A Mīmāṃsā-Based Framework for Disambiguating Compounds in Sanskrit

Devulapalli Koushik Suraj
Humanities and Social Sciences
Indian Institute of Technology
Roorkee, India
koushik.sd@hs.iitr.ac.in

Pavankumar Satuluri
Humanities and Social Sciences
Indian Institute of Technology
Roorkee, India
pavankumar.satuluri@hs.iitr.ac.in

Abstract

Compounds in Sanskrit (*samāsa*) often admit multiple grammatically valid internal analysis, each yielding a distinct meaning. While such multiplicity is valued in poetic contexts, technical and *śāstric* texts generally intend a single interpretation. For this, principles which can disambiguate and give a single meaning are essential. *Pūrva Mīmāṃsā* addresses this problem through interpretive maxims (*nyāyas*), among which the *Niṣādasthapati-nyāya* plays a role. This paper examines the theoretical basis of this nyāya and its relevance to resolving ambiguities in compounds (*samāsa*).

The *Niṣādasthapati-nyāya* prioritises interpretations that convey meaning through primary meaning (*abhidhā*) over those requiring secondary meaning (*lakṣaṇā*), provided no grammatical or contextual obstruction is present. Building on this principle, the paper proposes a structured framework that integrates grammatical filtering, identification of *lakṣaṇā* dependence, and contextual evaluation through *yogyatā*, *viśeṣya–viśeṣaṇa* relations, and *tātparya-liṅgas*. The study further outlines how this Mīmāṃsā-based framework can support computational tasks such as Word Sense Disambiguation and sentence interpretation.

1 Introduction

The Sanskrit language contains words whose meanings can be derived from their constituents (*avayavas*). All four *vṛttis* define meaning through such constituents, a process commonly referred to as etymological meaning (*yaugikārtha*). Among these, compound (*samāsa*) occupies a special position, since a single compound word (*samastapada*) may admit multiple internal analysis based on its *avayavas*. As a result, the same compound word can potentially yield different meanings depending on the *samāsa* analysis adopted. While this characteristic constitutes one of the distinctive strengths of Sanskrit, it also gives rise to a fundamental interpretive question: which meaning is intended in a given context? The Pūrva Mīmāṃsā tradition addresses this problem through specific interpretive principles (*nyāyas*), among which the *Niṣādasthapati-nyāya* plays a crucial role. This paper examines this *nyāya* and explains how it guides the selection of the appropriate *samāsa* when multiple analysis are possible. Compound words (*Samasta padas*) can be formed through multiple types of compounds (*samāsa*); consequently, a single compound word may admit several compound (*samāsa*) analysis, each yielding a distinct meaning. This feature has been extensively exploited in Sanskrit and Indian poetic traditions, where poets deliberately employ such multiplicity of meanings through *śleṣa*. In these literary contexts, multiple meanings are contextually fixed and collectively regarded as a poetic virtue (Ingalls, 1965). However, in many non-poetic and *śāstric* contexts, only one meaning is intended, while other possible interpretations are unintended. This raises the issue of on what basis one samāsa interpretation should be accepted while others are rejected.(Bronkhorst, 2001). Although compound words (*samasta padas*), like other *vṛttis*, derive meaning from their constituents, the interpretive challenge arises not from derivability itself but from the coexistence of multiple grammatically

valid analysis for a single compound. When more than one *samāsa* interpretation is grammatically possible, purely formal criteria are often insufficient to determine the intended meaning. In such cases, an additional interpretive framework becomes necessary. The Pūrva Mīmāṃsā tradition offers systematic solutions precisely to such problems of meaning determination (Jha, 1916; Verpoorten, 1987).

To address interpretive ambiguity in authoritative texts, Maharṣi *Jaimini* composed the *Pūrva Mīmāṃsā Śāstra*, within which a structured system of interpretive maxims (*nyāyas*) was developed. The *Pūrva Mīmāṃsā Śāstra* is traditionally known as the *Dvādaśalakṣaṇī*, as it consists of twelve chapters (*dvādaśa* meaning twelve, and *lakṣaṇa* meaning defining topic). These interpretive principles were formulated through close analysis of authoritative texts such as the Vedas, which are regarded within the Indian tradition as *pramāṇika* and free from error. Since the Vedas are considered unauthored and infallible, deriving interpretive principles from them was regarded as methodologically sound. Several of these maxims are sufficiently general to assist in the interpretation of texts beyond the Vedic corpus, thereby contributing to a systematic and consistent approach to linguistic interpretation (Jha, 1916; Clooney, 1990). For this reason, *Mīmāṃsāśāstra* is often described as the science of knowledge, the science of the study and interpretation of Vedic sentences, and the science of sentences. Despite its philosophical sophistication, it has not received the attention it deserves, largely due to the difficulty of mastering Vedic literature and the rigorous engagement required with both scriptural exegesis and ritual theory. The central aim of the Śāstra is to facilitate accurate interpretation and proper understanding of complex textual expressions (Kane, 1941; Clooney, 1990).Many core challenges related to sentence interpretation are systematically addressed within the *Mīmāṃsā* framework. By resolving sentence-level interpretive problems, the *Mīmāṃsā Śāstra* establishes a foundational methodology for understanding complex linguistic structures. In the modern era, with the rapid development of artificial intelligence, enabling machines to understand natural language has become increasingly important. Applications in natural language processing (NLP), such as machine translation and subtasks like Word Sense Disambiguation (WSD), have advanced significantly, and Sanskrit computational linguistics has also emerged as an active area of research (Hellwig, 2010; Kulkarni, 2012).

Interestingly, many interpretive challenges encountered in modern NLP closely parallel the problems addressed in the *Mīmāṃsā Śāstra*. Several *Mīmāṃsā nyāyas* can be directly applied to tasks such as WSD and machine translation in Sanskrit. The strength of these principles lies in their generality, as certain maxims are broad enough to support computational models across languages and domains. Applying these interpretive principles therefore offers a promising methodological bridge between traditional Indian linguistic philosophy and contemporary computational approaches to language understanding (Goyal, 2012; Kulkarni,2013).

