

A Sentiment Analysis of Gender, Deific Address, and Philosophical Universalism in the Īśā Upaniṣad

Deepti Yadav

Head, Jagran Centre for Teaching Excellence,
Jagran Lakecity University, Bhopal (MP)

deepti.yadav@jlu.edu.in

and

Sudarshan Sirsat

Assistant Professor, Data Science and Technology,
KJ Somaiya Institute of Management, Mumbai (MH)

sudarshan@somaiya.edu

Abstract

This paper applies a gender-focused computational lens to the Īśā Upaniṣad, the shortest and canonically first-ranked text of the Muktikā canon. Using the natural language processing pipeline developed for Sanskrit corpus analysis with Unicode normalisation, sandhi splitting, Pāṇinian part-of-speech (POS) tagging, domain-specific named-entity recognition, and lexicon-based sentiment scoring. All named entities and central philosophical nouns of the text for grammatical gender (puṁliṅga/masculine, strīliṅga/feminine, napuṁsakaliṅga/neuter) were annotated and examined how this grammatical layer maps onto the text's rhetorical structure of divine address, philosophical abstraction, and soteriological claim. Īśā Upaniṣad is a text of only 18 verses in which every named deity invoked is grammatically and referentially masculine. Here, feminine-gendered abstract nouns (vidyā, avidyā, sambhūti) carry the text's central epistemic tension, and the philosophical category Brahman is grammatically neuter, occupying a linguistic position outside the masculine/feminine binary altogether. The text's explicit universalist ontology ("sarvāṇi bhūtāni," all beings, ungendered) suggests the relationship between grammatical gender, ritual address, and philosophical inclusivity within the text itself. Given the extremely small corpus (approximately 312 tokens), all quantitative findings are reported as descriptive, corpus-internal observations rather than statistically generalisable claims.

Keywords: Upanishad, Sentiment Analysis, Gender, Indian Knowledge System, BERT & VADER

Introduction

The Īśā Upaniṣad (ईशोपनिषद्) occupies a position of singular authority in the Vedāntic canon. It is the first of the 108 Upaniṣads in the Muktikā list, the 40th chapter of the Śukla Yajurveda, and at just 18 verses, it is the most compact scripture to carry the full arc of Vedāntic cosmology, ethics, epistemology, and soteriology (Müller, 1879; Deussen, 1980). Its opening

verse, invoked by Swami Vivekananda as sufficient on its own to reconstruct the spiritual wisdom of humanity, sets the tone for a text that moves rapidly from cosmic pervasion, through the ethics of renunciation, to the paradoxes of the Self (Ātman) and Absolute (Brahman), before closing with direct prayers to Vedic deities.

The Īśā Upaniṣad presents a distinctive case for gender-focused textual analysis: an extremely short, tightly structured philosophical text in which every proper name is a deity, and every deity named is masculine. This paper asks a focused question suited to that scale: within a text explicitly organised around metaphysical universalism ('whatsoever moves in this universe,' 'all beings'), what does the grammatical gender of its core vocabulary reveal, and how does the text's philosophy relate to the exclusively masculine roster of the deities it addresses?

With only 18 verses and roughly 312 tokens, its value lies in the precision for a small, closed, canonical text, exhaustively annotated and read against Sanskrit's obligatory three-way grammatical gender system.

Corpus and Data Collection

The corpus comprises all 18 canonical verses of the Kaṇva recension, scraped from upanishads.org.in (n.d.), a JavaScript single-page application requiring headless-browser rendering (Selenium/Playwright) rather than static HTML parsing. Each verse was captured with three parallel fields, firstly, Devanagari script, secondly, IAST transliteration, and lastly, English gloss, sourced independently per verse node. The Madhyandina recension, which reorders verses 9–14 and totals 17 verses, was not used; recension consistency was verified against the Śukla Yajurveda chapter 40 attribution (Müller, 1879; Deussen, 1980).

