Advanced International Journal of Multidisciplinary Research
E-ISSN: 2584-0487
editor@aijmr.com
Volume 3, Issue 4, July-August 2025
CrossRef DOI: 10.62127/aijmr.2025.v03i04.1484

Selective Locality: Dependency-Length Minimization, Non-Projectivity, and the Ākāṅkṣā Account of Word Order in Vedic Sanskrit

Vishal Sharma

Research Scholar
Department of Vyakarana, Central Sanskrit University, Vedvyas Campus Balahar.

Abstract:
Sanskrit is traditionally described as a free-word-order language, and the indigenous grammatical tradition explains what constrains word combination through three notions: *ākāṅkṣā* (syntactic expectancy), *yogyatā* (semantic compatibility), and *sannidhi* (proximity or contiguity). This study asks whether *sannidhi*, read as a locality constraint, is empirically supported, and how it relates to the cross-linguistic principle of dependency-length minimization (DLM). Using the Universal Dependencies Treebank of Vedic Sanskrit (27,182 sentences; 206,440 tokens), I compared observed dependency lengths against a random-projective baseline and a minimal-arrangement heuristic, partitioned arcs by grammatical relation, quantified non-projectivity, and traced variation across the corpus's chronological layers. At the whole-sentence level, Vedic showed no dependency-length minimization beyond the projectivity constraint: observed mean length (1.996) was statistically indistinguishable from the random-projective baseline (2.002) and far above the minimal arrangement (1.512). This near-parity, however, masked a systematic relation-specific split. Core verb-argument relations—the *kāraka*-type expectancy relations—were placed reliably closer than their own random baseline (mean deviation −0.36 tokens, 95% CI [−0.38, −0.34]), whereas coordinate, appositional, and modifier relations were placed at or beyond chance distance (+0.16 tokens, 95% CI [+0.13, +0.18]). Non-projectivity was common (19.4% of sentences) but overwhelmingly mild and well-nested, and it declined sharply from the Ṛgvedic layer (35.9%) to the Sūtra layer (11.0%). An independent classical treebank reproduced both the aggregate parity and the non-projectivity rate. I argue that *sannidhi* is best understood not as global length optimization but as a selective, expectancy-scoped locality operating over *ākāṅkṣā*-linked pairs, and that Vedic word order becomes measurably more projective over time.

Keywords: dependency-length minimization, Vedic Sanskrit, word order, *sannidhi*, *ākāṅkṣā*, non-projectivity, Universal Dependencies, treebank

Advanced International Journal of Multidisciplinary Research
E-ISSN: 2584-0487
editor@aijmr.com
Volume 3, Issue 4, July-August 2025
CrossRef DOI: 10.62127/aijmr.2025.v03i04.1484

Selective Locality: Dependency-Length Minimization, Non-Projectivity, and the *Ākāṅkṣā* Account of Word Order in Vedic Sanskrit

The claim that Sanskrit has "free" word order is among the most repeated generalizations in descriptive Indology, and among the least precise. Speijer (1886) already observed that the order of words in a Sanskrit sentence is grammatically underdetermined relative to languages such as Latin or English, and Staal (1967) made the observation a theoretical problem by asking what, if not fixed constituent order, carries the syntactic information that configurational languages encode positionally. The answer offered by the Indian grammatical and philosophical tradition is not that word order is unconstrained but that its constraints are of a different kind. On the tradition's account, the well-formedness of a sentence rests on three conditions on the words that compose it: ākāṅkṣā, the mutual expectancy by which one word requires another to complete its meaning; yogyatā, the semantic compatibility that makes the combination coherent; and sannidhi, the proximity or contiguity in which the expectant elements must be presented (Matilal, 1990). Of these, sannidhi is the condition that speaks most directly to word order, because it is a claim about arrangement rather than about meaning or selection.

What kind of claim is it? In its strongest form, a proximity condition on syntactically linked words is a locality principle, and locality principles are exactly what a large cross-linguistic literature has proposed as a near-universal shaper of word order. Under the heading of dependency-length minimization (DLM), it has been argued that languages arrange words so as to keep syntactically dependent elements close together, thereby reducing the memory and integration cost of processing (Gibson, 1998; Liu, 2008). Gildea and Temperley (2010) showed that English and German order words to shorten dependencies well below chance, and Futrell et al. (2015) extended the result to 37 languages, reporting that observed dependency lengths are shorter than random baselines everywhere they looked. Temperley and Gildea (2018) reviewed the evidence and treated DLM as a candidate cognitive universal. If sannidhi is a locality condition, and if DLM is a universal, then the traditional Sanskrit account and the modern processing account would appear to describe the same force from two directions, and Vedic Sanskrit—an old, richly inflected, apparently free-order language—should exhibit dependency-length minimization.