Accordingly, this paper focuses on compound words and the ambiguity that arises from the availability of multiple *samāsa* interpretations for a single *samasta pada*. Such ambiguities are frequently encountered in the interpretation of Vedic and *śāstric* texts (Deshpande, 1991; Bronkhorst, 2001). Using examples from Vedic literature, this paper examines how such ambiguities are resolved within the framework of *Mīmāṃsā Śāstra,* explains the relevant *nyāyas*, and analyzes how the intended meaning is determined (Jha, 1916). Finally, it explores how these principles can be applied to contemporary Word Sense Disambiguation (WSD) tasks in computational linguistics (Goyal, 2012; Kulkarni, 2012).

2 Motivation and Research Gap

Substantial progress has been achieved in the computational processing of Sanskrit over the past two decades, particularly in foundational tasks such as sandhi splitting, morphological analysis, and *samāsa* generation (Huet, 2009; Kulkarni, 2013). Several rule-based and hybrid systems are capable of generating compound words in accordance with *Pāṇinian* grammatical

principles, often with user-specified inputs such as the intended *samāsa* type. These tools have proved effective for demonstrating grammatical derivations and for pedagogical purposes, where the user already possesses sufficient linguistic competence to guide the system. However, most existing approaches focus primarily on compound analysis rather than compound interpretation. They typically presuppose that the correct compound (*samāsa*) analysis is known in advance or explicitly supplied by the user. While this assumption is acceptable in controlled or educational settings, it becomes a serious limitation in higher-level natural language processing tasks such as sentence translation, *anvaya* construction, semantic parsing, where no prior human intervention can be assumed.

In naturally occurring Sanskrit texts particularly in Vedic, śāstric, and poetic literature *samasta padas* frequently admit multiple grammatically valid *samāsa* analysis, each yielding a distinct semantic interpretation. The intended meaning is often determined not at the word level, but through sentence-level semantics, the interaction between the compound and the *kriyā-pada*, *viśeṣya–viśeṣaṇa* relations, and broader indicators of speaker intention (*tātparya*). Consequently, merely enumerating all possible *samāsa* analysis is insufficient for tasks that require reliable semantic understanding; what is required is a principled mechanism for selecting the contextually intended interpretation. This limitation becomes particularly evident in contemporary computational processing of Sanskrit. Although such systems can handle large amounts of text at a surface level, they lack access to classical interpretive principles. As a result, *samāsa* ambiguity often leads to incorrect interpretations, unstable semantic representations, or inconsistent translations. In the absence of a structured interpretive framework, these systems tend to rely on surface-level patterns rather than linguistically and philosophically grounded reasoning.

2.1 Research Gap

Despite the availability of computational tools for compound (*samāsa*) generation and morphological analysis, there remains a significant gap in existing research with respect to the autonomous interpretation of compound words (*samasta padas*) at the sentence level. Current systems either assume prior knowledge of the intended compound (*samāsa*) or depend on explicit user intervention, thereby limiting their applicability to advanced NLP tasks such as machine translation, and *anvaya* construction. At the same time, although *Mīmāṃsā nyāyas* have been extensively discussed within traditional hermeneutics as principles for resolving semantic ambiguities (Jha, 1916; Clooney, 1990), their systematic formalization and application to disambiguate the compound words computationally remain largely unexplored. The interpretive insights developed in *Pūrva Mīmāṃsā* particularly those addressing ambiguity arising from multiple valid grammatical analysis have not yet been adequately integrated into contemporary computational frameworks.

The present work is motivated by this gap between grammatical generation and semantic interpretation. Drawing on the interpretive framework of Pūrva Mīmāṃsā, and in particular the *Niṣādasthapati-nyāya*, this paper proposes a decision-oriented framework for resolving ambiguity in compounds (*samāsa*) without relying on user intervention. By formalizing classical *nyāyas* as explicit interpretive principles, the proposed approach aims to bridge traditional Sanskrit hermeneutics and the requirements of modern computational linguistics. Thus, while existing *samāsa* generators are effective for producing compound forms, they are inadequate for tasks that require sentence-level interpretation and reliable semantic disambiguation. Addressing this limitation is crucial for the progress of Sanskrit NLP, particularly in applications such as machine translation and word sense disambiguation. This paper contributes toward this goal by showing how Mīmāṃsā-based interpretive principles can be systematically applied to *samāsa* interpretation in computational contexts.

3 Samāsa and the Problem of Interpretive Ambiguity

3.1 Pada: Semantic and Grammatical Definitions

In Sanskrit, the notion of *padam* is understood in two distinct but complementary ways: *śaktam padam* and *subantam–tiṅantam padam*. From a semantic perspective, *padam* is conceived as *śaktam padam*, that is, a linguistic unit endowed with meaning through *śakti* (specifically, *abhidhā-śakti*). In this sense, a word primarily functions as a bearer of meaning, and semantic interpretation takes precedence. At the grammatical level, however, particularly within the Pāṇinian tradition, *padam* is defined structurally as either *subanta* or *tiṅanta*, namely, a form ending in *sup* or *tiṅ* affixes (Cardona, 1997). This definition is central to the analysis of word formation and syntactic relations.

Since the discussion of *samāsa* belongs primarily to the grammatical domain, *padam* is understood here in the sense of *subantam* and *tiṅantam*. In this context, *samāsa* is traditionally defined as *anekapadānāṃ ekapadīkaraṇaṃ samāsaḥ* [1], that is, the process by which multiple words are combined into a single word. This well-established definition highlights the structural transformation involved in compound formation (Roodbergen, 1991). For example, in the expression *sītāyāḥ patiḥ*, two independent words occur. Through the process of *samāsa*, these are combined into a single compound form, *sītāpatiḥ*, which is referred to as a *samastapadam*. Such compounding plays a central role in Sanskrit grammar and significantly enhances the expressive capacity of the language.

