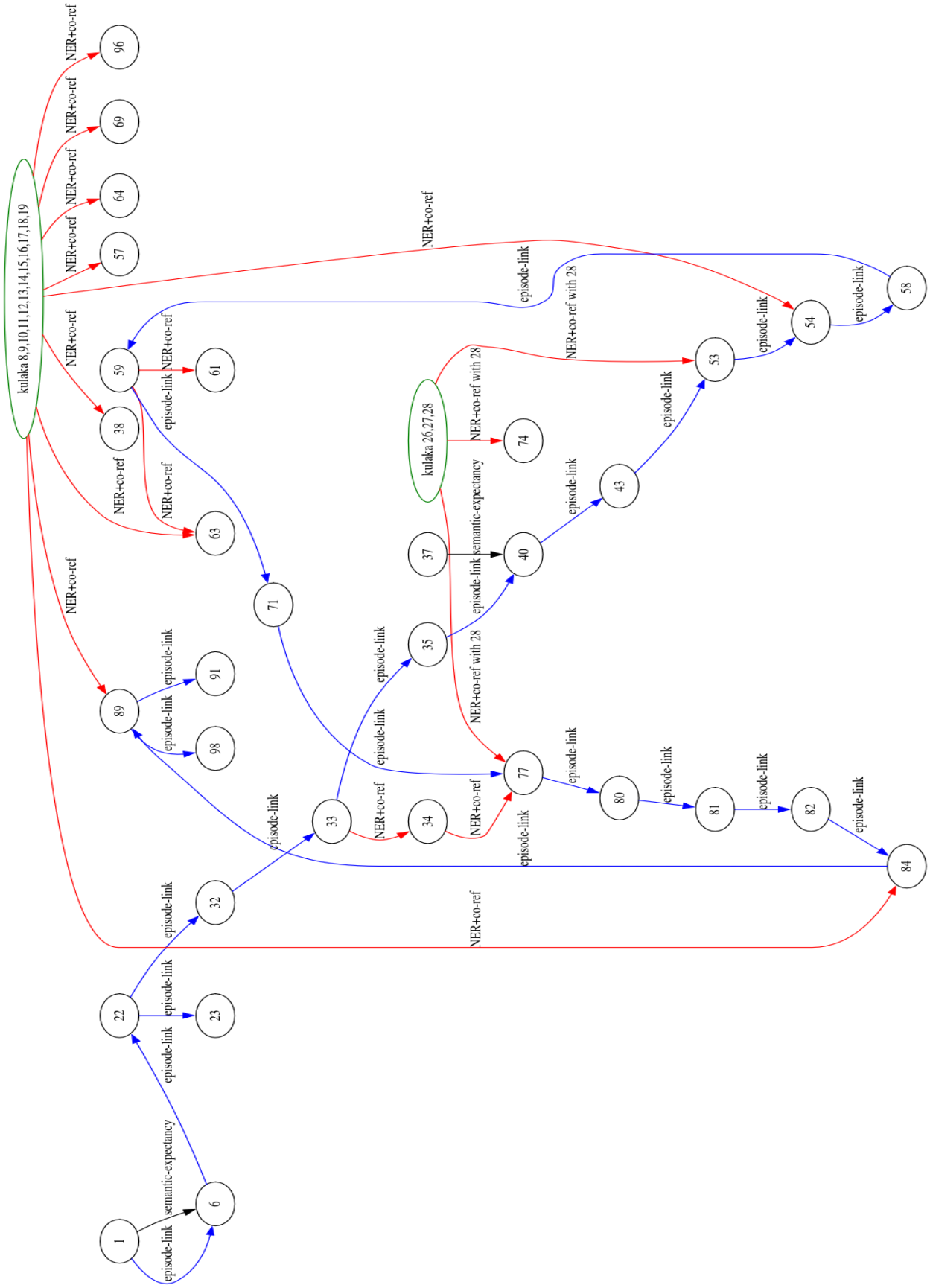

Figure 5.9: Coherence linkage of well-connected ślokas of Saṃkṣeparāmāyaṇam

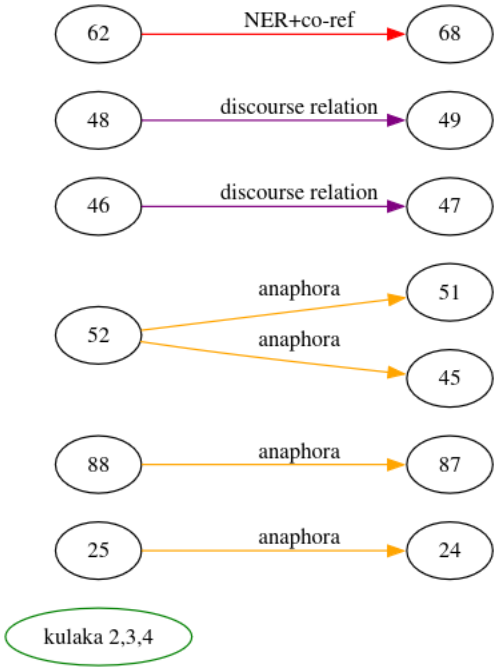

Figure 5.10: Coherence linkage of comparatively non-connected ślokas of Saṃkṣeparāmāyaṇam

We overviewed the principle of *Prabandhavākyatā* and different ways it can express discourse connection.

Using the Saṃkṣeparāmāyaṇa as a case study, the chapter introduces a set of coherence tools - anaphora resolution, lexically marked discourse relations, named entity recognition, kulaka grouping, semantic expectancies, and episodic segmentation- to model the internal structure of the text. These tools facilitate the construction of a comprehensive graph that visually encodes the interdependencies among discourse units. The chapter thus proposes an integrative and adaptable framework for analyzing Sanskrit texts, aligning both with traditional Indian linguistic theory and contemporary discourse analysis methodologies.

CHAPTER 6

Conclusion

6.1 Summary

This research has explored the concept of discourse from both Western and Indian theoretical perspectives, highlighting the need to engage with the largely unexplored and often implicit discourse theories found within the Indian intellectual tradition. A significant part of the study involved an examination of key Sanskrit knowledge systems- Nyāya (Logic), Vyākaraṇa (Grammar), Mīmāṃsā (Hermeneutics), and Sāhitya (Poetics)- and their relevance to modern discourse analysis. Traditional theories such as Saṅgati, Adhyāhāra, and Tatparya, among others, have been discussed, and the possibility of their application is pondered upon from the perspective of computational and linguistic analysis of discourse.

Central to our framework is the principle of Ekavākyatā (unified sententiality), which was demonstrated to function effectively at multiple levels of language analysis. At the word level, it aids in the resolution of anaphoric expressions, particularly pronouns and co-referential terms. At the sentence level, it provides cohesion through discourse markers, and at the text level, it offers techniques to assess internal coherence. Preliminary algorithms were developed and tested for discourse relations phenomena, achieving an F1-score of 0.90, marking an initial step toward computational implementation and automation of Sanskrit discourse analysis.

This chapter concludes the study by summarizing the key challenges and limitations encountered during the research, reflecting on the salience of the findings, and outlining potential future directions.

6.2 Challenges and Limitations

One of the primary challenges faced in this research is the difficulty of demarcating sentence boundaries in Sanskrit, particularly due to its relatively free word order and stylistic flexibility. Writers often exercise discretion in word arrangement, which complicates computational identification of sentence endings and beginnings. This significantly affects the resolution of anaphoric expressions and viśeṣaṇa-viśeṣya (modifier-modified) relations. See the following example:

Option 1:

Sanskrit: S1[Bhojanaṃ karoti] S2[saḥ bālakaḥ khādati tataḥ]

Gloss: S1- Meal[n,acc,sg] do[present,3p,sg]

S2- That[m,nom,sg] boy[m,nom,sg] eat[present,3p,sg] then[indcl].

English: (He) prepares a meal. And then that boy eats.