This convergence has been assumed more often than tested. The quantitative study of Sanskrit word order is recent and still thin. Kulkarni and colleagues have shown that Sanskrit order is far from free in any strong sense and that the dependency structures of Sanskrit verse are largely, though not entirely, projective; Vikram and Kulkarni (2020) analyzed the Bhagavadgītā and reported that roughly one sentence in six contains a crossing dependency, and that well-nestedness alone does not capture the attested non-projective configurations, which they suggest are better classified through the expectancy relations of the tradition. That study, however, was confined to a single classical text, measured non-projectivity rather than dependency length, and did not compare observed arrangements against a formal baseline. Whether Vedic prose and poetry minimize dependency length, whether any such tendency is uniform across the grammar or concentrated in particular relations, and whether the pattern changes across the long chronological span of the Vedic corpus, are open questions.

The present study addresses them using the Universal Dependencies (UD) Treebank of Vedic Sanskrit (Hellwig et al., 2020, 2023), the largest syntactically annotated corpus of early Sanskrit. I test four

Advanced International Journal of Multidisciplinary Research
E-ISSN: 2584-0487
editor@aijmr.com
Volume 3, Issue 4, July-August 2025
CrossRef DOI: 10.62127/aijmr.2025.v03i04.1484

hypotheses. First, if word order is globally shaped by locality, observed dependency lengths should be shorter than a baseline that holds the tree constant but randomizes the order of siblings (*H1*, global DLM). Second, if *sannidhi* operates specifically over *ākāṅkṣā*-linked pairs rather than over the syntactic graph as a whole, then the relations that encode verb-argument expectancy—the *kāraka*-type relations—should be more strongly minimized than adjunct, coordinate, and appositional relations (*H2*, selective locality). Third, if Sanskrit non-projectivity is disciplined rather than chaotic, crossing dependencies should be common but mild, with low gap degree and few genuinely ill-nested structures (*H3*). Fourth, if the well-known drift of Vedic from older poetry toward later technical prose involves a tightening of syntactic order, non-projectivity should decline across the corpus's chronological layers (*H4*). As will be seen, the aggregate prediction of *H1* fails, but its failure is informative: it holds together with strong support for *H2*, *H3*, and *H4*, and the combination points to a specific reinterpretation of *sannidhi*.

Method

Corpus and Materials

The primary data were the UD Treebank of Vedic Sanskrit (treebank identifier UD_Sanskrit-Vedic; the release current as of 2024), comprising 27,182 annotated sentences and 206,440 tokens (Hellwig et al., 2020, 2023). Each token carries a lemma, a universal part-of-speech tag, morphological features, and a labeled syntactic head under the UD scheme (de Marneffe et al., 2021). The corpus is drawn from across the Vedic period and is internally stratified into six ordered layers that the annotators use to periodize the material. Characterized by their dominant source texts, these are: an oldest layer of Ṛgvedic poetry (labeled 1-RV); a Mantra-language layer combining the Ṛgveda with the Atharvaveda and Vājasaneyi mantras (2-MA); a Saṃhitā- and Brāhmaṇa-prose layer dominated by the Aitareya-Brāhmaṇa and the prose of the Maitrāyaṇī, Kāṭhaka, and Taittirīya Saṃhitās (3-PO); a later prose layer dominated by the Bṛhadāraṇyaka and Chāndogya Upaniṣads and related Brāhmaṇa prose (4-PL); a Śrautasūtra layer of technical ritual prose (5-SU); and a small post-Vedic layer containing Mahābhārata and Arthaśāstra material (6-PV). Layers 1 through 5 follow the accepted relative chronology of Vedic; layer 6 is post-Vedic. For an independent replication I used the UD_Sanskrit-UFAL treebank, a small collection of 230 sentences of largely classical Sanskrit annotated under the same scheme.

Dependency Length and Baselines

Following standard practice (Ferrer-i-Cancho, 2004; Liu, 2008; Temperley & Gildea, 2018), the length of a dependency was defined as the absolute difference in linear position between a token and its syntactic head, and the dependency length of a sentence as the mean length of its arcs. Punctuation was ignored, and only sentences forming a single well-formed rooted tree were analyzed. To separate any ordering preference from the mechanical effects of sentence length and tree shape, each observed sentence was compared against two references. The *random-projective baseline* was generated by keeping the dependency tree fixed and, at each node, randomly permuting the relative order of the head and its dependent subtrees on either side; this yields a projective arrangement with the same tree but an order chosen without regard to locality. Fifty such linearizations were generated per sentence and averaged. The

Advanced International Journal of Multidisciplinary Research
E-ISSN: 2584-0487
editor@aijmr.com
Volume 3, Issue 4, July-August 2025
CrossRef DOI: 10.62127/aijmr.2025.v03i04.1484

minimal-arrangement heuristic placed each head's dependent subtrees closest-first and alternating outward from the head, the arrangement-heuristic of Gildea and Temperley (2010) that approximates the projective order minimizing total dependency length. The random-projective baseline thus represents order-indifference, and the minimal arrangement represents order fully optimized for locality; an observed corpus that minimizes dependency length should fall below the random baseline and toward the minimal one.