3.2 Samāsa Formation and Sources of Ambiguity

The formation of *samāsa* is governed by Pāṇinian grammatical rules and depends upon the *sāmarthya* (semantic compatibility) of the constituent *padārthas*. When these conditions are satisfied, multiple words may be combined into a single *samasta pada*. A characteristic feature of Sanskrit compounding, however, is that a single surface form may admit more than one valid *samāsa* analysis. In certain cases, the same compound can be analyzed as a *karmadhāraya*, *tatpuruṣa*, or *bahuvrīhi samāsa*. Occasionally, other types of *samāsa* may also be possible. This plurality of grammatical analysis gives rise to interpretive ambiguity.

In poetic and literary contexts, such ambiguity is often intentional. Through devices such as *śleṣa*, multiple meanings are simultaneously evoked and appreciated as a source of aesthetic richness. In such cases, the coexistence of meanings is not regarded as problematic but rather as a deliberate poetic strategy. In contrast, technical, ritual, and philosophical texts generally admit only a single intended meaning. In these contexts, alternative grammatical interpretations, although theoretically possible, must be excluded. This leads to an important interpretive question: on what grounds should one *samāsa* analysis be accepted while others are rejected?

3.3 Grammatical Determination of Samāsa

In many instances, the intended *samāsa* can be determined on the basis of grammatical indicators such as *liṅga* (gender), *svara* (accent), *samāsānta pratyayas*, and other morphological constraints. These features often provide reliable clues for identifying the correct type of compound. For example, consider the forms *adhanaḥ* and *adhanaṃ*. Since *dhana* is a *napuṃsaka* noun, the form *adhanaḥ* can be identified as a *bahuvrīhi samāsa*, yielding the meaning *na vidyate dhanaṃ yasya saḥ*, "one who has no wealth." Similarly, in the case of *avyayībhāva samāsa*, the resulting compound is typically *napuṃsaka*, as seen in *niṣkāmaṃ*, derived as *kāmanāyāḥ abhāvaḥ*, "absence of desire." By contrast, the form *niṣkāmaḥ* can be analyzed as a *Tatpuruṣasamāsa*, meaning *nirgataḥ kāmaḥ yasmāt saḥ*, "one whose desire has departed." In such examples, *liṅga* functions as a decisive factor in determining the intended *samāsa*. Similar determinations may be made using accentual patterns and specific suffixes associated with particular types of compounds (Cardona, 1997; Roodbergen, 1991).

[1] Samasa, p. 1

3.4 Limits of Grammatical Determination

Despite the usefulness of grammatical indicators, there are cases in which the intended *samāsa* cannot be conclusively determined on the basis of form, accent, or gender alone. In such cases, the same compound legitimately admits more than one grammatical interpretation. For instance, the compound *madhuradhvaniḥ* can be interpreted both as a *karmadhāraya* and as a *bahuvrīhi samāsa*. Similarly, the form *apuruṣaḥ* allows both a *tatpuruṣa* interpretation (*na puruṣaḥ*) and a *bahuvrīhi* interpretation (*na vidyate puruṣaḥ yasya saḥ*). In such cases, even *liṅga* fails to resolve the ambiguity. Here, the intended meaning is generally determined with reference to *tātparya-liṅga*, that is, indicators of speaker intention derived from contextual, semantic, and pragmatic considerations. While such cues are readily accessible to human interpreters, formalizing them within a computational framework poses considerable challenges.

3.5 Mīmāṃsā-Based Resolution of Samāsa Ambiguity

When grammatical analysis proves insufficient, the Mīmāṃsā tradition offers a set of systematic interpretive principles known as *nyāyas*. One such principle is the *Niṣādasthapati-nyāya*, which is applied in cases where multiple grammatical interpretations are possible but only one aligns with the intended purpose of the sentence or passage. These *nyāyas* were formulated through close analysis of authoritative texts, particularly the Vedas, which are regarded in the Indian tradition as an authoritative (pramāṇika) source of knowledge (*pramāṇika*) and free from error (Clooney, 1990). By invoking such principles, interpreters are able to move beyond surface-level grammatical analysis and arrive at the meaning that best fits the broader textual and ritual context. When even these principles fail to yield a decisive interpretation, final recourse is made to *tātparya-liṅgas* and overarching semantic considerations. This layered interpretive strategy, progressing from grammatical form to contextual intention and purposive reasoning, constitutes one of the distinctive strengths of the Mīmāṃsā approach. By situating the problem of *samāsa* ambiguity within this broader interpretive framework, the present study demonstrates how classical Mīmāṃsā principles can address long-standing issues in Sanskrit interpretation. At the same time, these principles offer valuable insights for contemporary computational tasks such as Word Sense Disambiguation and machine translation, where structurally similar ambiguities frequently arise.

4 Niṣādasthapati-nyāya and the Resolution of Samāsa Ambiguity

4.1 Source and Context of the Nyāya

The *Niṣādasthapati-nyāya* is discussed in the *Pūrva Mīmāṃsā Śāstra* under the *Niṣādasthapati-adhikaraṇa* (6.1.12). The discussion is grounded in the Vedic injunction:

etayā niṣādasthapatiṃ yājayet

"By this rite, one should make the *Niṣāda-sthapati* perform the sacrifice."

The compound *niṣādasthapati* admits multiple *samāsa* analysis and, consequently, multiple meanings. This semantic indeterminacy prompted the Mīmāṃsā tradition to formulate a principle for determining the intended interpretation when several grammatically valid possibilities coexist (Clooney, 1990).