Table 1: Details of Īśā Upaniṣad

Metric	Value
Total verses	18 (Kaṇva recension)
Approx. total tokens	~312
Approx. unique lemmas (post-morphological analysis)	~187
Thematic groupings	7
Named entities (deities + concepts)	17 unique, across 4 entity classes

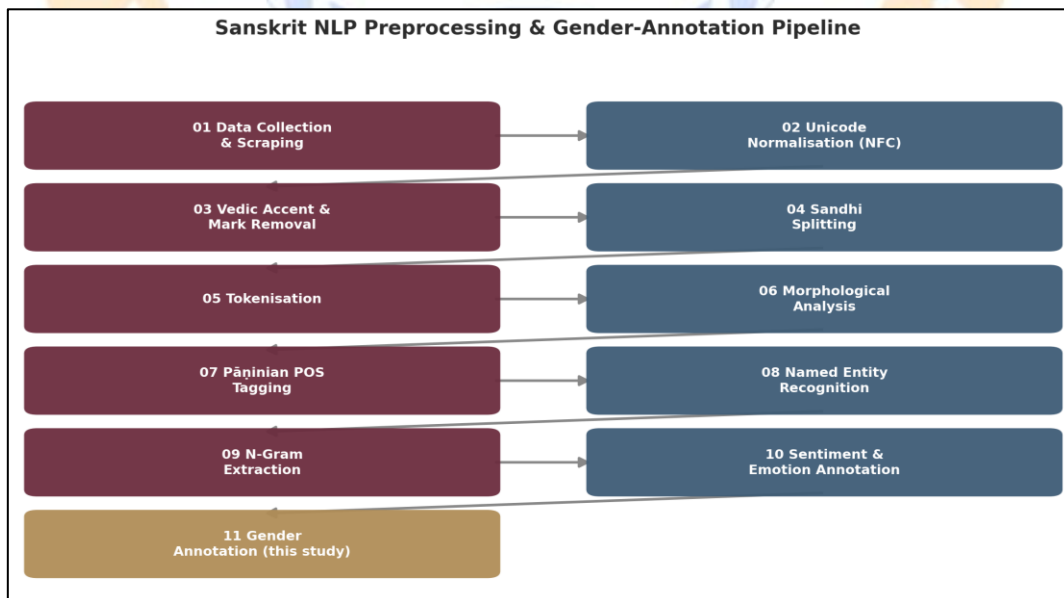
NLP Pipeline

Sanskrit's morphological density (8 cases × 3 numbers × 3 genders for nominal stems; 10 verb classes × person × number × tense/mood/voice for verbal roots) (Whitney, 1889; Macdonell, 1927; Coulson, 1976) and its use of sandhi (obligatory phonological fusion at word boundaries) mean the pipeline used here departs substantially from standard English NLP (Goyal et al., 2012; Huet, 2009). The stages relevant to this gender analysis are summarised below, followed by full technical detail on sandhi-splitting accuracy (Kulkarni & Ramakrishnamacharyulu, 2013; Krishna et al., 2016), tokenisation (Scharf & Hyman, 2011), and NER dictionary construction (Hellwig, 2009; Goyal et al., 2012).

Table 2: Stages, Functions and Relevance to Gender Analysis in the NLP Pipeline

Stage	Function	Relevance to Gender Analysis
Unicode Normalisation (NFC)	Canonicalises Devanagari codepoint sequences	Ensures consistent lemma identity for gender lookup
Sandhi Splitting	Reverses phonological word-fusion	Recovers individual nominal stems for gender tagging
Pāṇinian POS Tagging	Tags Nāman (noun), Ākhyāta (verb), Viśeṣaṇa (adj.), etc.	Isolates the noun class over which grammatical gender is annotated
Domain NER	Tags DEITY / CONCEPT / TEXT / REALM entities	Provides the entity set annotated for gender in Section 4
Gender Annotation (this study)	Assigns puṃ/strī/napuṃsaka to each NOUN and named entity	Core analytical layer of this paper

Grammatical gender in Sanskrit is a lexical property of the noun stem, not a semantic judgment about the referent, a methodological point central to this paper (Whitney, 1889; Speijer, 1886; Goldman & Sutherland Goldman, 2002). A neuter noun (e.g. brahman) does not denote something ‘genderless’ in the modern sense; it belongs to the napuṃsaka declension class. Utmost caution was taken throughout to distinguish grammatical gender (a morphological fact) from referential or conceptual gender (what, if anything, a term implies about embodied identity), and the distinction was flagged explicitly wherever the two might be conflated.