Operationalizing Expectancy

Testing *H2* requires an operational counterpart to *ākāṅkṣā*. Expectancy in the tradition is the requirement a word imposes on another to complete it, prototypically the requirement a verb imposes on its *kāraka* participants (agent, patient, instrument, recipient, and the like) and the corresponding requirement of governed complements (Kiparsky & Staal, 1969; Cardona, 1997; Bharati et al., 1995). As the treebank is annotated in UD rather than in *kāraka* labels, I approximated the expectancy set by the UD relations that most nearly encode verb-argument and governed-complement dependencies: nsubj, obj, iobj, obl, ccomp, xcomp, and csubj. All remaining relations—coordination, apposition, nominal and adjectival modification, determiners, discourse particles, and the like—formed the comparison set. This mapping is deliberately conservative and admittedly imperfect; its limitations, and in particular the special behavior of the subject relation, are taken up in the Discussion. For each relation I computed the observed mean arc length and the mean arc length of the same relation under the random-projective baseline, and expressed minimization as their ratio and as their signed difference (observed minus baseline, in tokens).

Non-Projectivity

A sentence is non-projective when at least one dependency arc crosses another. For each sentence I recorded whether any crossing occurred and the proportion of arcs involved in a crossing. To gauge the *severity* of non-projectivity I computed the gap degree of each sentence—the maximum number of discontinuities in the set of positions spanned by any subtree—and flagged genuinely ill-nested sentences, in which two disjoint subtrees interleave (Kuhlmann, 2013). Well-nested structures of low gap degree are the mildest, most computationally tractable form of non-projectivity; ill-nested and high-gap structures are the pathological cases.

Statistical Analysis

Uncertainty was quantified by nonparametric bootstrap. Confidence intervals for corpus-level means used 2,000 resamples of sentences; the comparison of the expectancy and non-expectancy partitions used a clustered bootstrap that resampled whole sentences (3,000 replicates), so that the dependence among arcs within a sentence is respected. Analyses were restricted to sentences of 3 to 60 tokens, leaving 24,955 sentences (198,259 tokens) with a mean length of 7.9 and a median of 6 tokens. All procedures were deterministic given a fixed seed and are reproducible from the released code. Because the corpus is a near-complete enumeration of the annotated material rather than a random sample, inferential statistics are reported as descriptive summaries of stability rather than as tests against a hypothetical superpopulation.

Advanced International Journal of Multidisciplinary Research
E-ISSN: 2584-0487
editor@aijmr.com
Volume 3, Issue 4, July-August 2025
CrossRef DOI: 10.62127/aijmr.2025.v03i04.1484

Results

Aggregate Dependency Length

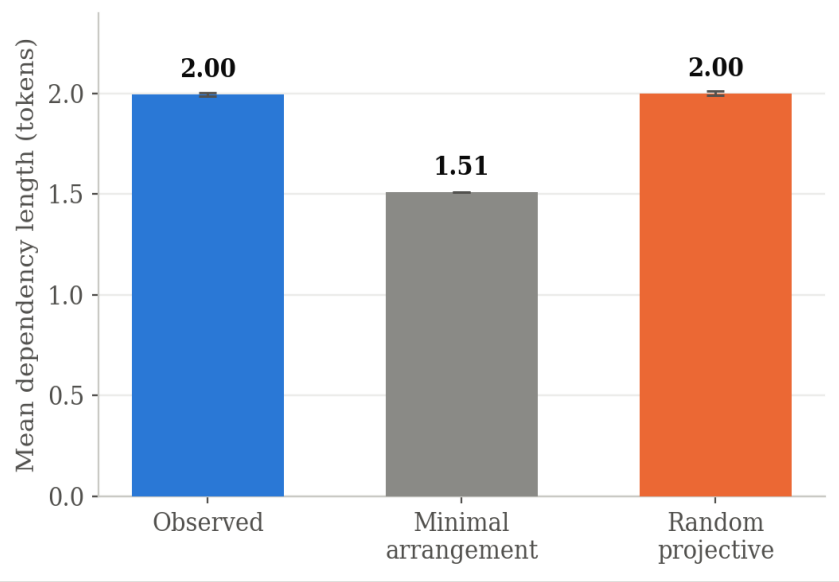
At the level of the whole sentence, Vedic word order showed no dependency-length minimization beyond what the projectivity of its trees already guarantees. The observed mean dependency length was 2.00 tokens (95% CI [1.99, 2.00]), essentially identical to the random-projective baseline of 2.00 tokens (95% CI [1.99, 2.01]) and far above the minimal arrangement of 1.51 tokens. The ratio of observed to random-projective length was 0.997, and observed length fell below its random baseline in only 47.3% of sentences—no more often than chance. *H1* is therefore not supported: as an aggregate, Vedic is order-indifferent with respect to dependency length, sitting at the non-minimizing end of the continuum between random and optimal arrangement (Figure 1, Table 1). This contrasts with the sub-random lengths reported for the languages surveyed by Futrell et al. (2015) and marks Vedic as a language in which whole-sentence locality is weak.