4.2 Nature of the Ambiguity

The compound *niṣādasthapati* allows more than one legitimate *samāsa* interpretation. As a *bahuvrīhi samāsa*, it can be analyzed as *niṣādaḥ sthapatiḥ yasya saḥ*, meaning "one whose *sthapati* is a Niṣāda." This interpretation requires the use of *lakṣaṇā* at more than one level. As a *ṣaṣṭhī-tatpuruṣa samāsa*, it can be interpreted as *niṣādānāṃ sthapatiḥ*, meaning "the leader or chief of the Niṣādas." This interpretation proceeds without recourse to *lakṣaṇā*. A third possibility is the *karmadhāraya samāsa*, *niṣādaś ca asau sthapatiś ca*, meaning "a person who is both a Niṣāda and a *sthapati*." While grammatically possible, this interpretation is contextually

constrained. The compound is therefore an instance of *anekavigraha-sambhava*, that is, a form capable of yielding multiple meanings through different *samāsa* analysis.

4.3 Statement of the Nyāya

From the analysis of this case, Mīmāṃsā formulates the following interpretive principle:

When a compound word is capable of expressing multiple meanings through different *samāsa* analysis, the interpretation that conveys meaning without resorting to *lakṣaṇā* is to be preferred, provided there is no contextual obstruction (*bādhaka*). In brief, among multiple possible *samāsa* interpretations, the one that avoids secondary or forced meaning is to be accepted (Clooney, 1990).

4.4 Which *Samāsas* Require *Lakṣaṇā* and Which Do Not

A central issue in applying the *Niṣādasthapati-nyāya* to *samāsa* interpretation is determining which types of *samāsa* necessarily involve *lakṣaṇā* (secondary signification) and which yield meaning purely through *abhidhā* (direct signification). This distinction is crucial, since the *nyāya* explicitly gives preference to interpretations that do not require *lakṣaṇā*, unless such avoidance is blocked by grammatical or contextual constraints. According to Pūrva Mīmāṃsā and allied interpretive traditions, a principled distinction can be drawn among the major *samāsa* types. In general, *karmadhāraya* and *dvandva samāsas* are regarded as *lakṣaṇā*-free, whereas *bahuvrīhi*, *tatpuruṣa*, and *avyayībhāva samāsas* characteristically involve *lakṣaṇā*. This observation gives rise to two related questions: (1) On what basis can we determine whether a given *samāsa* requires *lakṣaṇā*? (2) When multiple *lakṣaṇā* based *samāsas* are grammatically possible, how should ambiguity be resolved?

4.5 Basis for Identifying *Lakṣaṇā* in *Samāsa*

In *samāsa* interpretation, meaning is fundamentally derived from the *śakti* (primary semantic power) of the constituent words (*avayavas*). If the meaning of the compound can be fully obtained from the primary meanings of its constituents together with their syntactic relation, no appeal to *lakṣaṇā* is required. Conversely, when the compound conveys a meaning that cannot be compositionally derived from the *avayava* meanings alone, secondary signification becomes unavoidable. Thus, whenever the meaning of a *samasta-pada* exceeds, shifts away from, or is external to the literal meanings of its constituents, *lakṣaṇā* must be invoked. This criterion provides a consistent basis for classifying *samāsas* with respect to their semantic mode.

Bahuvrīhi Samāsa: *Bahuvrīhi samāsas* represent the clearest and most systematic use of *lakṣaṇā*. In such compounds, the referent is external to the compound itself, and the meaning cannot be obtained without semantic extension.

For example: *citragauḥ*: *citrāḥ gāvaḥ yasya saḥ*

Although the compound consists of *citra* and *go*, the intended meaning is not "a variegated cow," but "a person who possesses variegated cows." The referent is neither *citra* nor *go* but an external entity related to them through possession. This shift from a literal nominal meaning to a relational one necessarily involves *lakṣaṇā*. A similar mechanism operates in expressions such as: *ārūḍhavanaraḥ vṛkṣaḥ*. Here, the tree is described as "one that has been climbed by monkeys." The word *vanara* does not directly denote the tree; instead, it conveys an associated action or relation. Such indirect attribution is characteristic of *bahuvrīhi* and confirms its dependence on *lakṣaṇā*.

Tatpuruṣa Samāsa: *Tatpuruṣa samāsas* also generally involve *lakṣaṇā*, though in a more restrained and relational manner. Consider the compound: *rājapuruṣaḥ*, the compound consists of *rājan* and *puruṣa*. Its intended meaning is "a man connected with the king," such as a servant, official, or associate (*rājasambandhī puruṣaḥ*). This relational meaning is not directly expressed by the primary sense of *rājan* alone. Instead, the first member undergoes semantic extension

to convey an implicit relation. This reliance on relational inference indicates the operation of *lakṣaṇā*. While *tatpuruṣa* meanings remain internal to the compound, they often depend on unstated relational concepts, distinguishing them from *karmadhāraya* compounds, where such extension is unnecessary.

In the case of *dvandva samāsa*, *lakṣaṇā* does not operate. For example:

dhavakhadirau chindhi
("Cut the *dhava* and *khadira* trees.")

Here, the compound *dhavakhadirau* expresses a straightforward coordination of *dhava* and *khadira*. The meaning of the compound is completely obtained from the primary meanings of its constituent words (*avayavas*). No indirect or extended meaning is introduced by the compound as a whole. Consequently, *dvandva samāsas* are regarded as *lakṣaṇārahita*.

Similarly, *karmadhāraya samāsa* does not require *lakṣaṇā*. Consider the example:

nīlotpalam
("a blue lotus")

In this case, the compound meaning is directly derived from the primary meanings of *nīla* and *utpala*, which stand in *sāmānādhikaraṇya*. The compound neither exceeds nor departs from the literal meanings of its members. Since no semantic extension is required, *lakṣaṇā* does not arise in *karmadhāraya samāsa*.