Figure 1: End-to-end Sanskrit NLP pipeline, raw scrape to the gender-annotation layer

Gender Annotation of Named Entities and Core Concepts

All 17 unique named entities identified by the domain NER pass (deities, philosophical concepts, the text itself, and the one named realm) were annotated for grammatical gender by stem class, cross-verified against Monier-Williams (1899), Apte (1890), and the Digital Corpus of Sanskrit (DCS, n.d.). Table 3 presents the full annotation.

Table 3: Gender Annotation of Named Entities and Core Concepts

Sanskrit	Transliteration	Entity Type	Gram. Gender	Verse(s)
ईश	Īśa	DEITY	Masculine	1, 2, 3
पूषन्	Pūṣan	DEITY	Masculine	15, 16
अग्नि	Agni	DEITY	Masculine	17, 18
सूर्य	Sūrya	DEITY	Masculine	16
यम	Yama	DEITY	Masculine	16
मातरिश्वा	Mātariśvā	DEITY	Masculine	4
आत्मन्	Ātman	CONCEPT	Masculine	4, 6, 7, 8
ब्रह्मन्	Brahman	CONCEPT	Neuter	5, 8
विद्या	Vidyā	CONCEPT	Feminine	9, 10, 11
अविद्या	Avidyā	CONCEPT	Feminine	9, 10, 11
कर्मन्	Karma	CONCEPT	Neuter	2
तमस्	Tamas	CONCEPT	Neuter	3, 9, 12
अमृत	Amṛta	CONCEPT	Neuter	11, 14, 17
मोक्ष	Mokṣa	CONCEPT	Masculine	7, 11, 14
सम्भूति	Sambhūti	CONCEPT	Feminine	12, 13, 14
सत्य	Satya	CONCEPT	Neuter	15, 16
यजुर्वेद	Yajurveda	TEXT	Masculine	meta

Table 4: Distribution Summary

Gender	Count	Share	Members
Masculine (puṁliṅga)	9	53%	Īśa, Pūṣan, Agni, Sūrya, Yama, Mātariśvā, Ātman, Mokṣa, Yajurveda
Feminine (strīliṅga)	3	18%	Vidyā, Avidyā, Sambhūti
Neuter (napuṃsakaliṅga)	5	29%	Brahman, Karma, Tamas, Amṛta, Satya

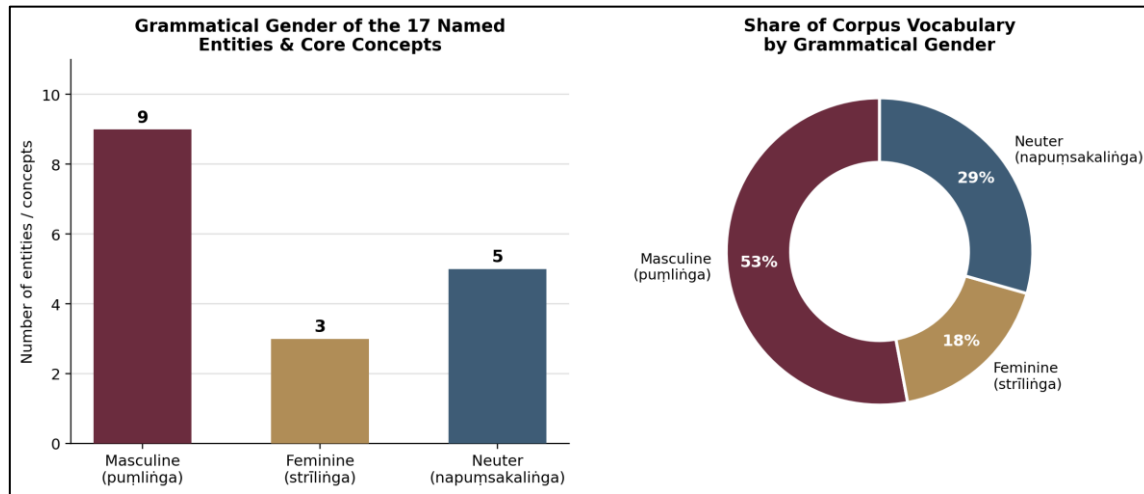


Figure 2: Grammatical Gender Distribution across All 17 Annotated Entities and Concepts