Table 1- *Aggregate Dependency Length Against Two Baselines (Vedic Treebank)*

Arrangement	M (tokens)	95% CI	Ratio to random
Observed order	2.00	[1.99, 2.00]	0.997
Minimal-arrangement heuristic	1.51	—	0.755
Random-projective baseline	2.00	[1.99, 2.01]	1.000

Note. Values are means over 24,955 sentences. Confidence intervals are 2,000-replicate sentence bootstraps. The minimal arrangement follows the closest-first alternating heuristic of Gildea and Temperley (2010).

Figure 1- *Observed Dependency Length Relative to Optimal and Random-Projective Baselines*



Advanced International Journal of Multidisciplinary Research
E-ISSN: 2584-0487
editor@aijmr.com
Volume 3, Issue 4, July-August 2025
CrossRef DOI: 10.62127/aijmr.2025.v03i04.1484

Note. Error bars are 95% bootstrap confidence intervals; the minimal arrangement is a deterministic heuristic and has none. Observed order coincides with the random-projective baseline and lies well above the minimal arrangement.

Selective Locality Across Relations

The aggregate parity is misleading, because it is the sum of two opposing tendencies. When arcs are separated by grammatical relation, a clear and orderly pattern emerges (Figure 2, Table 2). Core verb-argument relations are placed markedly closer than their random baseline: direct objects (obj) at a ratio of 0.72, open clausal complements (xcomp) at 0.71, indirect objects (iobj) at 0.81, and oblique arguments (obl) at 0.86. At the opposite extreme, coordination (conj, ratio 1.62), compounding of the annotated multiword type (1.50), determiners (1.25), adnominal clauses (acl, 1.21), and subordinating markers (mark, 1.13) are placed farther apart than chance. Modifiers and appositives cluster near unity. The verb-argument relations, in other words, are the locus of whatever locality Vedic exhibits, while the material that binds clauses and coordinates phrases actively runs long.

Table 2- *Mean Arc Length by Dependency Relation, With Random-Projective Baseline*

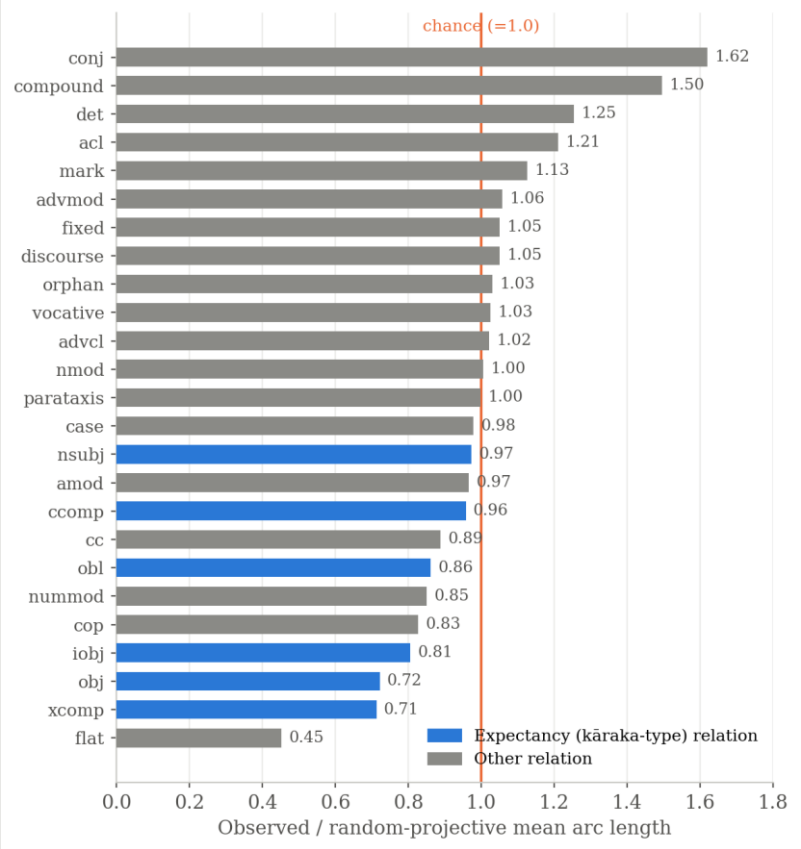
Relation	n	Observed M [95% CI]	Random M	Ratio	Type
flat	13075	1.01 [1.01, 1.01]	2.22	0.45	other
obj	15077	1.82 [1.80, 1.85]	2.52	0.72	*kāraka*
xcomp	1986	1.86 [1.79, 1.94]	2.61	0.71	*kāraka*
iobj	2557	2.22 [2.16, 2.28]	2.76	0.81	*kāraka*
obl	16072	2.22 [2.19, 2.25]	2.58	0.86	*kāraka*
cop	1423	1.72 [1.66, 1.78]	2.08	0.83	other
nsubj	17607	2.36 [2.33, 2.38]	2.42	0.97	*kāraka*
ccomp	2758	4.51 [4.37, 4.65]	4.70	0.96	*kāraka*
nmod	11070	1.63 [1.60, 1.66]	1.62	1.00	other
advmod	13462	2.41 [2.37, 2.45]	2.28	1.06	other
mark	8015	1.71 [1.68, 1.74]	1.52	1.13	other
acl	8366	2.86 [2.80, 2.92]	2.37	1.21	other
det	5567	1.73 [1.70, 1.76]	1.38	1.25	other
conj	13443	4.40 [4.33, 4.49]	2.72	1.62	other