In contrast, *avyayībhāva samāsa* characteristically involves *lakṣaṇā*. For example:

upakumbham
("something located near a pot")

Although *kumbha* appears as the *uttarapada*, the compound does not denote a pot itself. Instead, it conveys the meaning "that which is near or related to a pot" (*kumbhasambandhini*). This relational meaning cannot be obtained from the primary meaning of *kumbha* alone. Therefore, *lakṣaṇā* necessarily operates on the *uttarapada* in *avyayībhāva samāsa*.

On this basis, *dvandva* and *karmadhāraya samāsas* may be classified as *lakṣaṇā*-free, whereas *avyayībhāva samāsa* requires *lakṣaṇā*. Consequently, when *karmadhāraya* is among the competing analysis of an ambiguous compound, the decision is straightforward: since it does not involve *lakṣaṇā*, it is preferred under the *Niṣādasthapati-nyāya*. However, when ambiguity exists exclusively among *lakṣaṇā*-based *samāsas*, such as between *tatpuruṣa* and *bahuvrīhi*, the *nyāya* alone is insufficient to determine a unique interpretation. In such cases, further criteria must be invoked, including *yogyatā*, *viśeṣya–viśeṣaṇa* compatibility, and ultimately *tātparya-liṅgas*. The resolution of such cases is addressed in the subsequent discussion.

4.6 Implications for *Niṣādasthapati-nyāya*

From the perspective of the *Niṣādasthapati-nyāya*, these distinctions are decisive. When a compound admits both a *lakṣaṇā*-free interpretation (such as *karmadhāraya*) and a *lakṣaṇā*-based interpretation (such as *bahuvrīhi* or *tatpuruṣa*), the former is to be preferred, provided no grammatical or contextual obstruction (*bādhaka*) intervenes. When all available interpretations require *lakṣaṇā*, further resolution must proceed through contextual indicators and *tātparya-liṅgas*.

4.7 Conditions for the Application of the Nyāya

The *Niṣādasthapati-nyāya* is not universally applicable. Its operation presupposes specific conditions. First, the compound must be *anekasamāsa-sambhava*, that is, capable of admitting multiple *samāsa* analysis. Second, each interpretation must be grammatically well-formed. Third,

at least one interpretation must require *lakṣaṇā*, while another must convey meaning directly. Finally, there must be no overriding contextual contradiction (*asati bādhake*). Only when these conditions are jointly satisfied does the nyāya become operative.

4.8 Illustrative Examples

Consider the compound *āmravṛkṣaḥ*. This form admits both a *karmadhāraya* interpretation, *āmraḥ eva vṛkṣaḥ*, meaning "a mango tree," and a *tatpuruṣa* interpretation, *āmrasya vṛkṣaḥ*, meaning "a tree of mango." Although both analysis are grammatically possible, the *karmadhāraya* interpretation directly expresses the intended meaning without invoking *lakṣaṇā*. Accordingly, following the *Niṣādasthapati-nyāya*, it is preferred. A similar case is *adhanaḥ*. A *tatpuruṣa* analysis yields *na dhanaḥ*, "not wealth," which is semantically weak or contextually inappropriate. The *bahuvrīhi* analysis, *na vidyate dhanaṃ yasya saḥ*, meaning "one who has no wealth," provides a more natural and contextually suitable interpretation. In the absence of contextual obstruction, the latter is therefore preferred (Cardona, 1997).

4.9 Interpretive Significance

The *Niṣādasthapati-nyāya* establishes economy of interpretation as a decisive criterion in resolving semantic ambiguity. When multiple interpretations are available, *Mīmāṃsā* privileges the one that is grammatically sound, semantically direct, avoids unnecessary extension of meaning, and aligns with the intended purpose (*tātparya*) of the sentence. This principle enforces interpretive discipline and prevents arbitrary or excessively forced readings of textual expressions.

4.10 Relevance for Computational Frameworks

From a computational perspective, the *Niṣādasthapati-nyāya* can be modeled as a structured decision rule. All grammatically valid *samāsa* interpretations may first be generated. Among these, interpretations that require *lakṣaṇā* can be identified and deprioritized. Interpretations with direct semantic mapping are then preferred, with contextual or *tātparya*-based reasoning invoked only when necessary. Such a procedure offers an explainable and rule governed mechanism for resolving ambiguity. As a result, the *Niṣādasthapati-nyāya* provides a conceptually robust foundation for tasks such as Word Sense Disambiguation and compound interpretation in Sanskrit computational linguistics.

5 Framework for Disambiguating Samāsa

This section presents a principled framework for resolving ambiguity in Sanskrit compound words when multiple valid *samāsa* analysis are possible. The framework is inspired by classical *Mīmāṃsā* interpretive hierarchy and is designed to operate without user intervention, making it suitable for computational applications.

5.1 Input Representation

The input to the proposed framework consists of Sanskrit sentences containing one or more *samasta padas*. As a preliminary step, compound words are automatically identified within the sentence using standard morphological analysis techniques. Once a *samasta pada* is recognized, the system enumerates all grammatically valid *samāsa* analysis permitted by Pāṇinian grammar. At this stage, analysis that can be conclusively accepted or rejected on the basis of grammatical indicators such as *liṅga* agreement, *svara* constraints, and *samāsānta* markers—are filtered out. If this grammatical filtering yields a single surviving analysis, it is accepted as the intended interpretation without further processing. When more than one grammatically valid *samāsa* analysis remains, the compound is marked as ambiguous. Only under this condition is the *Niṣādasthapati-nyāya* invoked. The framework therefore applies this nyāya strictly in cases of genuine interpretive competition, where a single compound word can convey distinct meanings through different *samāsa* types. It is important to note that the present framework addresses ambiguity at the level of major *samāsa* categories such as *karmadhāraya*, *tatpuruṣa*,

and *bahuvrīhi*. Finer distinctions within a single *samāsa* class (for example, between different subtypes of *tatpuruṣa*) are beyond the scope of this study.