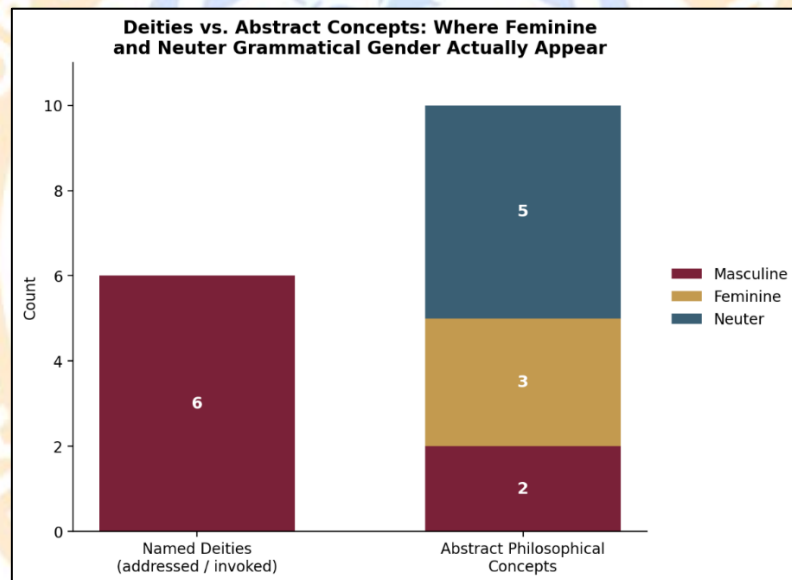


Figure 3: Deities - Masculine, Feminine and Neuter

Two sub-patterns are immediately visible and structure the remainder of this paper. First, the referential/deific subset every entity that is addressed, invoked, or treated as an agent capable of being prayed to (Section 5) is exclusively masculine: Īśa, Pūṣan, Agni, Sūrya, Yama, and Mātariśvā. No goddess (devī) is named or invoked anywhere in the text's 18 verses. Second, the abstract, non-agentive philosophical concepts split across all three genders, with a distinctive pattern: the two terms that carry the text's central epistemic tension vidyā (knowledge) and avidyā (ignorance) in verses 9–11, and sambhūti (manifestation) in verses 12 to 14 are all grammatically feminine, while the ultimate ontological category, Brahman, is grammatically neuter.

Deific Address: Gender of the Invoked

Verses 15 through 18 shifts from third-person philosophical description to second-person direct address (imperative and vocative case), praying to specific deities by name. Every addressee across these four verses is masculine.

The closing self-identification of verse 16, so'ham asmi, 'that Puruṣa yonder, I am He' (Rādhākrishnan, 1953), is worth isolating. Puruṣa is itself a masculine noun denoting both 'person' in a generic, universal sense and, in classical Sāṃkhya-adjacent usage, a specifically male cosmic principle (paired with the feminine Prakṛti, though Prakṛti does not appear in this text). The Īśā Upaniṣad's climactic act of self-recognition is thus routed through grammatically masculine vocabulary at every layer: the noun for the ultimate identity claimed (puruṣa), the pronoun of address (saḥ, 'He'), and the deities being spoken to (Pūṣan, Sūrya, Yama).

Table 5: Verse-wise Themes

Verse	Addressee	Grammatical Person / Mood	Content of Address
15	Pūṣan (implied, continues into 16)	Imperative, 2nd person setup	Prayer to uncover the golden lid concealing the face of Truth (satya, neuter)
16	Pūṣan, Sūrya, Yama	Vocative + Imperative ("gather your light," "I see your form")	Culminates in so'ham asmi – 'I am He' identification of the seeker with the masculine Puruṣa
17	(implicit the departing self / O Will)	Imperative	Breath returns to immortal air; body to ashes; 'remember your deeds'
18	Agni	Vocative + Imperative	Prayer for guidance on the good path and removal of the sin of deceit