Note. Relations with at least 1,000 arcs, ordered by ratio. Ratio is observed mean divided by random-projective mean; values below 1.00 indicate placement closer than chance. Confidence intervals are 800-

Advanced International Journal of Multidisciplinary Research
E-ISSN: 2584-0487
editor@aijmr.com
Volume 3, Issue 4, July-August 2025
CrossRef DOI: 10.62127/aijmr.2025.v03i04.1484

replicate bootstraps. "kāraka" marks the expectancy set used to test H2. The subject relation nsubj, though a core kāraka, patterns near chance; see Discussion.

Figure 2- *Locality by Dependency Relation*



Note. Bars show the ratio of observed to random-projective mean arc length for every relation with at least 1,000 arcs; the vertical line marks chance (1.0). Expectancy (kāraka-type) relations are shaded. Values below 1.0 indicate placement closer than a projectivity-preserving random order. flat is a fixed multiword relation whose members are adjacent by definition.

The Expectancy Partition

Aggregated into the two partitions, the contrast is unambiguous and stable. Expectancy (*kāraka*-type) arcs were placed, on average, 0.36 tokens closer than their relation-specific random baseline (95% CI [−0.38, −0.34]), whereas all other arcs were placed 0.16 tokens *farther* than baseline (95% CI [+0.13, +0.18]). The difference between the partitions, −0.52 tokens (95% CI [−0.55, −0.49]), is far from zero and stable under a clustered bootstrap that resamples whole sentences. *H2* is supported: the locality that the aggregate hides is real and is concentrated exactly on the relations that encode *ākāṅkṣā*. The one conspicuous exception within the expectancy set is the subject relation, nsubj (ratio 0.97), which behaves almost like chance and is discussed below.

Advanced International Journal of Multidisciplinary Research
E-ISSN: 2584-0487
editor@aijmr.com
Volume 3, Issue 4, July-August 2025
CrossRef DOI: 10.62127/aijmr.2025.v03i04.1484

Non-Projectivity

Crossing dependencies were common but mild (Table 3). Of the 24,955 sentences, 19.4% contained at least one crossing arc, and 8.9% of all arcs were involved in a crossing. The overwhelming majority of non-projectivity was of the gentlest kind: 20.5% of sentences had gap degree exactly one, only 1.9% had gap degree two or higher, and just 1.4% of all sentences—about 7% of the non-projective ones—were genuinely ill-nested. *H3* is supported: Vedic non-projectivity is not disorder but a disciplined, mostly well-nested, low-gap phenomenon of the kind that mildly context-sensitive grammars are designed to capture (Kuhlmann, 2013). The crossing arcs were not evenly distributed across relations: adnominal clauses (acl, 26.9% crossing), determiners (22.8%), vocatives (24.5%), and nominal modifiers (14.1%) crossed most often, while the fixed and compound relations almost never crossed and the core arguments crossed relatively rarely (nsubj 6.0%, obj 5.2%). The elements that run long and the elements that cross are, revealingly, the same non-expectancy relations.

Table 3- *Non-Projectivity in the Vedic and Classical (UFAL) Treebanks*

Measure	Vedic	UFAL
Sentences with ≥1 crossing arc	19.4%	17.5%
Arcs involved in a crossing	8.9%	—
Sentences with gap degree = 1	20.5%	—
Sentences with gap degree ≥ 2	1.9%	—
Ill-nested sentences (of all)	1.4%	—
Observed / random-projective ratio	0.997	1.054