5.2 Decision Procedure

First, all grammatically permissible *samāsa* analysis are generated using rule-based systems grounded in Pāṇinian grammar. These include major categories such as *tatpuruṣa*, *karmadhāraya*, *bahuvrīhi*, *dvandva*, and *avyayībhāva*, subject to morphological compatibility.Analysis that violate grammatical constraints such as incompatible *liṅga*, inappropriate *samāsānta* markers, or prohibited *svara* patterns are eliminated. If this step results in a unique analysis, it is selected as the intended interpretation. When multiple analysis survive grammatical elimination, the framework evaluates the semantic mode through which each interpretation conveys meaning. Specifically, it determines whether the meaning is obtained through *abhidhā* (direct signification) or requires *lakṣaṇā* (secondary signification). At this stage, interpretations are classified but not immediately rejected.

In accordance with the *Niṣādasthapati-nyāya*, preference is given to interpretations that yield meaning through *abhidhā* alone, provided there is no grammatical or semantic obstruction (*bādhaka*). Interpretations requiring *lakṣaṇā* are deprioritised unless direct signification leads to incoherence or contradiction. If more than one *abhidhā*-based interpretation remains, the framework evaluates semantic compatibility with the surrounding sentence. Since *samasta padas* frequently function as modifiers, the analysis must support appropriate *viśeṣya–viśeṣaṇa* relations and maintain *sāmānādhikaraṇya*. Interpretations that fail to integrate with sentence-level semantics or result in *tātparyānupapatti* are rejected. Only when ambiguity persists after these evaluations does the framework rely on broader indicators of speaker intention (*tātparya-liṅgas*), such as compatibility with the *kriyā-pada*, overall sentence coherence, and discourse context. Because *tātparya* is inherently context-dependent, this stage is applied conservatively. In rare cases where no single interpretation can be conclusively selected, the compound is marked as context-dependent, reflecting the limits of rule-based resolution.

5.3 Mapping *Mīmāṃsā Nyāya* to Computational Logic

Although the framework is presented at a conceptual level, the underlying Mīmāṃsā principles align naturally with computational reasoning models. Within this mapping, a *nyāya* functions as a preference constraint that guides interpretation among competing analysis. Interpretations requiring *lakṣaṇā* can be understood as incurring higher semantic cost, as they involve indirect or extended meaning. A *bādhaka* functions as a hard constraint, eliminating analysis that result in grammatical or semantic contradiction. *Tātparya* operates as a soft contextual constraint, invoked only when stronger grammatical and semantic criteria fail to yield a unique result.

Viewed in this manner, the *Niṣādasthapati-nyāya* operates as a structured preference rule within a hierarchy of constraints, prioritising interpretations that minimise semantic indirection while preserving contextual coherence. This alignment enables the integration of classical *Mīmāṃsā* reasoning into modern Sanskrit NLP pipelines such as sentence parsing, semantic disambiguation, and large language model–based reasoning without reducing interpretation to purely statistical heuristics.

5.4 Limits of Nyāya Application and the Role of Tātparya-liṅgas

A natural question arises in the application of Mīmāṃsā nyāyas to *samāsa* disambiguation: how should conflicts be handled in cases where a nyāya is not applicable? This question leads to a broader clarification of the role of nyāyas in resolving linguistic ambiguity. Mīmāṃsā nyāyas are not rigid rules that operate mechanically in every context. Rather, they function as interpretive maxims derived from repeated patterns of successful meaning determination in authoritative texts. Just as grammatical rules describe regular linguistic behavior rather than enforce meaning externally, nyāyas capture the general reasoning strategies employed by

competent language users. Their application therefore presupposes contextual judgment and interpretive sensitivity. In the case of the *Niṣādasthapati-nyāya*, the preference for *lakṣaṇārahita* interpretations is not an absolute rule but a conditional guideline.

It reflects the observation that, in comparable cases, interpretations that rely on direct signification (*abhidhā*) are preferred unless constrained otherwise. Consequently, situations may arise where this nyāya cannot yield a decisive interpretation. In such cases, interpretation proceeds through *tātparya-liṅgas*. The *Pūrva Mīmāṃsā Śāstra* enumerates six *tātparya-liṅgas*, which serve as indicators for determining the intended meaning:

upakrama-upasaṃhārau abhyāso'pūrvata phalam|
arthavāda-upapattī ca liṅgaṃ tātparya-nirṇaye[2] ||

These indicators function in a manner similar to nyāyas, as they reflect the general interpretive reasoning employed by human readers. However, not all *tātparya-liṅgas* are relevant for compound disambiguation. In the context of *samāsa* ambiguity, two indicators are especially useful: *abhyāsa* (repetition) and *upakrama–upasaṃhāra* (beginning and conclusion). Both may be understood broadly as aspects of contextual coherence, including sentence-level and paragraph-level context.

Three major situations of ambiguity may be identified. First, ambiguity between *bahuvrīhi* and *tatpuruṣa*, both of which involve *lakṣaṇā*. Second, cases where a *lakṣaṇārahita* interpretation is possible, but no clear *tātparya-liṅga* is available. Third, situations where both a nyāya and one or more *tātparya-liṅgas* are applicable simultaneously. In the first case, disambiguation relies on contextual indicators such as the governing verb (*dhātu*), the qualified noun (*viśeṣya*), or another syntactically related word (*anyat padam*). For example, in *pītāmbaraṃ dharati*, the verb *dharati* indicates an object that can be worn, supporting a *karmadhāraya* interpretation. In *pītāmbaraṃ viṣṇuṃ paśyati*, the noun *viṣṇum* functions as the *viśeṣya*, favoring a *bahuvrīhi* interpretation. Similarly, in *pītāmbarasya śaṅkhaḥ asti*, the word *śaṅkhaḥ* serves as an external indicator, guiding interpretation.