In this interpretation, *saḥ bālakaḥ* appears within the same sentence, forming a direct modifier-modified relationship.

Option 2:

Sanskrit: S1[Bhojanaṃ karoti saḥ] S2[bālakaḥ khādati tataḥ]

Gloss: S1, Meal[n,acc,sg] do[present,3p,sg] he[m,nom,sg]

S2, boy[m,nom,sg] eat[present,3p,sg] then[indcl]

English: He prepares the meal. And then the boy eats.

Here, *saḥ* and *bālakaḥ* belong to separate sentences, making *bālakaḥ* a potential referent (cataphor) for the pronoun *saḥ*. This shifts the relation to a referential one *(paramarśa)*, which is more complex to resolve computationally.

Such distinctions are made more difficult due to poetic constructions, where sentence segmentation often does not align with syntactic or semantic boundaries. In poetry, *pādas* (metrical units) often replace full stops, and sentence constituents may be distributed across multiple verses. This presents significant challenges in identifying referents, resolving anaphora, and interpreting

discourse particles, especially when attempting to match protasis and apodosis across metrical boundaries.

Additionally, the interpretive openness of poetic texts and the multiplicity of valid commentary traditions further complicate the task. A particular sentence may be parsed or interpreted differently depending on the commentary tradition being followed. For instance:

Sanskrit: S1- Tadā mayūraḥ nṛtyati

S2- Yadā meghaḥ varṣati

Gloss: S1- Then[indcl] peacock[m,nom,sg] dances [present,3p,sg]

S2- when[indcl] clond[m,nom,sg] rain[present,3p,sg]

English: When it rains. (then) peacock dances

While the yadā-tadā pair can mark temporal simultaneity, it may also imply a cause-and-effect relationship. Such ambiguities are challenging to resolve automatically.

Resolving personal pronouns requires inference about speaker, addressee, or referent identity, which is often not explicitly stated. In contrast, demonstrative pronoun resolution is highly dependent on contextual, ontological, and named entity information. Yet, we have formulated a basic algorithm for resolving both types of pronouns, laying the groundwork for further development.

6.3 Relation Representation Paradigms

Distinguishing between pronoun referencing, co-referential expressions, and sentence-level discourse particle relations is another nuanced difficulty, especially in the context of dependency parsing. These relationships differ in directionality and scope: See figure 6.1 for the following observations:

1. Pronoun Resolution: Unidirectional (e.g., from pronoun to noun), with no connection between respective verbs.

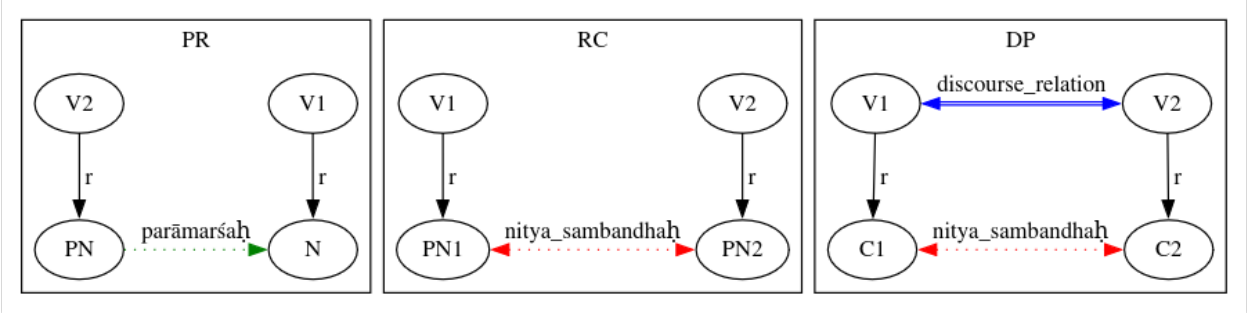

Figure 6.1: Graphical difference between Pronoun Resolution, Co-reference Structures and Discourse Particle Relations

2. Co-reference Structures: Bidirectional, but restricted to the noun level; sentence-level structures remain unlinked.

3. Discourse Particle Relations: Operate at the sentence level, connecting verbs (sentence heads) and forming bilateral dependencies.

6.4 Future Directions

While this work represents an initial step, it opens several promising avenues for future research:

1. Addressing the challenges and limitations identified, particularly regarding sentence boundary detection, implicit relations, and poetic constructions.

2. Expanding the annotated corpus and fine-tuning the algorithms developed thus far through broader evaluation.

3. Advancing toward automation of textual coherence analysis, leveraging the annotated data and rule-based algorithms.

4. Extending the current focus beyond explicit relations to encompass implicit coherence, and scaling the analysis to larger discourse units (i.e., beyond two-sentence clusters).

5. Investigating other traditional concepts from classical texts, beyond Ekavākyatā, including

those mentioned or hinted at in chapter 2 of this thesis, which could prove valuable in discourse-level interpretation and disambiguation.

6. Exploring the applicability of these techniques across other classical, Indian, or foreign languages, highlighting the broader relevance of traditional Indian linguistic theories.

6.5 Significance of the Study

This work represents a unique and pioneering attempt to integrate traditional Indian discourse theories- particularly from Sanskrit śāstric traditions-into the modern field of discourse analysis and computational linguistics. Despite being a preliminary effort, the study is both interdisciplinary and applicable in nature.

Rooted in tradition yet directed toward contemporary problems, this research demonstrates the untapped potential of ancient Indian linguistic frameworks. Concepts derived from *Nyāya, Vyākaraṇa, Mīmāṃsā*, and *Sāhitya-śāstra* have been repurposed to solve real-world issues in sentence and discourse analysis, highlighting their enduring relevance.

The techniques introduced in this study not only open new possibilities for Sanskrit language processing but also suggest a broader framework for cross-linguistic and cross-cultural discourse analysis. This work, therefore, marks the beginning of a new research domain, wherein traditional Indian knowledge systems may offer insightful solutions to modern computational challenges.

6.6 Conclusion

This thesis presents a pioneering and interdisciplinary approach to anaphora resolution and discourse analysis, grounded in the indigenous grammatical and philosophical traditions of India. By engaging with classical Sanskrit knowledge systems, such as *Nyāya, Vyākaraṇa, Mīmāṃsā,* and

Sāhitya, and re-purposing concepts like *Ekavākyatā, Saṅgati*, and *Tātparya*, etc., this work offers a novel framework for understanding textual coherence across lexical, syntactic, and discourse levels. The study systematically analyzed the referential behavior of Sanskrit pronouns, classifying relations into intra-sentential (*kāraka*), inter-sentential discourse, and inter-sentential anaphora relations, all of which were modeled within a unified dependency structure based on the corpus, such as *Bhagvadgītā*. A pilot study on *Saṅkṣeparāmāyaṇam* further demonstrated the practical applicability of the proposed framework, reinforcing the feasibility of integrating traditional linguistic insights with contemporary computational approaches.

While the thesis establishes a strong foundation, it also highlights challenges like sentence boundary ambiguity and context-dependent interpretation, which complicate automated annotation and require advanced algorithms. The developed models provide a crucial starting point for addressing implicit discourse relations, extending analysis to multi-sentence clusters, and exploring frameworks beyond *Ekavākyatā*. With cross-linguistic potential, this work reaffirms the value of classical knowledge systems in advancing modern linguistic research.

Bibliography

Iso 24617-8:2016 – language resource management – semantic annotation framework (semaf) – part 8: Semantic relations in discourse, core annotation schema (dr-core), 2016. Core annotation schema for local discourse relations. 142

Sachin Agarwal, Manoj Srivastava, Pallavi Agarwal, and Ratna Sanyal. Anaphora resolution in hindi documents. In *2007 International Conference on Natural Language Processing and Knowledge Engineering*, pages 452–457. IEEE, 2007.