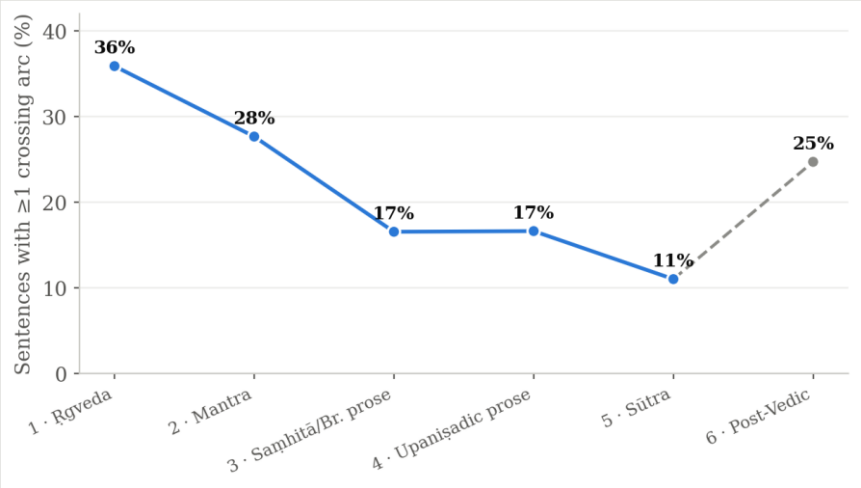
Note. UFAL comprises 211 analyzed sentences of largely classical Sanskrit; dashes mark measures not reported for the smaller corpus. The UFAL observed-to-random ratio above 1.0 reproduces the Vedic finding that whole-sentence order is not length-minimizing.

Diachronic Variation

Non-projectivity declined steeply and monotonically across the five Vedic layers, from 35.9% of sentences in the Ṛgvedic layer to 27.7% in the Mantra layer, about 16.6% in the two prose layers, and 11.0% in the Śrautasūtra layer; the post-Vedic layer rose again to 24.7% (Figure 3). *H4* is supported for the Vedic sequence. Crucially, this trend is not an artifact of sentence length. If crossings were simply a by-product of longer sentences, the Sūtra layer—whose sentences are the longest in the corpus, averaging about nine tokens—should cross most; instead it crosses least, while the shorter Ṛgvedic poetry crosses most. The decline therefore reflects a genuine tightening of syntactic order over the history of the language, from the freer arrangement of the oldest poetry toward the more configurational prose of the ritual and philosophical literature. The rebound in the post-Vedic layer, dominated by epic and *śāstra* material, is consistent with the metrical freedom of epic verse.

Advanced International Journal of Multidisciplinary Research
E-ISSN: 2584-0487
editor@aijmr.com
Volume 3, Issue 4, July-August 2025
CrossRef DOI: 10.62127/aijmr.2025.v03i04.1484

Figure 3- *Non-Projectivity Across the Chronological Layers of the Corpus*



Note. Points show the percentage of sentences containing at least one crossing arc in each layer; the solid segment traces the Vedic sequence (layers 1–5) and the dashed segment marks the post-Vedic layer (6). The layer with the longest sentences (Sūtra) shows the fewest crossings, ruling out sentence length as the driver.

Replication on Classical Sanskrit

The independent UFAL treebank, though small (211 analyzable sentences of largely classical prose), reproduced the two central findings. Observed dependency length there was 2.25 tokens against a random-projective baseline of 2.13, a ratio of 1.05—again no minimization below the projective baseline—while 17.5% of sentences contained a crossing arc, closely matching both the Vedic rate and the one-in-six figure reported by Vikram and Kulkarni (2020) for the Bhagavadgītā. The convergence across corpora, periods, and registers suggests that the aggregate non-minimization and the moderate, disciplined non-projectivity are properties of Sanskrit syntax broadly, not of one text or layer.

Discussion

The findings sort into a single coherent picture. Vedic Sanskrit does not minimize dependency length as a global property of the sentence; its observed arrangements are, on average, indistinguishable from a random order that respects only the constraint of projectivity. Yet beneath that null result lies a strong and orderly locality effect confined to the relations that the indigenous tradition would classify as expectancy relations. Verb arguments are pulled toward their governors; coordinators, subordinators, and modifiers are not, and in fact run longer than chance. Non-projectivity is frequent but mild and well-nested, it is carried by the same non-expectancy relations that run long, and it recedes as the language moves from archaic poetry to classical prose. I take these results to motivate a specific reinterpretation of *sannidhi*.

Advanced International Journal of Multidisciplinary Research
E-ISSN: 2584-0487
editor@aijmr.com
Volume 3, Issue 4, July-August 2025
CrossRef DOI: 10.62127/aijmr.2025.v03i04.1484

Sannidhi as Selective, Expectancy-Scoped Locality

If *sannidhi* were a global minimization principle—a claim that all syntactically linked words are kept close—the aggregate result would refute it. But the tradition does not state *sannidhi* over the syntactic graph as a whole. In the classical analysis, *sannidhi* is the third condition alongside *ākāṅkṣā* and *yogyatā*, and it is the proximity of *expectant* elements specifically: the words bound by mutual expectancy must be presented without disruptive separation so that the expectancy can be satisfied and verbal cognition (*śābdabodha*) can arise (Matilal, 1990; Kiparsky & Staal, 1969). Read this way, *sannidhi* is a locality condition scoped to *ākāṅkṣā*-linked pairs, not a global one—and that is precisely the pattern in the data. The relations that carry expectancy are minimized; the relations that do not carry it are free to spread. The traditional triad, in other words, predicts selective locality, and selective locality is what Vedic exhibits. The contribution of the present study is to show that this is not merely a coherent reading of the tradition but a measurable property of the largest Vedic corpus.