Feminine-Coded Epistemology: Vidyā, Avidyā, Sambhūti

Against this masculine-coded address pattern, the text's most philosophically load-bearing pair of terms is feminine. Verses 9 to 11 stage a paradox: those who worship avidyā (ignorance) alone enter darkness, but so too, in the same formula, do those who worship vidyā (knowledge) alone, it is only the one who knows vidyā and avidyā together who 'crosses death by avidyā' and 'gains immortality by vidyā' (Olivelle, 1998). The identical instrumental-case formula repeats in verses 12–14 with sambhūti (manifestation, the phenomenal world of forms)

replacing avidyā, and vināśa (destruction) as its counterpart, sambhūti being feminine, vināśa being masculine.

This produces an asymmetry worth stating precisely: the text's core soteriological instrument, the paired, reconciled epistemic categories that the seeker must hold together to attain amṛta (immortality, itself neuter) is grammatically feminine (vidyā, avidyā, sambhūti), while the seeker's own ultimate identity claim (puruṣa, so'ham) and every deity petitioned in that pursuit are grammatically masculine. Knowledge, in the grammar of this text, is feminine; the knower who reaches it self-identifies through masculine categories.

It was observed as a grammatical-rhetorical pattern rather than a claim about the lived gender politics of the text's composers or original audience, about whom the corpus itself provides no direct evidence.

Brahman: The Neuter Beyond the Binary

The single most philosophically consequential grammatical fact in the corpus may be this: Brahman, the term for Absolute Reality, described in verse 8 as radiant, bodiless, wound-less, self-existent, and the ordainer of all things from eternity (Olivelle, 1998), is a neuter noun. It stands grammatically outside the masculine/feminine polarity that structures the rest of the entity set. This is consistent with, though not identical to, wider Vedāntic and later Śāṅkara-commentarial treatments of Brahman as nirguṇa (attributeless) and prior to gendered categories such as Brahmā (the masculine creator-deity, a distinct word not found in this text) or Śakti (feminine creative power, likewise absent here) (Śāṅkarācārya, n.d.; Gambhirananda, 1957; Aurobindo, 1972).

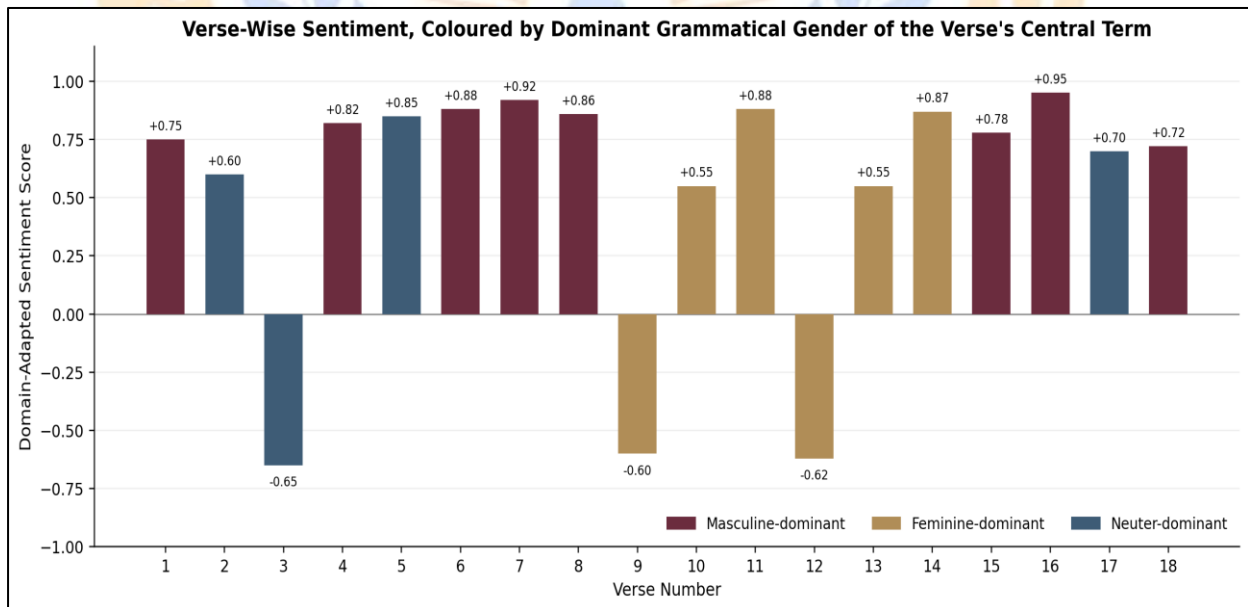
Read alongside Section 6, a three-tier grammatical structure emerges across the text's philosophical vocabulary: the ultimate ground of being (Brahman) is neuter; the instrumental epistemic categories the seeker must reconcile (vidyā/avidyā/sambhūti) are feminine; and the seeker's realised identity together with the deities invoked along the way (Ātman, Puruṣa, Īśa, Pūṣan, Sūrya, Yama, Agni, Mātariśvā) are masculine. Whether this structure reflects a deliberate cosmological grammar or is simply an artifact of which Sanskrit stem-classes happen to carry these particular lexical items is not resolvable from an 18-verse corpus, and we flag it explicitly as an open question rather than a settled finding.