Vinay Annam, Nikhil Koditala, and Radhika Mamidi. Anaphora resolution in dialogue systems for south asian languages. In *Proceedings of arXiv Preprint*, 2019. URL `https://arxiv.org/abs/1911.09994`.

Annambhatta. *Tarkasamgraha*. Central Library, Baroda, 1917. 33

Vinay Annan et al. Anaphora resolution in dialogue systems for south asian languages. In *Proceedings of the International Conference on South Asian Languages and Linguistics*, 2021. 62

Regina Barzilay and Mirella Lapata. Modeling local coherence: An entity-based approach. *Computational Linguistics*, 34(1):1–34, 2008. 150

Bharata. *Nāṭyaśāstra: A Treatise on Hindu Dramaturgy and Histrionics*, volume 1. Asiatic Society, Calcutta, 1961. 37

Akshar Bharati, Dipti Misra Sharma, and Rajeev Sangal. Towards discourse annotation for indian languages. In *Proceedings of the 11th International Conference on Natural Language Processing (ICON-2013)*. International Institute of Information Technology, Hyderabad, 2013. 99

Bhartrhari. *Vākyapadīya*. University of Travancore, 1942.

Laugakshi Bhaskar. *the Arthasaṅgraha*. Chaukhamb Sanskrit Series, Banaras, 1934. 23

Kumaril Bhatta. *Mimamsa Darshan Of Jaimini With Tantra Vartik Explanation*. Ananda Ashram Sanskrit Series, Pune, 1920.

Kumaril Bhatta. *Mīmāṃsā Ślokavārttika*. Kameshwar Singh Darbhanga Sanskrit Vishva Vidyalaya, 1979.

V P Bhatta. *Gadadhara's Saktivada*. Eastern Book Linkers,Dehli, 1994. 64, 65, 67, 71

Gillian Brown and George Yule. *Discourse Analysis*. Cambridge University Press, 1983.

Patricia L. Carrell. Cohesion is not coherence. *TESOL Quarterly*, 16(4):479–488, 1982.

Dan Cristea, Nancy Ide, and Laurent Romary. Veins theory: A model of global discourse cohesion and coherence. *Proceedings of the 17th International Conference on Computational Linguistics (COLING '98)*, pages 281–285, 1998a. 150

Dan Cristea, Nancy Ide, and Laurent Romary. Veins theory: A model of global discourse cohesion and coherence. *Proceedings of COLING-ACL*, pages 281–285, 1998b.

Debopam Das and Manfred Stede. Developing the bangla rst discourse treebank. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan, 2018. European Language Resources Association (ELRA). URL https://aclanthology.org/L18-1288/. 6, 98

Debopam Das, Manfred Stede, Soumya Sankar Ghosh, and Lahari Chatterjee. Dimlex-bangla: A lexicon of bangla discourse connectives. In *Proceedings of the Twelfth Language Resources and Evaluation Conference (LREC 2020)*, pages 1097–1102, Marseille, France, 2020. European Language Resources Association (ELRA). 7, 151

Monali Das. *discourse analysis of sanskrit texts: first attempt towards computational processing*. PhD thesis, University of Hyderabad, 2016. 9, 14, 17, 46, 47, 100, 102

Monali Das and Amba Kulkarni. Discourse level tagger for mahābhāṣya -a sanskrit commentary on pāṇini's grammar. In *ICON 2013, Noida*, 2013. 152

H. David. Definitions of the sentence (vakya) in the sanskrit tradition: Between grammar and exegesis. *Langages*, 205:27–41, 01 2017. 19

Robert Alain de Beaugrande and Wolfgang U. Dressler. *Introduction to Text Linguistics*. Longman, London and New York, 1981. 149

Deborah Tannen Deborah Schiffrin and Heidi E. Hamilton. *The Handbook of Discourse Analysis*. Blackwell Publishers (P) Ltd, 2001.

R. Deepthi and M. Prabhakar. Anaphora resolution in malayalam texts for discourse coherence. In *Proceedings of the International Conference on Natural Language Processing (ICON)*, India, 2014. 61

Sobha Lalitha Devi and Sobha Rani. Resolution for pronouns in tamil using crf. 2012. URL https://aclanthology.org/W12-5610.pdf. 61

Sobha Lalitha Devi and Sobha Rani. Anaphora resolution in tamil novels. In *Lecture Notes in Computer Science*, volume 8921, 2014. URL https://link.springer.com/chapter/10.1007/978-3-319-13817-6_26.

Sobha Lalitha Devi, Vijay Sundar Ram, and Pattabhi R. K. Rao. A generic anaphora resolution engine for indian languages. In *Proceedings of COLING 2014: Technical Papers*, pages 1172–1181. Association for Computational Linguistics, 2014. URL https://aclanthology.org/C14-1172/.

Bhattoji Dikshita. *Siddhāntakaumudī*. Nirnaysagar Press, Mumbai, 1915.

Dinkar Vishnu Gokhale, editor. *Bhagavadgītā–Bhāṣya of Śrī Śaṅkarācārya*. Oriental Book Agency, Pune, 1931.

Madhav Gopal. *Anaphor Resolution in the Sanskrit Text Panchatantra: A Rule Based Approach to Resolve Lexical Anaphors in Sanskrit*. LAP Lambert Academic Publishing, 2012. ISBN 978-3848492725. 62

Madhav Gopal and Girish Nath Jha. Resolving anaphors in sanskrit. In *Lecture Notes in Artificial Intelligence (LNAI)*, volume 8387, pages 83–92. Springer, 2014. 63

Barbara J. Grosz, Aravind K. Joshi, and Scott Weinstein. Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics*, 21(2):203–225, 1995. doi: 10.1162/jcol.1995.21.2.203. 58

Hasan R. Halliday, M. A. K. *Cohesion in English*. English Language Series, London, 1976. 95, 149

Oliver Hellwig. Detecting sentence boundaries in sanskrit texts. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 288–297, Osaka, Japan, 2016. The COLING 2016 Organizing Committee. URL `https://aclanthology.org/C16-1028`. 124

Jerry R. Hobbs. Resolving pronoun references. *Lingua*, 44(4):311–338, 1978. ISSN 0024-3841. doi: https://doi.org/10.1016/0024-3841(78)90006-2. URL `https://www.sciencedirect.com/science/article/pii/0024384178900062`. 58

Arnab Bhattacharya Hrishikesh Terdalkar. Framework for question-answering in sanskrit through automated construction of knowledge. In *6th International Sanskrit Computational Linguistics Symposium (ISCLS)*, pages 98–117, 2019. 152

Jaimini. *The Purva Mimamsa Sutras Of Jaimini*. THE PANINI office, bhuvaneswari asrama, Allahbad, 1916. 20

S. Jain, S. Agarwal, and A. Joshi. Explicit argument identification for discourse parsing in hindi. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 2010–2015. Association for Computational Linguistics, 2016. doi: 10.18653/v1/D16-1210. 6, 97

A. Jha. Algorithms for anaphora resolution. *Journal of Natural Language Processing*, 15(3): 123–145, 2007. doi: 10.1234/jnlp.2007.12345.

Girish Nath Jha, Sobha Lalitha Devi, Diwakar Mishra, S. K. Singh, and Pravin Pralayankar. Anaphors in sanskrit. In Christer Johansson, editor, *Proceedings of the Second Workshop on Anaphora Resolution (WAR II)*, volume 2 of *NEALT Proceedings Series*, pages 11–25, Bergen, Norway, 2008. URL `http://hdl.handle.net/10062/7361`.

R. Jha. Anaphora resolution algorithm for sanskrit. In *Proceedings of the 5th International Conference on Natural Language Processing*, pages 231–240, 2008. 62

Feng Jiang, Sheng Xu, Xiaomin Chu, Peifeng Li, Qiaoming Zhu, and Guodong Zhou. MCDTB: A macro-level Chinese discourse TreeBank. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 3493–3504, Santa Fe, New Mexico, USA, 2018. Association for Computational Linguistics. URL https://aclanthology.org/C18-1296. 96

M. Joshi, K. Patel, and P. Desai. Improving semantic coherence of gujarati text topic model using inflectional forms reduction and single-letter words removal. *ACM Transactions on Asian and Low-Resource Language Information Processing*, 20(2):1–18, 2021. 7, 151

Daniel Jurafsky and James H. Martin. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Pearson, 3rd edition, 2021.