These elements function as *tātparya-liṅgas*. In the second case, where a *lakṣaṇārahita* interpretation exists and contextual indicators are minimal, the nyāya itself may suffice. For instance, in *sarasi nīlotpalam paśya*, the compound *nīlotpalam* naturally admits a *karmadhāraya* analysis. The locative *sarasi* supports this reading but does not override the nyāya; rather, it confirms the preference for a non-*lakṣaṇā* interpretation. In rare cases where no clear *tātparya-liṅga* is available, interpretation may proceed solely on the basis of the nyāya. Finally, when both a nyāya and relevant *tātparya-liṅgas* are applicable, they operate jointly. The nyāya provides the initial preference, while the *tātparya-liṅgas* confirm and strengthen the chosen interpretation. This layered approach reflects actual human interpretive practice and enables ambiguity to be resolved in the majority of compound constructions.

[2] Vedāntasāra, p. 109.

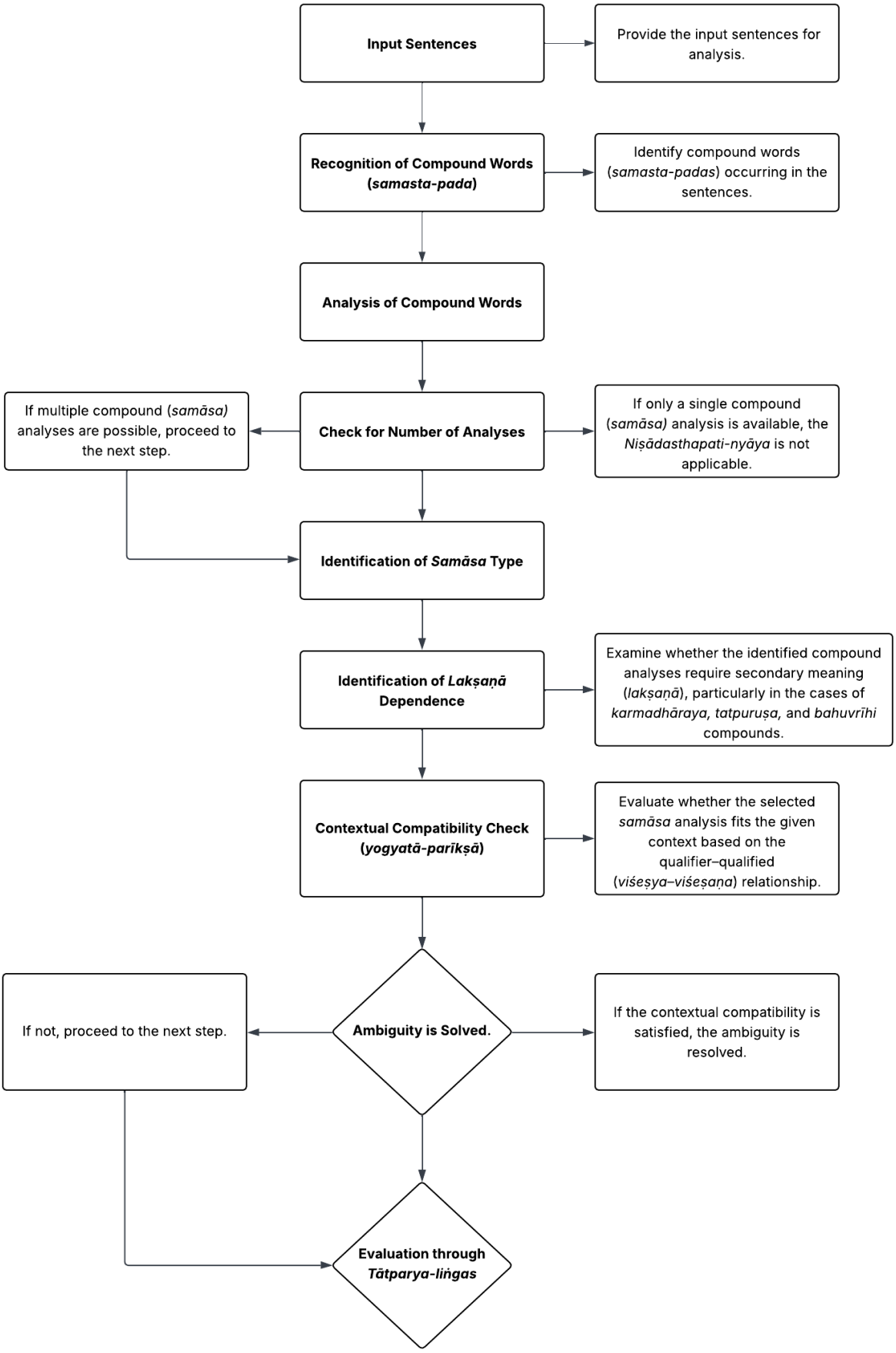

Figure 1: Methodology flowchart

6 Conclusion and Future Work

This paper proposed a principled framework for resolving *samāsa* ambiguity in Sanskrit by drawing on the *Niṣādasthapati-nyāya* of *Mīmāṃsā*. While existing computational approaches are capable of generating multiple grammatically valid *samāsa* analysis, they generally lack reliable mechanisms for selecting the intended interpretation in context. The framework presented here addresses this limitation by introducing a structured decision procedure that prioritizes interpretations yielding meaning through *abhidhā* (direct signification) over those requiring *lakṣaṇā* (secondary signification), subject to grammatical validity and the absence of contextual obstruction.

The proposed approach integrates multiple layers of interpretation, including grammatical filtering, identification of *lakṣaṇā* dependence, and contextual evaluation through *yogyatā*, *viśeṣya–viśeṣaṇa* relations, and *tātparya-liṅgas*. By organizing these factors hierarchically, the framework closely mirrors the interpretive priorities articulated in classical *Mīmāṃsā* while remaining compatible with computational reasoning. A central contribution of this study lies in clarifying the *Niṣādasthapati-nyāya* as a conditional preference rule rather than a universally applicable principle, thereby enabling its systematic and restrained use in ambiguity resolution.

