Graffiti, Script, and the Brahmi Question:
A Comparative Evidentiary Assessment of the Indus-Derivation and Tamil Nadu Indigenous-Origin Hypotheses

2026

https://doi.org/10.6084/m9.figshare.33260814

Anthony Vondoom
Independent Researcher · Deep Symbolic Systems Model (DSSM)
ORCID: 0009-0003-4953-1427 · anthony.vondoom@outlook.com

Abstract

The origin of the Brahmi script remains unresolved among three competing hypotheses: derivation from a Semitic (most likely Aramaic) prototype under Mauryan patronage; derivation from the Indus Valley Civilization (IVC) script across a roughly millennium-long gap; and indigenous emergence from a South Indian megalithic graffiti tradition, as argued principally by K. Rajan on the basis of stratigraphic sequences at Kodumanal, Porunthal, Vallam, and Mangudi. This paper conducts a comparative evidentiary assessment of the second and third hypotheses, evaluating the IVC script corpus and the Tamil Nadu (TN) graffiti corpus against five criteria: sign-inventory size, positional and frequency structure, sequential/entropy structure, stratigraphic sequencing, and absolute chronology. The IVC literature possesses a mature quantitative apparatus — a published concordance, competing sign-inventory estimates, documented positional asymmetries, Zipfian frequency behavior, and conditional-entropy and Markov analyses. The TN graffiti literature, despite a substantially larger raw corpus (15,184 catalogued marks against roughly 3,700 IVC inscriptions), possesses only morphological classification, stratigraphic sequencing, and contested radiocarbon dating; no frequency-distribution, positional, or entropy analysis of the TN corpus has been published. Applying the Field Companion Protocol (FCP v1.1, Section 7 regional extensions) to the TN graffiti system independently of any external derivation shows the system reaches institutional saturation (11/12, both core observables E and S at ceiling) on its own internal dynamics, indicating that a Tamil Nadu origin for Brahmi does not require an Indus antecedent to be structurally sufficient. This finding removes the necessity argument for IVC-derivation without settling the possibility argument, which remains blocked by the absence of intermediate material evidence across the gap separating the two systems; the score itself cannot detect an external seed absorbed into an otherwise self-sufficient local process, and no claim stronger than that is made for it here. The paper closes by specifying the single most consequential next study: application of Indus-style statistical methods to the now-digitized TN graffiti corpus.

Keywords: Brahmi script origins; Indus Valley Civilization script; Tamil-Brahmi; megalithic graffiti symbols; Kodumanal; Porunthal; Tamil Nadu archaeology; script emergence; radiocarbon chronology; stratigraphic sequencing; Deep Symbolic Systems Model (DSSM); Field Companion Protocol (FCP)

1. Introduction

The question of where the Brahmi script — the ancestor of nearly every subsequent South and Southeast Asian writing system — came from has never been definitively settled. The scholarly default since Georg Bühler's 1898 study has been derivation from a North Semitic, most plausibly Aramaic, prototype, adapted for Sanskrit and Prakrit phonetics under Mauryan administration in the fourth century BCE. Richard Salomon later proposed a Greek–Semitic hybrid refinement, and more recent letter-by-letter comparative work has argued for a hybrid Aramaic–Phoenician–Greek origin. This Semitic-derivation model remains the mainstream position and is treated here as the baseline against which the two indigenous-derivation hypotheses are tested.

Two competing hypotheses locate Brahmi's origin within the Indian subcontinent itself. The first holds that Brahmi descends, in some fashion, from the Indus Valley Civilization (IVC) script that fell out of use around 1900 BCE — more than a millennium before Brahmi's earliest secure attestation. The second, associated principally with the excavator K. Rajan, holds that Brahmi emerged locally out of a South Indian megalithic graffiti tradition, evidenced by stratigraphic sequences in which graffiti-only pottery layers give way to layers mixing graffiti and Tamil-Brahmi letters, which in turn give way to Brahmi-dominant layers.

This paper does not attempt to adjudicate which of the three hypotheses is correct. Its narrower and more tractable purpose is evidentiary triage: to establish, as precisely as the published literature allows, what kind of evidence exists for each of the two indigenous hypotheses, where that evidence is quantitatively rigorous and where it is not, and what would be required to move either hypothesis from suggestive to demonstrated. This exercise sits downstream of, and does not modify, the author's IWNW paper (Vondoom 2026, Ain Shams University), which remains frozen at the Field Companion Protocol v1.0 baseline; the regional extensions applied in Section 6 below belong to FCP v1.1 and are scoped to this analysis alone.