R. P. Kangle. *The Kautīliya Arthaśāstra*. University of Bombay, 1960.

R. Kar, S. Bandyopadhyay, P. Das, and A. Ghosh. Corpus creation and annotation for discourse relations in bangla. In *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC 2013)*, pages 2254–2261. European Language Resources Association (ELRA), 2013. URL http://www.lrec-conf.org/proceedings/lrec2013/pdf/1035_Paper.pdf. 6, 98

Kawaljit Kaur, Vishal Goyal, and Kamlesh Dutta. Design and implementation of anaphora resolution in punjabi language. In *Proceedings of the 17th International Conference on Natural Language Processing (ICON 2020)*, pages –, 2020. URL https://aclanthology.org/2020.icon-demos.14/. 61

Kalpana Khandale and C. Namrata Mahender. Rule-based design for anaphora resolution of marathi sentence. In *2019 IEEE 5th International Conference for Convergence in Technology (I2CT)*, pages 1–7, 2019. doi: 10.1109/I2CT45611.2019.9033823. 61

Kalpana B. Khandale and C. Namrata Mahender. Resolving anaphora using gender and number agreement in marathi text. In *Proceedings of the International Conference on Data Engineering and Communication Technology*, pages 1–6. Springer, 2021.

Keshav Kolluru, Martin Rezk, Pat Verga, William W. Cohen, and Partha Talukdar. Multilingual fact linking. In *Proceedings of the 3rd Conference on Automated Knowledge Base Construction (AKBC 2021)*, 2021. URL https://openreview.net/forum?id=jbnn1iNNsUj. 7, 151

Amba Kulkarni. *Sanskrit Parsing based on the Theories of śābdabodha*. Indian Institute of Advanced Study, Shimla, 2019.

Amba Kulkarni and Monali Das. Discourse analysis of Sanskrit texts. In *Proceedings of the Workshop on Advances in Discourse Analysis and its Computational Aspects*, pages 1–16, Mumbai, India, 2012. The COLING 2012 Organizing Committee. URL https://aclanthology.org/W12-4701. 100

Amba Kulkarni and K. V. Ramakrishnamacharyulu. Parsing sanskrit texts: Some relation specific issues. In *Malhar Kulkarni, editor, Proceedings of the 5th International Sanskrit Computational Linguistics Symposium. D. K. Printworld(P) Ltd*, 2016. 16, 100

Sheeja S. Kumari and Sobha Lalitha Devi. Annotations of connectives and arguments in malayalam language. *Procedia Technology*, 25:280–285, 2016. 1st Global Colloquium on RAEREST 2016. 6, 99

Kālidāsa. *The Raghuvaṃśa of Kālidāsa: With the Commentary (Sanjīvinī) of Mallinātha*. Motilal Banarsidass, Delhi, 1972.

John Lafferty, Andrew McCallum, and Fernando CN Pereira. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Proceedings of the Eighteenth International Conference on Machine Learning (ICML)*, pages 282–289. Morgan Kaufmann Publishers Inc., 2001. 60

Sobha Lalitha Devi, S. Lakshmi, and Sindhuja Gopalan. Discourse tagging for indian languages. In *Computational Linguistics and Intelligent Text Processing*, pages 469–480. Springer, 2014. doi: 10.1007/978-3-642-54906-9_38. 99

Shalom Lappin and Herbert J. Leass. An algorithm for pronominal anaphora resolution. *Computational Linguistics*, 20(4):535–561, 1994. URL `https://aclanthology.org/J94-4002`.

Dahee Lee, M. Wen Chang, and Kristina Toutanova. End-to-end neural coreference resolution with model stacking. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing (EMNLP 2017)*, pages 2026–2036, 2017. doi: 10.18653/v1/D17-1204. 59

J. Lee and D. Chang. Sri's coreference resolution system. In *Proceedings of the 2008 Conference on Computational Natural Language Learning (CoNLL-2008)*, Manchester, UK, 2008.

W K Lele. *Methodology of Ancient Indian Sciences*. Chaukhamba Surabharati Prakashan, Varanasi, 2006. 48

Jiaqi Li, Ming Liu, Bing Qin, and Ting Liu. A survey of discourse parsing. *Frontiers of Computer Science*, 16(6):166307, 2022. doi: 10.1007/s11704-021-1115-1.

Šárka Zikánová Lucie Mladová and Eva Hajičová. From sentence to discourse: Building an annotation scheme for discourse based on prague dependency treebank. In *Proc. of LREC-2008*, "2008".

John Lyons. *Semantics*, volume 2. Cambridge University Press, New York, 1977. 63, 66

P. R. Madhavan and S. Bandyopadhyay. Anaphora resolution in sanskrit: issues and challenges. In *Proceedings of the 2014 International Conference on Asian Language Processing*, pages 16–23, 2014.

Mammaṭa. *Kāvyaprakāśa of Mammaṭa with the Commentary of Jayadeva*. Nirnaya Sagar Press, Bombay, 2 edition, 1928. 37

William C. Mann and Sandra A. Thompson. Rhetorical structure theory: Toward a functional theory of text organization. *Text: Interdisciplinary Journal for the Study of Discourse*, 8(3):243–281, 1988. 95, 138, 149

Eleni Miltsakaki, Rashmi Prasad, Aravind Joshi, and Bonnie Webber. The Penn Discourse Treebank. In *Proceedings of the Fourth International Conference on Language Resources and Evaluation*, 2004.

S. Mohan, R. Prasad, and K. Kumar. A rule-based anaphora resolution system for telugu language. In *Proceedings of the International Conference on Natural Language Processing (ICON 2010)*, pages 142–150, 2010. 61

Bhatt Nagesh. *vaiyākaraṇasiddhāṃtalaghumañjūṣā*, volume 2. Caukhambā Surabhāratī, Banaras, 1989. 64, 77

Nageshbhatta. *Laghushabdendushekhar*. Central Library, Baroda, 1917.

Sneha Nallani, Manish Shrivastava, and Dipti Sharma. A fully expanded dependency treebank for telugu. In *Proceedings of the WILDRE5 – 5th Workshop on Indian Language Data: Resources and Evaluation*, 2020. 6, 98

Gopal Raghunath Nandargikar, editor. *The Raghuvansha of Kalidasa*. Arya-bhushana Press, Poona, 1897.

Poonam A. Navelker and Jyoti Pawar. Pronominal anaphora resolution in konkani language incorporating gender agreement. In *Proceedings of the 21st International Conference on Natural Language Processing (ICON 2024)*, pages –, 2024. URL https://aclanthology.org/2024.icon-1.27/. 60

T. Neville. Using gazetteer lists for anaphora resolution. In *Proceedings of the International Conference on Natural Language Processing*, pages 123–129, 1999. 59

Umangi Oza, Rashmi Prasad, Sudheer Kolachina, Dipti Misra Sharma, and Aravind Joshi. The hindi discourse relation bank. In Manfred Stede, Chu-Ren Huang, Nancy Ide, and Adam Meyers, editors, *Proceedings of the Third Linguistic Annotation Workshop (LAW III)*, pages 158–161, Suntec, Singapore, August 2009. Association for Computational Linguistics. URL https://aclanthology.org/W09-3029/. 6, 97

Vishwanathan Panchukrishnan. Decoding the systematic construction of valmiki ramayana through tantrayukti principles. In *Journal of Applied Consciousness Studdies*, 2024. 53, 153

Rāmānanda Pandeya. *Vyākaraṇacandrikā*. Chowkhamba Vidyabhawan, Varanasi, India, 2006. A treatise on Sanskrit grammar, especially syntactic structures. 102

Panini. *Aṣṭādhyāyīsūtrapāṭha*. Chaukhamba Publication, Banaras, 1937.