Future work may extend this framework in several directions. One avenue is the treatment of finer distinctions within individual *samāsa* categories, such as conflicts among different subtypes of *tatpuruṣa*. Another promising direction is the incorporation of discourse-level and pragmatic information, which may further refine *tātparya* based interpretation. Finally, implementing the proposed framework within computational systems will allow empirical evaluation of its effectiveness in tasks such as Sanskrit parsing, semantic interpretation, and word sense disambiguation. More broadly, the present study suggests that *Mīmāṃsā*-based interpretive principles offer a rich and largely untapped resource for advancing Sanskrit natural language processing.

References

Sri G. Mahabaleshwara Bhatt. 2006. *Samāsaḥ Saṃskṛtabhāratī*.

Johannes Bronkhorst. 2001. *Language and Reality: On an Episode in Indian Thought*. Brill, Leiden.

George Cardona. 1997. *Pāṇini: His Work and Its Traditions*, volume 1. Motilal Banarsidass, Delhi.

Andrea Ciabattoni, Björn Lellmann, and Elisa Freschi. 2019. A deontic logic for mīmāṃsā. *Journal of Philosophical Logic*, 48(6):1127–1160.

Andrea Ciabattoni, Janneke Dik, and Elisa Freschi. 2025. Mīmāṃsā on "better-not" permissions. *Journal of Philosophical Logic*, 54(3):651–689.

Francis X. Clooney. 1990. *Thinking Ritually: Rediscovering the Pūrva Mīmāṃsā of Jaimini*. De Nobili Research Library, Vienna.

Madhav M. Deshpande. 1991. *Sanskrit and Prakrit: Sociolinguistic Issues*. Motilal Banarsidass, Delhi.

Pawan Goyal, Gérard Huet, Amba Kulkarni, and Shripad Sarma. 2012. A natural language processing framework for sanskrit. In *Proceedings of the 24th International Conference on Computational Linguistics (COLING 2012)*, pages 25–34, Mumbai.

Oliver Hellwig. 2010. Design and development of a sanskrit treebank. In *Proceedings of the Fifth International Sanskrit Computational Linguistics Symposium*, pages 161–170, Kyoto.

Oliver Hellwig. 2015. Sanskrittagger: A pos tagger for sanskrit. In *Proceedings of the 12th International Conference on Natural Language Processing (ICON)*, pages 252–259.

M. Hiriyanna. 1952. *Outlines of Indian Philosophy*. Allen & Unwin, London.

Gérard Huet. 2009. Sanskrit computational linguistics: A survey. *Journal of Indian Philosophy*, 37(4):403–428.

Daniel H. H. Ingalls. 1965. *An Anthology of Sanskrit Court Poetry (Vidyākara's Subhāṣitaratnakoṣa)*. Harvard University Press, Cambridge, MA.

Ganganatha Jha. 1916. *The Pūrva Mīmāṃsā Sūtras of Jaimini*. Panini Office, Allahabad. Translation.

Shivram D. Joshi and Johannes A. F. Roodbergen. 1991. *Patañjali's Vyākaraṇa-Mahābhāṣya*. Sahitya Akademi, New Delhi.

Pandurang Vaman Kane. 1941. *History of Dharmaśāstra*, volume 1. Bhandarkar Oriental Research Institute, Poona.

Amba Kulkarni, Soma Paul, Malhar Kulkarni, Anil Kumar, and Nitesh Surtani. 2012. Semantic processing of compounds in indian languages. In *Proceedings of COLING 2012*, pages 1489–1502, Mumbai, India. The COLING 2012 Organizing Committee.

Anil Kumar, Vipul Mittal, and Amba Kulkarni. 2010. Sanskrit compound processor. In *Proceedings of the International Symposium on Sanskrit Computational Linguistics*, Lecture Notes in Computer Science, pages 57–69. Springer. Available: `https://sanskrit.uohyd.ac.in/faculty/amba/PUBLICATIONS/papers/scl_anil.pdf`.

S. Nehrdich, O. Hellwig, and K. Keutzer. 2024. By t5-sanskrit: A unified model for sanskrit nlp tasks. *arXiv preprint arXiv:2409.13920*.

S. B. Pancoli. 2021. *Vaiyākaraṇabhūṣaṇasāraḥ*. Chaukhambha Sanskrit Sansthan, Varanasi.

Amba Kulkarni pavankumar Satuluri. 2013. Generation of sanskrit compounds. In *Proceedings of the International Conference on Sanskrit Computational Linguistics*, pages 19–28. AISC. Available at: `https://www.researchgate.net/profile/Amba-Kulkarni/publication/259655936_Generation_of_Sanskrit_Compounds/links/00b4952d2d36bc0602000000/Generation-of-Sanskrit-Compounds.pdf`.

Kunjunni Raja. 1969. *Indian Theories of Meaning*. The Adyar Library and Research Centre, Madras.

K. K. Raja. 1977. *Indian Theories of Meaning*. The Adyar Library and Research Centre, Madras.

P. Ramanujan. 1993. *Principles of Mīmāṃsā for Natural Language Processing*. Centre for Development of Advanced Computing (C-DAC), Pune.

Peter Scharf and Malcolm Hyman. 2009. *Linguistic Issues in Encoding Sanskrit*. Motilal Banarsidass, Delhi.

Jean-Marie Verpoorten. 1987. Mīmāṃsā hermeneutics and the interpretation of vedic sentences. *Indo-Iranian Journal*, 30:1–22.

Sadānanda Yogīndra. 1931. *Vedāntasāra of Sadānanda*. Advaita Āśrama, Almora.

Āpadeva. 1972. *Mīmāṃsā Nyāya Prakāśa: With Sarva Vivecinī Explanation*. Chowkhamba Sanskrit Series, Varanasi.

Dharmendranātha Śāstrī. 1953. *Nyāya Siddhānta Muktāvalī*. Rashtriya Sanskrit Sansthan.