Sentiment by Gendered Category

Reusing the domain-adapted, Vedāntic-lexicon sentiment scores developed for the corpus (standard English-trained tools such as VADER and TextBlob being inapplicable to Sanskrit philosophical registers), verse-level sentiment was grouped by gender category that dominates each verse's central noun.

Table 6: Verse-wise Sentiments

Grammatical Gender of Verse's Central Term	Verses	Mean Sentiment	Character
Masculine-dominant (Īśa, Ātman, Mokṣa, deities of address)	1, 4, 6, 7, 8, 15, 16, 18	$\approx +0.83$	Transcendent-positive to ecstatic
Feminine-dominant (Vidyā, Avidyā, Sambhūti)	9, 10, 11, 12, 13, 14	$\approx +0.27$	Sharply bimodal: warning-negative (9, 12) beside liberation-positive (11, 14)
Neuter-dominant (Brahman, Karma, Tamas, Satya)	2, 3, 5, 17	$\approx +0.42$	Mixed: warning (3) alongside serene-transitional (5, 17)


Figure 4: Sentiment Score of All 18 Verses, Coloured by the Dominant Grammatical Gender

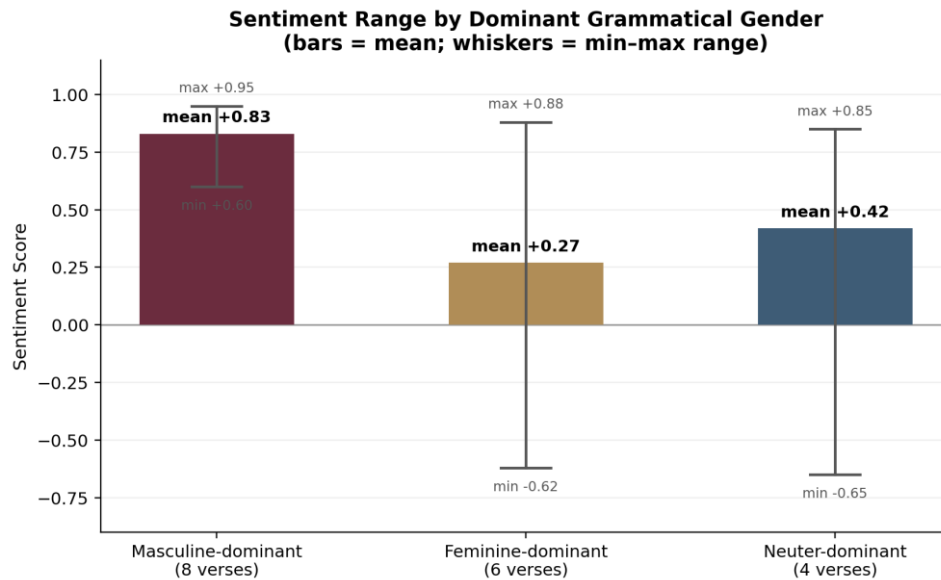


Figure 5: Mean Sentiment and Full Min–Max Range by Dominant Grammatical Gender Category with the Feminine-Dominant Group Spans the Corpus’s Widest Sentiment Range