Patañjali. *Vyākaraṇa-Mahābhāṣya of Patañjali*. University of Poona, 1968–1986. Critical edition in multiple volumes.

Soma Paul and Pushpak Bhattacharyya. Discourse relation annotation for indian languages. In Alexander Clark, Chris Fox, and Shalom Lappin, editors, *The Handbook of Computational Linguistics and Natural Language Processing*, pages 649–672. Springer, 2019. doi: 10.1007/978-3-030-15768-5_35. URL https://link.springer.com/chapter/10.1007/978-3-030-15768-5_35. 99

Livia Polanyi. The linguistic structure of discourse. In *The Handbook of Discourse Analysis*, pages 265–281, 2008. 96

Pravin Pralayankar and Devi Sobha, Lalitha. Anaphora resolution algorithm for sanskrit. In *International Sanskrit Computational Linguistics Symposium*, volume 6465 of *Lecture Notes in Computer Science*, pages 209–217, 2010. 62

Rashmi Prasad and Michael Strube. Discourse salience and pronoun resolution in hindi. In *Proceedings of the Discourse Anaphora and Anaphor Resolution Colloquium (DAARC)*, pages 1–10, 2000. 6, 60

Rashmi Prasad, Nikhil Dinesh, Alan Lee, Eleni Miltsakaki, Livio Robaldo, Aravind Joshi, and Bonnie Webber. *The Penn Discourse TreeBank - Annotation Manual 1.0*. University of Pennsylvania, 2006. 96, 134

Rashmi Prasad, Nikhil Dinesh, Alan Lee, Eleni Miltsakaki, Livio Robaldo, Aravind Joshi, and Bonnie Webber. The Penn Discourse TreeBank 2.0. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation*, 2008. 96, 134

Rashmi Prasad, Nikhil Dinesh, Alan Lee, Eleni Miltsakaki, Livio Robaldo, Aravind Joshi, and Bonnie Webber. *The Penn Discourse TreeBank - Annotation Manual 3.0*. University of Pennsylvania, 2017. 96, 134

Soumya Ray Pritish M. Ranjan. A survey of discourse tagging for indian languages. In *Proceedings of the 2022 International Conference on Language Resources and Evaluation (LREC)*, 2022. URL https://aclanthology.org/2022.lrec-1.230/.

Varkheḍī Śrīnivāsa Pāṇḍuraṅgī Vīranārāyaṇa. *Śaktivādaḥ*. Kavikulaguru Kālidāsa Saṃskṛta Viśvavidyālaya, Nagpur, 2020. 66, 77, 78, 80

Ravi Teja Rachakonda and Dipti Misra Sharma. Creating an annotated tamil corpus as a discourse resource. In *Proceedings of the 5th Linguistic Annotation Workshop*, pages 119–123, Portland, Oregon, USA, June 2011. Association for Computational Linguistics. URL https://aclanthology.org/W11-0414. 6, 99

V Raghavan. *Critical Analysis Of Shringar Prakash Of Bhoja*. Madhya Pradesh Hindi Granth Academy, Bhopal, 1963. 40, 42

Pritish M Rajan et al. A survey of discourse tagging of indian languages. In *Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC 2015)*, pages 3151–3158. European Language Resources Association (ELRA), 2015. URL http://www.lrec-conf.org/proceedings/lrec2015/pdf/798_Paper.pdf. 99

R. V. S. Ram and Sobha Lalitha Devi. Pronominal resolution in tamil using tree crfs. In *Proceedings of the 6th Language and Technology Conference*, Poznań, Poland, 2013. 61

K V Ramakrishnamacharyah. vīvakṣātātparyañca. In *Sanskritavidya*. Banaras Hindu University, 2010.

Loganathan Ramasamy et al. Building a discourse annotated corpus in tamil. In *Proceedings of the COLING Workshop on Linguistic Annotation*, 2012.

Sovan Kumar Sahoo, Saumajit Saha, Asif Ekbal, and Pushpak Bhattacharyya. A platform for event extraction in hindi. In *Proceedings of the 12th Language Resources and Evaluation Conference (LREC 2020)*, pages 2241–2250, Marseille, France, 2020. ELRA. 7, 151

Madhusūdana Sarasvatī. *Vaiyākaraṇa Siddhānta Mañjūṣā*, volume 1. Chowkhamba Sanskrit Series, 1961.

P.M. Scharf and H.H. Hock. *Sanskrit Syntax: Selected Papers Presented at the Seminar on Sanskrit Syntax and Discourse Structures, 13-15 June, 2013, Université Paris Diderot*. Sanskrit Library, 2015. ISBN 9781943135004. URL `https://books.google.co.in/books?id=NabejgEACAAJ`.

Apurbalal Senapati and Utpal Garain. Anaphora resolution in bangla using global discourse knowledge. In *Proceedings of the 2012 International Conference on Asian Language Processing*, pages 49–52. IEEE, 2012. 60

Apurbalal Senapati and Utpal Garain. Guitar-based pronominal anaphora resolution in bengali. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 126–130. Association for Computational Linguistics, 2013. URL `https://aclanthology.org/P13-2023/`. 59, 60

Suryakanta Shukla. *Avyayakośa*. Chowkhamba Sanskrit Series Office, Varanasi, India, 2008. A dictionary of Sanskrit indeclinables (avyayas). 102

Utpal Kumar Sikdar, Asif Ekbal, Sudip Kumar Saha, Olga Uryupina, and Massimo Poesio. Anaphora resolution for bengali: An experiment with domain adaptation. *Computación y Sistemas*, 17(2):161–172, 2013. 60

Utpal Kumar Sikdar, Asif Ekbal, Sivaji Saha, Olga Uryupina, and Massimo Poesio. A generalized framework for anaphora resolution in indian languages. *Knowledge-Based Systems*, 109:147–159, 2016. doi: 10.1016/j.knosys.2016.06.033. 6, 60, 61

A. Singh and R. Sharma. Textual cohesion in hindi: A comparative study. *Indian Linguistics*, 2017. 7, 151

Smita Singh, Priya Lakhani, and Pratishtha. Anaphora resolution in hindi language using gazetteer method. *International Journal on Computational Sciences and Applications (IJCSA)*, 4(3):71–78, 2014.

Lalitha Devi Sobha and B. B. Patnaik. Vasisth: An anaphora resolution system for malayalam and hindi. In *Proceedings of the 11th International Conference on Natural Language Processing (ICON 2014)*, pages 121–130, Goa, India, 2014. NLP Association of India. 61

Lalitha Devi Sobha and Dipti Misra Patnaik. Vasisth: An anaphora resolution system for indian languages. In *Proceedings of the DAARC Workshop on Discourse Anaphora and Anaphor Resolution Colloquium*, pages 1–10, 2000. 6, 61

S. Sobha. Algorithms for anaphora resolution in sanskrit. In *Proceedings of the 3rd International Conference on Sanskrit Linguistics*, pages 98–105. Publisher Name, 2010a.

V. Sobha. Anaphora resolution in sanskrit: Rule-based approaches. *International Journal of Sanskrit Computing*, 4(2):23–36, 2010b.