The chief complication is the subject relation. On the *kāraka* analysis the agent (*kartṛ*) is a paradigm case of an expectant participant, yet nsubj showed almost no minimization (ratio 0.97). Two considerations bear on this. First, subjects in Vedic are frequently the topic and are independently attracted to sentence-initial position by information structure, which can pull them away from the verb; the position of an overt subject may therefore be governed by discourse prominence rather than by *sannidhi*. Second, subjects are often unexpressed, so the annotated nsubj arcs are a biased, overt-only sample. The behavior of nsubj is thus better read as a limitation of the UD-to-*kāraka* approximation than as evidence against selective locality, and it identifies the single most important place where a genuinely *kāraka*-annotated corpus would sharpen the test.

Reconciling the Aggregate Null With Cross-Linguistic DLM

Does the aggregate result contradict the DLM literature? Not straightforwardly. The cross-linguistic studies report that observed lengths fall below random baselines (Futrell et al., 2015; Temperley & Gildea, 2018), and the present study finds the observed length exactly at the random-projective baseline. Several features of the Vedic material temper the comparison. The annotated sentences are very short—a median of six tokens—which sharply limits the room in which dependency length can vary and attenuates any minimization signal; DLM effects in other languages grow with sentence length. The older layers are metrical, and the demands of Vedic meter compete directly with locality for control of word order. And the corpus contains a high proportion of the coordinate and appositive structures that, as shown here, run systematically long. The right conclusion is not that Vedic violates a processing universal but that whole-sentence DLM is a weak and uneven force in this language, real only where expectancy concentrates it. This localizes the universal rather than denying it, and it is consistent with the observation that DLM is strongest for exactly the argument relations that dominate the effect here.

Advanced International Journal of Multidisciplinary Research
E-ISSN: 2584-0487
editor@aijmr.com
Volume 3, Issue 4, July-August 2025
CrossRef DOI: 10.62127/aijmr.2025.v03i04.1484

Non-Projectivity, Well-Nestedness, and Diachrony

The non-projectivity results both confirm and refine prior Sanskrit work. The crossing rate of roughly one sentence in five, and the classical rate near one in six, agree closely with Vikram and Kulkarni's (2020) finding for the Bhagavadgītā and with the general expectation that Sanskrit is largely but not wholly projective. The present analysis departs from theirs on well-nestedness: whereas they reported that well-nested and ill-nested crossings were comparably frequent in the Gita, I find that in the far larger Vedic corpus genuine ill-nesting is rare—about 7% of crossing sentences—and that almost all non-projectivity is well-nested and of gap degree one. The discrepancy is plausibly a matter of corpus and annotation scheme (classical verse under a *kāraka*-style analysis versus Vedic under UD) and of the sentence-level definition of ill-nesting used here, and it deserves direct study. The diachronic decline is, to my knowledge, a new quantitative result: syntactic order tightens measurably across the Vedic period, independent of sentence length. It offers a syntactic parallel to the long-noted stylistic movement from the paratactic freedom of the hymns to the more rigid periods of *sūtra* prose.

Relation to the Pāṇinian and Computational Tradition

The interpretation offered here connects the processing literature to a computational reading of the Indian grammarians. The *kāraka* system is, in modern terms, an argument-linking level mediating between meaning and morphology (Kiparsky & Staal, 1969; Cardona, 1997), and the Pāṇinian framework for parsing free-word-order languages built by Bharati et al. (1995) treats *ākāṅkṣā* and *yogyatā* as the constraints that recover structure when order does not encode it. The present results give that framework an empirical footing: the expectancy relations it relies on are exactly the relations that are spatially privileged in the corpus, which is what one would expect if the parser's constraints track a real property of the language. They also suggest a concrete refinement for computational parsers of Sanskrit—that locality features be weighted by relation type, strongly for argument relations and weakly or negatively for coordinate and clausal ones—rather than applied uniformly.

Limitations

Four limitations qualify these conclusions. First and most important, *ākāṅkṣā* was approximated by a fixed set of UD relations rather than measured on *kāraka* annotation; the anomalous behavior of nsubj shows that this proxy is imperfect, and a *kāraka*-annotated treebank—such as those developed in the Saṃsādhanī and Pāṇinian dependency traditions—would allow a direct test. Second, the annotated Vedic sentences are short and, by the corpus's own documentation, sometimes fragmentary, which limits the dynamic range of the dependency-length measures and may understate minimization. Third, the layer-based diachrony conflates chronology with register and genre, since the older layers are also the more metrical ones; disentangling age from meter would require matched prose and verse within a period. Fourth, meter itself was not modeled, so the competition between metrical and syntactic pressures in the poetic layers remains to be quantified. None of these threatens the central contrast between expectancy and non-expectancy relations, which is large and stable, but each marks a boundary of the present evidence.