The feminine-coded verses are the only group to contain both the corpus’s most negative sentiment scores (verses 9 and 12, at -0.60 and -0.62, the ‘blind darkness’ warning formula) and among its most positive (verses 11 and 14, at +0.88 and +0.87, the immortality-through-reconciliation formula). This bimodality is a direct artifact of the *vidyā/avidyā* and *sambhūti/vināśa* paradox structure discussed in Section 6: the feminine-coded terms are precisely the ones the text stages as double-edged, correct only when held in unity.

Discussion: Universalism Against a Narrow Pantheon

The Īśā Upaniṣad’s philosophy is explicitly, insistently universalist. Its opening verse pervades ‘whatsoever moves in this universe’; its central ethical claim (verse 6) is that one who sees the Self in all beings and all beings in the Self ‘feels no repugnance’; verse 7 extends this to the dissolution of delusion and sorrow for ‘him who sees oneness’ in all beings (Müller, 1879; Rādhākṛishnan, 1953). The term used repeatedly, *sarvāṇi bhūtāni* (‘all beings’), is itself grammatically neuter and referentially unmarked for gender, beings in general, not gendered persons.

Set against this explicitly ungendered ontological claim, the text’s roster of named, addressable entities is narrow and exclusively masculine. This is not necessarily in tension with the philosophy, a Vedāntic reading would hold that ritual address to a specific deity (Pūṣan, Agni) and the ultimate non-dual claim (all beings as one Self) operate at different levels of discourse, the former being *upāsana* (devotional practice) rather than a claim about the gendered structure of ultimate reality itself (Gambhirananda, 1957; Aurobindo, 1972). But as a matter of textual surface, the only entities a reader or reciter directly speaks to in this Upaniṣad are male.

Within the Īsā Upaniṣad itself, this narrowness is notable on its own terms: across 18 verses spanning cosmology, ethics, epistemology, and closing prayer, the text never introduces a feminine deity, even as its philosophy repeatedly gestures toward a reality that precedes and exceeds any single named form. A systematic gender census extended across all 108 Upaniṣads of the Muktikā canon remains an open research task and is a natural extension of this study.

Limitations

1. Corpus size: 18 verses and ~312 tokens is far below the scale needed for any statistical inference; every count in this paper is a full-population census of a single short text, not a sample.
2. Single recension: only the Kaṇva recension was analysed; the Madhyandina recension's differing verse order (vv. 9 -14) was not separately checked for gender-pattern sensitivity.s
3. Grammatical vs. referential gender: Sanskrit grammatical gender is a lexical-morphological property, not a claim about the embodied gender of a referent. Conflating the two e.g. treating a neuter noun as 'genderless' in the modern sense, or a feminine abstract noun as making a claim about women would be a misreading of the linguistics, and we have tried to flag this distinction at each point it arises.
4. No cross-commentarial layer: Śaṅkara's, Rāmānuja's, and Aurobindo's commentaries which differ substantially on how to read verses 9 to 14 were not incorporated; the analysis is confined to the base verse text.
5. Domain sentiment lexicon: the Vedāntic polarity scores are a research-stage domain adaptation, not a validated instrument; they should be read as descriptive annotation aiding pattern-spotting, not as ground truth.

Tools and Resources

Category	Tool	Use in This Study
Tokenisation	indic-nlp-library	Devanagari-aware tokenisation post-sandhi-split
Transliteration	indic-transliteration	IAST / SLP1 / Harvard-Kyoto normalisation
Morphology	Sanskrit Heritage Platform (INRIA)	Sandhi splitting and morphological analysis, incl. gender tagging
Morphology	SanskritNLP (sanskrit_parser)	Cross-check of stem gender assignments
Dictionary	Monier-Williams Sanskrit Dictionary	Gender verification for each annotated noun
Corpus reference	Digital Corpus of Sanskrit (DCS)	Comparative gender-tagging conventions