Xiaojun Song, Xintong Li, and Xiang He. Coreference resolution with pointer networks. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP 2018)*, pages 3456–3466, 2018. doi: 10.18653/v1/D18-1371. 59

J. S Speyer. *Sanskrit Syntax*. Leyden: E.J. Brill, University of Cornell, 1886. 102, 130

C.N. Subalalitha and Ranjani Parthasarathi. An approach to discourse parsing using sangati and Rhetorical Structure Theory. In *Proceedings of the Workshop on Machine Translation and Parsing in Indian Languages*, pages 73–82, Mumbai, India, December 2012. The COLING 2012 Organizing Committee. URL `https://aclanthology.org/W12-5607`. 47, 97, 152

Dr Koraḍ Subrahmaṇyaḥ. *Mahā-vākya-vicāraḥ*. lakṣmaṇārya bhavyasmṛti granthamālā, Hyderabad, 2012. 18, 19, 26, 27, 28, 29

Pushpak Bhattacharyya Sweta Khandelwal. Design and development of a discourse tagging scheme for hindi. In *Proceedings of the 2013 International Conference on Natural Language Processing*, 2013. URL `https://www.researchgate.net/publication/307541429`. 6, 97

Daya Shankar Tadav et al. Anaphora resolution for indian languages: A state of the art. *International Journal of Computer Science and Information Security (IJCSIS)*, 19(12):–, 2021.

Hrishikesh Terdalkar and Arnab Bhattacharyya. Semantic annotation and querying framework based on semi-structured ayurvedic text. *Journal of Biomedical Semantics*, 13(1):1–20, 2022. doi: 10.1186/s13326-022-00272-w. 152

Lakshmi Sireesha Vakada, Anudeep Chaluvadi, Mounika Marreddy, Subba Reddy Oota, and Radhika Mamidi. Gae-isumm: Unsupervised graph-based summarization of indian languages. In *Proceedings of the International Joint Conference on Neural Networks (IJCNN 2023)*, pages 1–8. IEEE, 2023. doi: 10.1109/IJCNN54540.2023.10191588. 7, 151

Valmīki. *Saṃkṣepa Ramayanam*. Giri Trading Agency Private Limited, 2022. ISBN 9788179509036. Condensed version of the Rāmāyaṇa in dual language edition.

Teun A. van Dijk and Walter Kintsch. *Strategies of Discourse Comprehension*. Academic Press, New York, 1983. 150

A. Vaswani, A. K. Joshi, and S. Bandyopadhyay. Anaphora resolution in hindi using machine learning. In *Proceedings of the 2015 International Conference on Natural Language Processing and Knowledge Engineering (NLP-KE 2015)*, pages 125–134, 2015. doi: 10.1109/NLP-KE.2015.7374728. 60

Saee Vaze and Amba Kulkarni. Inter sentential discourse relations. In Arnab Bhattacharya, editor, *Proceedings of the 7th International Sanskrit Computational Linguistics Symposium*, pages 67–83, Auroville, Puducherry, India, 2024. Association for Computational Linguistics. URL https://aclanthology.org/2024.iscls-1.6.

Chaitanya Vempaty, Viswanatha Naidu, Samar Husain, Ravi Kiran, Lakshmi Bai, Dipti M. Sharma, and Rajeev Sangal. Issues in analyzing Telugu sentences towards building a Telugu treebank. In Alexander Gelbukh, editor, *Computational Linguistics and Intelligent Text Processing (CICLing 2010)*, volume 6008, pages 50–59. Springer, Berlin, Heidelberg, 2010. 6, 98

S. Verma and N. Gupta. Review and implementation of topic modeling in hindi. *International Journal of Speech Technology*, 22(1):125–134, 2019. 7, 151

P. Vijayalakshmi and Dipti Misra Sharma. Discourse segmentation and marker identification in Telugu. In *Proceedings of the International Joint Conference on Natural Language Processing (IJCNLP)*, 2007. 98

Viśvanātha. *Sāhityadarpaṇa*. Chaukhamba Vidya Bhawan, 1957. 38

Miśra Viśvanātha. *Paribhāṣenduśekharaḥ*. Caukhambā Surabhāratī, Banaras, 1990. 29

Azmine Toushik Wasi, Taki Hasan Rafi, Raima Islam, and Dong-Kyu Chae. Banglaautokg: Automatic bangla knowledge graph construction with semantic neural graph filtering. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 2100–2106, Torino, Italia, 2024. ELRA and ICCL. URL https://aclanthology.org/2024.lrec-main.189. 7, 151

Bonnie Lynn Webber and Aravind K. Joshi. Anchoring a Lexicalized Tree-Adjoining Grammar for discourse. In *Discourse Relations and Discourse Markers*, 1998. 96

Paul Werth. *Text Worlds: Representing Conceptual Space in Discourse*. Longman, London, 1999. 150

Florian Wolf and Edward Gibson. Representing discourse coherence: A corpus-based study. *Computational Linguistics*, 31(2):249–287, 2005. 96

Deniz Zeyrek, Işin Demirşahin, Ayişiği Sevdik-Çalli, Hale Ögel Balaban, İhsan Yalçinkaya, and Ümit Deniz Turan. The annotation scheme of the Turkish discourse bank and an evaluation of inconsistent annotations. In *Proceedings of the Fourth Linguistic Annotation Workshop*, pages 282–289, Uppsala, Sweden, July 2010. Association for Computational Linguistics. URL https://aclanthology.org/W10-1844. 96

Ānandavardhana. *The Dhvanyāloka of Ānandavardhana with the Locana of Abhinavagupta*. Harvard University Press, Cambridge, MA, 1990. 37, 38

Śrī Mādhavācārya. *Jaiminīya Nyāyamālāvistaraḥ*. Kṛṣṇadās Academy, Varanasi, 1989. 24, 46

Ananta Śāstrī, editor. *Kārikāvalī with the Commentaries Muktāvalī, Dinakarī, Rāmarudrī*. Nirṇaya Sāgara Press, Bombay, 1916. 45

Appendices

APPENDIX A

Transliteration Schemes

देवनागरी	IAST*	WX**	देवनागरी	IAST	WX
अ	a	a	ए	e	e
आ	ā	A	ऐ	ai	E
इ	i	i	ओ	o	o
ई	ī	I	औ	au	O
उ	u	u	अं	aṃ	aM
ऊ	ū	U	अः	aḥ	aH
ऋ	ṛ	q	अँ		az
ॠ	ṝ	Q	ऽ	’	Z
ऌ	ḷ	L			
ॡ	ḹ				

* International Alphabet for Sanskrit Transliteration

** WX Scheme

Figure A.1

APPENDIX B

Input Sample for Discourse Relations Algorithm

(avy (sid 1)(id 1)(cid 1)(word aWa)(rt aWa)(pUrvapaxa n)(uwwarapaxa n)(rel_nm sambanXaH)(relata_pos_id 5) (relata_pos_cid 1))

(sup (sid 1)(id 2)(cid 1)(word ciwwam)(rt ciwwa)(pUrvapaxa n)(uwwarapaxa n)(lifgam napuM)(viBakwiH 2)(vacanam eka)(rel_nm karma)(relata_pos_id 3) (relata_pos_cid 1))

(avykqw (sid 1)(id 3)(cid 1)(word samAXAwum)(rt XA1)(pUrvapaxa n)(uwwarapaxa n)(upasarga sam_Af) (sanAxi_prawyayaH X)(kqw_prawyayaH wumun)(gaNaH juhowyAxiH)(rel_nm wumunkarma)(relata_pos_id 5) (relata_pos_cid 1))

(avy (sid 1)(id 4)(cid 1)(word na)(rt na)(pUrvapaxa n)(uwwarapaxa n)(rel_nm prawiReXaH)(relata_pos_id 5) (relata_pos_cid 1))

(wif (sid 1)(id 5)(cid 1)(word SaknoRi)(rt Sak3)(pUrvapaxa n)(uwwarapaxa n)(upasarga X)(sanAxi_prawyayaH X)(prayogaH karwari)(lakAraH lat)(puruRaH ma)(vacanam eka)(paxI parasmEpaxI)(gaNaH svAxiH)(rel_nm X)(relata_pos_id 0) (relata_pos_cid 0))

(sup (sid 1)(id 6)(cid 1)(word mayi)(rt asmax)(pUrvapaxa n)(uwwarapaxa n)(lifgam a)(viBakwiH 7)(vacanam eka)(rel_nm xeSAXikaraNam)(relata_pos_id 5) (relata_pos_cid 1))