Advanced International Journal of Multidisciplinary Research
E-ISSN: 2584-0487
editor@aijmr.com
Volume 3, Issue 4, July-August 2025
CrossRef DOI: 10.62127/aijmr.2025.v03i04.1484

Conclusion

Sanskrit word order is neither free nor globally optimized for locality. In the largest corpus of Vedic, dependency length at the level of the whole sentence is indistinguishable from a projectivity-respecting random order, yet the relations that encode verb-argument expectancy are placed reliably closer than chance while coordinate, subordinate, and modifying relations run long. Non-projectivity is common but mild and well-nested, borne by the same non-expectancy relations, and it declines across the Vedic period as syntactic order tightens. Taken together, these results support reading *sannidhi* not as a claim that all linked words are kept close but as a selective locality operating over *ākāṅkṣā*-linked pairs—a reinterpretation under which the ancient tripartite account of what binds a sentence and the modern quantitative study of what shapes word order describe, from two directions, the same fact.

Advanced International Journal of Multidisciplinary Research
E-ISSN: 2584-0487
editor@aijmr.com
Volume 3, Issue 4, July-August 2025
CrossRef DOI: 10.62127/aijmr.2025.v03i04.1484

REFERENCES:

1. Bharati, A., Chaitanya, V., & Sangal, R. (1995). *Natural language processing: A Pāṇinian perspective*. Prentice-Hall of India.
2. Cardona, G. (1997). *Pāṇini: His work and its traditions. Vol. 1: Background and introduction* (2nd rev. ed.). Motilal Banarsidass.
3. de Marneffe, M.-C., Manning, C. D., Nivre, J., & Zeman, D. (2021). Universal Dependencies. *Computational Linguistics*, *47*(2), 255–308. https://doi.org/10.1162/coli_a_00402
4. Ferrer-i-Cancho, R. (2004). Euclidean distance between syntactically linked words. *Physical Review E*, *70*(5), 056135. https://doi.org/10.1103/PhysRevE.70.056135
5. Futrell, R., Mahowald, K., & Gibson, E. (2015). Large-scale evidence of dependency length minimization in 37 languages. *Proceedings of the National Academy of Sciences*, *112*(33), 10336–10341. https://doi.org/10.1073/pnas.1502134112
6. Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, *68*(1), 1–76. https://doi.org/10.1016/S0010-0277(98)00034-1
7. Gildea, D., & Temperley, D. (2010). Do grammars minimize dependency length? *Cognitive Science*, *34*(2), 286–310. https://doi.org/10.1111/j.1551-6709.2009.01073.x
8. Hellwig, O., Scarlata, S., Ackermann, E., & Widmer, P. (2020). The treebank of Vedic Sanskrit. In *Proceedings of the Twelfth Language Resources and Evaluation Conference* (pp. 5137–5146). European Language Resources Association.
9. Hellwig, O., Nehrdich, S., & Sellmer, S. (2023). Data-driven dependency parsing of Vedic Sanskrit. *Language Resources and Evaluation*, *57*(3), 1173–1206. https://doi.org/10.1007/s10579-023-09636-5
10. Kiparsky, P., & Staal, J. F. (1969). Syntactic and semantic relations in Pāṇini. *Foundations of Language*, *5*(1), 83–117.
11. Kuhlmann, M. (2013). Mildly non-projective dependency grammar. *Computational Linguistics*, *39*(2), 355–387. https://doi.org/10.1162/COLI_a_00125
12. Liu, H. (2008). Dependency distance as a metric of language comprehension difficulty. *Journal of Cognitive Science*, *9*(2), 159–191. https://doi.org/10.17791/jcs.2008.9.2.159
13. Matilal, B. K. (1990). *The word and the world: India's contribution to the study of language*. Oxford University Press.
14. Scharf, P. M. (Ed.). (2015). *Sanskrit syntax: Selected papers presented at the seminar on Sanskrit syntax and discourse structures*. The Sanskrit Library.
15. Speijer, J. S. (1886). *Sanskrit syntax*. E. J. Brill.
16. Staal, J. F. (1967). *Word order in Sanskrit and universal grammar* (Foundations of Language Supplementary Series, Vol. 5). D. Reidel.
17. Temperley, D., & Gildea, D. (2018). Minimizing syntactic dependency lengths: Typological/cognitive universal? *Annual Review of Linguistics*, *4*, 67–80. https://doi.org/10.1146/annurev-linguistics-011817-045617

Advanced International Journal of Multidisciplinary Research

E-ISSN: 2584-0487
editor@aijmr.com

Volume 3, Issue 4, July-August 2025

CrossRef DOI: 10.62127/aijmr.2025.v03i04.1484

18. Vikram, S., & Kulkarni, A. (2020). Free word order in Sanskrit and well-nestedness. In *Proceedings of the 17th International Conference on Natural Language Processing (ICON)* (pp. 308–316). NLP Association of India.