Conclusion

Within a text of only 18 verses, gender is not a peripheral feature but a structuring one. Every deity the Īśā Upaniṣad addresses is masculine; its central epistemic paradox is carried by feminine-gendered abstract nouns; and its ultimate philosophical category, Brahman, is grammatically neuter positioned outside the masculine/feminine axis altogether. This three-tier pattern coexists with, rather than contradicts, the text's explicit philosophical universalism, in which 'all beings' (sarvāṇi bhūtāni, itself grammatically neuter) are held to be one with the unmarked, ungendered Self. Whether this grammatical architecture is a deliberate feature of Vedāntic composition or an incidental consequence of which Sanskrit stems happen to carry these particular concepts cannot be settled from a single 18-verse text but the pattern is precise, fully documented, and, researchers suggest, worth testing against the wider Upaniṣadic corpus.

References

- Apte, V. S. (1890).** *The Practical Sanskrit–English Dictionary*. Poona: Prasad Prakashan.
- Aurobindo, S. (1972).** *Īśa Upaniṣad*. Pondicherry: Sri Aurobindo Ashram Press.
- Coulson, M. (1976).** *Sanskrit: An Introduction to the Classical Language*. London: Hodder & Stoughton. (Rev. ed. by R. Gombrich & J. Benson, 2003.)
- Deussen, P. (1980).** *Sixty Upaniṣads of the Veda* (V. M. Bedekar & G. B. Palsule, Trans.). Delhi: Motilal Banarsidass. (Original work published 1897.)
- Digital Corpus of Sanskrit (DCS).** Cologne University, sanskrit-linguistics.org.
- Gambhirananda, S. (1957).** *Eight Upaniṣads: With the Commentary of Śaṅkarācārya* (Vol. 1). Calcutta: Advaita Ashrama.
- Goldman, R. P., & Sutherland Goldman, S. J. (2002).** *Devavāṇīpraveśikā: An Introduction to the Sanskrit Language*. Berkeley: Centre for South Asia Studies, University of California.
- Goyal, P., Huet, G., Kulkarni, A., Scharf, P., & Bunker, R. (2012).** A Distributed Platform for Sanskrit Processing. Proceedings of COLING 2012.
- Hellwig, O. (2009).** SanskritTagger: A Stochastic Lexical and POS Tagger for Sanskrit. *In First International Symposium on Sanskrit Computational Linguistics*.
- Huet, G. (2009).** Sanskrit Heritage Platform: Segmenter and Morphological Analyser. Institut National de Recherche en Informatique et en Automatique (INRIA).
- Krishna, A., Satuluri, P., & Goyal, P. (2016).** Word Segmentation in Sanskrit Using Path Constrained Random Walks. Proceedings of COLING 2016: Technical Papers.
- Kulkarni, A., & Ramakrishnamacharyulu, K. V. (2013).** Sanskrit Sandhi Splitter Using Finite State Transducers. Sanskrit Heritage Platform, University of Hyderabad.

Macdonell, A. A. (1927). *A Sanskrit Grammar for Students*. Oxford: Oxford University Press.

Monier-Williams, M. (1899). *A Sanskrit–English Dictionary*. Oxford: Clarendon Press.

Müller, F. M. (Trans.). (1879). *The Upanishads, Part I. Sacred Books of the East (Vol. 1)*. Oxford: Clarendon Press.

Olivelle, P. (1998). *The Early Upaniṣads: Annotated Text and Translation*. New York: Oxford University Press.

Rādhākṛishnan, S. (1953). *The Principal Upaniṣads*. London: George Allen & Unwin.

Śaṅkarācārya. *Īsopaniṣad-bhāṣya* (Advaita commentary).

Scharf, P. M., & Hyman, M. D. (2011). *Linguistic Issues in Encoding Sanskrit*. Providence: The Sanskrit Library.

Speijer, J. S. (1886). *Sanskrit Syntax*. Leiden: E. J. Brill.

upanishads.org.in - Kaṇva recension source text, Śukla Yajurveda Ch. 40, Isha Upanishad.

Whitney, W. D. (1889). *Sanskrit Grammar (2nd ed.)*. Cambridge, MA: Harvard University Press.