(sup (sid 1)(id 7)(cid 1)(word sWiram)(rt sWira)(pUrvapaxa n)(uwwarapaxa n)(lifgam puM)(viBakwiH 2)(vacanam eka)(rel_nm karma)(relata_pos_id 3) (relata_pos_cid 1))

(avy (sid 1)(id 8)(cid 1)(word .)(rt X)(pUrvapaxa n)(uwwarapaxa n)(rel_nm -)(relata_pos_id 0) (relata_pos_cid 0))

(sup (sid 2)(id 1)(cid 1)(word aByAsa-)(rt aByAsa)(pUrvapaxa y)(uwwarapaxa n)(lifgam a)(viBakwiH 0)(vacanam a)(rel_nm bahuvrIhiH)(relata_pos_id 1) (relata_pos_cid 2))

(sup (sid 2)(id 1)(cid 2)(word -yogena)(rt yoga)(pUrvapaxa n)(uwwarapaxa y)(lifgam napuM)(viBakwiH

3)(vacanam eka)(rel_nm hewuH)(relata_pos_id 5) (relata_pos_cid 1))

(avy (sid 2)(id 2)(cid 1)(word wawaH)(rt wawaH)(pUrvapaxa n)(uwwarapaxa n)(rel_nm hewuH)(relata_pos_id 5) (relata_pos_cid 1))

(sup (sid 2)(id 3)(cid 1)(word mAm)(rt asmax)(pUrvapaxa n)(uwwarapaxa n)(lifgam a)(viBakwiH 2)(vacanam eka)(rel_nm karma)(relata_pos_id 5) (relata_pos_cid 1))

(wif (sid 2)(id 4)(cid 1)(word icCa)(rt iR2)(pUrvapaxa n)(uwwarapaxa n)(upasarga X)(sanAxi_prawyayaH X)(prayogaH karwari)(lakAraH lot)(puruRaH ma)(vacanam eka)(paxI parasmEpaxI)(gaNaH wuxAxiH)(rel_nm X)(relata_pos_id 0) (relata_pos_cid 0))

(avykqw (sid 2)(id 5)(cid 1)(word Apwum)(rt Ap1)(pUrvapaxa n)(uwwarapaxa n)(upasarga X)(sanAxi_prawyayaH X)(kqw_prawyayaH wumun)(gaNaH svAxiH)(rel_nm iRkarma)(relata_pos_id 4) (relata_pos_cid 1))

(sup (sid 2)(id 6)(cid 1)(word XanaFjaya)(rt XanaFjaya)(pUrvapaxa n)(uwwarapaxa n)(lifgam puM)(viBakwiH 8)(vacanam eka)(rel_nm samboXyaH)(relata_pos_id 4) (relata_pos_cid 1))

(avy (sid 2)(id 7)(cid 1)(word .)(rt X)(pUrvapaxa n)(uwwarapaxa n)(rel_nm -)(relata_pos_id 0) (relata_pos_cid 0))

APPENDIX C

Output Sample for Discourse Relations Algorithm

2 2 1 kArya_xyowakaH 2 4 1 rule13 1

1 5 1 kArya_kAraNa_BAvaH 2 4 1 rule12 1

APPENDIX D

Algorithm Rules for Discourse Relations

Table D.1: Complete Rule Table (Rules 1–63)

Rule	**Marker(s)**	**Relation Identified**
1	(preprocess)	Verb identification: finite verbs and kṛdanta forms with `toid=0, tocid=0`.
2	(preprocess)	Map syntactic relations (kartā, karman, hetuḥ, etc.) to semantic encodings.
3	yaxi, cew	Necessary consequence (Avaśyakatā–pariṇāma–saṃbandhaḥ).
4	yaxi, cew	Conditional introducer.
5	warhi	Necessary consequence.
6	warhi	Consequence-indicator.
7	yawaḥ	Cause–effect (rel = sambandhaḥ).
8	yawaḥ	Cause-indicator.
9	yasmAw, yena	Cause–effect (rel = sambandhaḥ/hetuḥ).
10	yasmAw, yena	Cause-indicator.
11	yasmAw, yena	Cause–effect (rel = kāraṇa-dyotakaḥ).
12	wawaḥ, awaḥ	Cause–effect (rel = sambandhaḥ).
13	wawaḥ, awaḥ	Effect-indicator.
14	wawaḥ, awaḥ	Cause–effect (rel = kārya-dyotakaḥ).
15	wasmAw, wena	Cause–effect (rel = sambandhaḥ/hetuḥ).

Rule	Marker(s)	Relation Identified
16	wasmAw, wena	Effect-indicator.
17	wasmAw, wena	Cause–effect (rel = kārya-dyotakaḥ).
18	hi	Cause–effect (id = 2).
19	yaxyapi, cexapi, sannapi	Concession (vyabhicāraḥ).
20	yaxyapi, cexapi, sannapi	Cause-indicator.
21	waWApi, warhyapi	Concession.
22	waWApi, warhyapi	Effect-indicator.
23	aWApi	Concession.
24	aWApi	Concessive introducer.
25	paranwu, kinwu	Contrast.
26	paranwu, kinwu	Contrast-indicator.
27	ananwaram	Sequential time.
28	yaxA	Simultaneous time (first).
29	waxA	Simultaneous time (paired).
30	yawra	Same location.
31	wawra	Same location (paired).
32	waWA	Sequence (no preceding yaWA).
33	waWA	Sequential-indicator.
34	waWA	Similarity (paired with yaWA).
35	waWA	Verb-modifier.
36	yaWA	Similarity (rel kriyāviśeṣaṇam).
37	yaWA	Modifier.
38	yaWA	Similarity (when kriyāviśeṣaṇam).
39	aWa	Sequential time.
40	aWa	Sequential-indicator.
41	ca, api, cApi, aWaca, evaFca, iwoZpi	Coordination.

Rule	Marker(s)	Relation Identified
42	ca, api, cApi, aWaca, evaFca, iwoZpi	Coordinator-indicator.
43	vA, vApi, uwa, yaxvA, aWavA, uwApi, uwasviw	Alternation.
50	yaxA → waxA	Fixed temporal pairing.
51	waxA → yaxA	Fixed temporal pairing.
52	yawra → wawra	Fixed locative pairing.
53	wawra → yawra	Fixed locative pairing.
54	yaWA → waWA	Fixed similarity pairing.
55	waWA → yaWA	Fixed similarity pairing.
56	yaxvaw → waxvaw	Fixed pairing.
57	waxvaw → yaxvaw	Fixed pairing.
58	yawaH → wawaH	Fixed pairing.
59	wawaH → yawaH	Fixed pairing.
60	yena wena	Fixed pairing (rel = karaṇam/hetuḥ/viśeṣaṇam).
61	yena wena	Same as above.
62	yasmAw wasmAw	Fixed pairing (rel = apādānam/hetuḥ/viśeṣaṇam).
63	yasmAw wasmAw	Same as above.

The algorithm is available as a part of the *Samsaadhanii* toolkit on the github at https://github.com/samsaadhanii/scl

Sl No. 40442

Notification No. 1

हैदराबाद विश्वविद्यालय

UNIVERSITY OF HYDERABAD

(A Central University Established by an Act of Parliament)

P.O. CENTRAL UNIVERSITY, GACHIBOWLI, HYDERABAD - 500 046 (INDIA)

SEMESTER GRADE TRANSCRIPT

REGULAR EXAMINATION

REG. NO. **20HSPH03**

NAME OF THE STUDENT **VAZE SAEE CHANDRASHEKHAR**

MONTH AND YEAR OF **FEB 2021** **SEMESTER 1**

COURSE **Ph.D. SANSKRIT STUDIES**

PARENT'S NAME **CHANDRASHEKHAR VAZE / MEDHA VAZE**

COURSE NO	TITLE OF THE COURSE	CREDITS	RESULTS
SK801	RESEARCH METHODOLOGY	4	PASS
SK814	SPECIAL READING IN NAVYA NYAYA	4	PASS

SEMESTER GRADE POINT AVERAGE (SGPA) :**10.0**

(In words) :TEN POINT ZERO

DEPUTY REGISTRAR
(ACAD & EXAMS)

Date Of Result Notification : **Mar 10, 2021**

Date Of Download : **Aug 16, 2025**

This is system generated certificate issued by O/o The Controller of Examinations, University of Hyderabad

Sl No. 47831

Notification No. 1

हैदराबाद विश्वविद्यालय

UNIVERSITY OF HYDERABAD

(A Central University Established by an Act of Parliament)

P.O. CENTRAL UNIVERSITY, GACHIBOWLI, HYDERABAD - 500 046 (INDIA)

SEMESTER GRADE TRANSCRIPT

REGULAR EXAMINATION

REG. NO. **20HSPH03**

NAME OF THE STUDENT **VAZE SAEE CHANDRASHEKHAR**

MONTH AND YEAR OF **JUL 2021** **SEMESTER 2**

COURSE **Ph.D. SANSKRIT STUDIES**

PARENT'S NAME **CHANDRASHEKHAR VAZE / MEDHA VAZE**

COURSE NO	TITLE OF THE COURSE	CREDITS	RESULTS
EN875	RESEARCH AND PUBLICATION ETHICS	2	PASS
SK813	INDIAN AND WESTERN LOGICAL SYSTEMS	4	PASS

SEMESTER GRADE POINT AVERAGE (SGPA) :**10.0**

(In words) :TEN POINT ZERO

DEPUTY REGISTRAR
(ACAD & EXAMS)

Date Of Result Notification : **Aug 5, 2021**

Date Of Download : **Aug 16, 2025**

This is system generated certificate issued by O/o The Controller of Examinations, University of Hyderabad

केन्द्रीयसंस्कृतविश्वविद्यालयः
संसदः अधिनियमेन स्थापितः
(प्राक्तनं राष्ट्रियसंस्कृतसंस्थानम्)
गङ्गानाथझापरिसरः प्रयागराजः

Central Sanskrit University
Established by an Act of Parliament
(Formerly Rashtriya Sanskrit Sansthan
Ganga Nath Jha Campus, Prayagraj

Ref./पत्राङ्क :

Date / दिनाङ्क :

के. सं. वि./उशती/2025/291

14.08.2025

प्रमाण-पत्र

प्रमाणित किया जाता है कि सई वझे (शोधछात्रा, संस्कृत अध्ययन विभाग, हैदराबाद विश्वविद्यालय) और प्रो. अम्बा कुलकर्णी (आचार्या, संस्कृत अध्ययन विभाग, हैदराबाद विश्वविद्यालय) का "Indian Perspective on Pronominal Anaphora Resolution" नामक लेख उशती पत्रिका के 22वे अंक में प्रकाशन हेतु स्वीकृत है।

अपराजिता मिश्रा

(प्रो. अपराजिता मिश्रा)
सम्पादक

Certificate of attendance

This is to confirm that:

SAEE VAZE

has participated in the

7th ISCLS in Auroville

14-17th February 2024

Auroville, India

17 February 2024 — *Dated this*

Martin Gluckman — *for Sanskrit Research Institute*

Inter Sentential Discourse Relations

Saee Vaze
Department of Sanskrit Studies,
University of Hyderabad
20hsph03@uohyd.ac.in

Amba Kulkarni
Department of Sanskrit Studies,
University of Hyderabad
ambakulkarni@uohyd.ac.in

Abstract

In this paper, we present a tagging scheme for inter-sentential discourse relations that is developed, based on the insights from Indian Grammatical Tradition. We rely on the three factors - ākāṅkṣā, yogyatā and sannidhi to decide the connectivity between two consecutive sentences. Various clues are identified that bind the two consecutive sentences. A tag set is presented based on the explicit discourse markers. Implementation of discourse level analysis based on the explicit discourse markers is tested on the *Śrīmadbhagvadgītā* corpus. It is observed that some discourse markers are ambiguous and it is not trivial to develop a disambiguation module for such markers.

1 Introduction

The term discourse analysis has gained a lot of attention in the recent past. It typically refers to a linguistic unit that goes beyond a sentence.[1] Thus, the discourse analysis goes beyond the scope of sentence boundaries and looks at the text as a unit of language. In understanding the meaning of a discourse, both the linguistic and non-linguistic factors contribute. The linguistic factors include coherence markers while non-linguistic background includes, speaker-listener dynamics, situationality etc. In Natural Language Processing, the core interest is in producing computer-processable models of discourse at different levels such as sentence, paragraph, text, etc. Varied work has been done on the topic already on different levels and languages.

From the theoretical point of view the work by Indologists and Sanskrit scholars in the field of discourse analysis is very rich and valuable. Scharf and Hock (2015) provides an exhaustive bibliography of works in the field of general discourse and formal syntax. However, there is very little work from the perspective of computational linguistics with regards to Sanskrit language. There are several efforts in the West especially in the field of computational linguistics. Treatment of cohesion by Halliday and Hasan (1976) attempts to look at the text as a linguistic phenomenon. Rhetorical Structure Theory (RST) (Mann and Thompson, 1988) was the first effort towards establishing the discourse structure in the form of a graph, by connecting two adjacent units by a discourse relation. Another seminal effort was made by the team lead by Arvind Joshi in the project Penn Discourse Tree-Bank (Prasad et al., 2006) which focuses on the structure of arguments and how a connective enables a certain discourse relation, implicit or explicit. The discourse tree-banks were created from a huge data from Wall Street Journal (Mann and Thompson, 1988) following the RST framework. The other discourse databanks include Linguistic Discourse Model (Polanyi, 2008), and the Discourse Graphbank (Wolf and Gibson, 2005). The Discourse-Lexicalized Tree Adjoining Grammar (Webber and Joshi, 1998) was developed following the Penn Discourse Treebank guidelines. With the emergence of such computational guidelines and resources for several languages discourse tagged datasets were developed for languages other than English such as Czech (Mladová et al., 2008), Chinese (Jiang et al., 2018) and Turkish (Zeyrek et al., 2010) to name a few. Similar efforts were made

[1] https://www.merriam-webster.com/dictionary/discourse


समीक्षा-9
Sponsored by:
Dept. of Science and Technology, Govt. of India
Organized by
Samskriti Foundation®, Mysore
www.samskritifoundation.org
samIkSha-n
National Seminar On Sanskrit-IKS & ICT
(Information & Communication Tools)
CERTIFICATE
This is to certify that
Sri / Smt / Dr / Prof. Ms. Saee Chandrashekar Daze
has Participated / been a Resource person / Chaired a Technical Session /Presented a valued Paper
Titled Inter-Sentential Discourse Relations: An Overview - IGT Perspective
In the 2day National Seminar on Sanskrit-IKS & ICT (Information & Communication Tools)
held, on 17th & 18th February 2023 at the National Institute of Advanced Studies (NIAS), Bangalore
एस. विद्याश्री :
Dr. S. Vidyashree
Deputy Director
Samskriti Foundation, Mysore

Indian Traditional Perspective on Discourse Analysis

by Saee Vaze

Submission date: 15-Dec-2025 10:21AM (UTC+0530)
Submission ID: 2846471888
File name: Vaze_Saee_Chandrashekhar.pdf (1.91M)
Word count: 41505
Character count: 247881

Indian Traditional Perspective on Discourse Analysis

ORIGINALITY REPORT

0%	0%	0%	0%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Middle East Technical University Student Paper	<1%
2	aclanthology.org Internet Source	<1%
3	ufal.mff.cuni.cz Internet Source	<1%

Exclude quotes	On	Exclude matches	< 14 words
Exclude bibliography	On		