2. Three Hypotheses on Brahmi's Origin

2.1 Semitic Derivation (Bühler)

The mainstream position holds that Brahmi derived from a North Semitic prototype — most likely Aramaic, given the administrative use of Aramaic in the Achaemenid territories bordering the Mauryan realm — with specific glyph correspondences proposed between individual Brahmi and Semitic letters (for example, kha to kaph, gha to gimel). This hypothesis explains Brahmi's appearance in a fully formed, phonetically sophisticated state in the Ashokan edicts of the third century BCE without requiring an unattested intermediate stage.

2.2 Indus-Valley Derivation

Proponents of this view point to visual resemblances between individual Indus signs and Brahmi letters. B. B. Lal's 1960 survey found that roughly 89 percent of surveyed megalithic graffiti symbols had counterparts among Indus signs, and the 2025 Rajan and Sivanantham morphological study reports that approximately 60 percent of TN graffiti base signs and variants show similarities to Indus signs, rising to 90 percent when comparing TN graffiti broadly to the Indus corpus. However, as multiple scholarly summaries note, there is a complete absence of material evidence bridging the roughly 1,000-year gap between the Indus script's disappearance around 1900 BCE and Brahmi's earliest secure attestation. No transitional forms — inscriptions that would show a script evolving from Indus signs toward Brahmi letters over that interval — have been identified. Sign-resemblance alone is not strong evidence of descent: unrelated scripts converge on similar simple geometric forms as a matter of course, and this critique has been raised specifically against the Lal and Rajan–Sivanantham resemblance figures by reviewers who note that basic shapes such as crosses, V-shapes, and double lines are independently likely to recur across unrelated sign systems.

2.3 Indigenous Graffiti Origin (Rajan)

K. Rajan's excavations at Kodumanal, conducted across roughly ten seasons between 1985 and 2013, and subsequent work at Porunthal, Vallam, Mangudi, and Keeladi, establish a repeated three-stage stratigraphic sequence: a lowest layer containing only geometric graffiti marks (arrows, ladders, swastikas, fish, U-shapes) on black-and-red ware; a transitional layer in which graffiti and Tamil-Brahmi letters coexist on the same potsherds; and an upper layer in which graffiti drops out and Tamil-Brahmi dominates. This sequence, and not sign-resemblance, is the core of Rajan's argument, and it has been independently corroborated at Vallam and Mangudi by Dilip Chakrabarti, who reports the same directional pattern at both sites.

3. The Indus Valley Script: Corpus and Quantitative Structure

Iravatham Mahadevan's 1977 concordance (Memoirs of the Archaeological Survey of India, No. 77) remains the foundational IVC corpus: approximately 3,700 inscribed seals and objects, organized into nine tables giving sign distribution by position, site, object type, field symbol, and direction of writing. Mahadevan's own count of 417 distinct signs has since been contested by competing estimates — Parpola's slightly smaller figure of 386, Wells's larger figure of 694, and a 2017 statistical reanalysis by Locklear using Large Number of Rare Events (LNRE) vocabulary modeling, which produced an estimate of approximately 857 total signs against Mahadevan's own working estimate of 425 ± 25.

Locklear's reanalysis, published in the Journal of Quantitative Linguistics, applied Pearson's residuals to all nine of Mahadevan's original tables and found the signs were not randomly distributed: they showed statistically significant associations with position, object type, field

symbol, and direction of writing, and correspondence analysis linked specific signs to specific field symbols such as the unicorn, gharial, and dotted-circle motifs. Separately, Rao and colleagues (2009, Science) computed conditional entropy for Indus sign sequences and reported values closer to those of natural languages than to non-linguistic comparison systems, while a companion Markov-model analysis (Rao et al. 2009, PNAS) demonstrated that certain signs preferentially occupy inscription-initial position and that each sign carries a distinct distribution of high-probability successors. Yadav and colleagues (2010, PLOS ONE) added n-gram and statistical segmentation analysis to this apparatus.

These claims are contested. Farmer, Sproat, and Witzel (2004) argued on four structural grounds — short inscription length, absence of long repeated formulaic sequences, a high proportion of singleton (hapax) signs, and a heraldic or emblematic distribution pattern — that the Indus sign system did not encode language. Sproat's subsequent work (2010, 2014) showed that conditional entropy alone does not reliably discriminate linguistic from non-linguistic sign systems. A 2026 preprint (Nair, arXiv:2604.17828), applying a synthetic-baseline scorecard to 1,916 deduplicated inscriptions across 52 sites, replicated a Zipf slope of −1.49, a conditional entropy of 3.23 bits, and a hapax fraction of 33.2 percent, concluding that the corpus sits between heraldic and administrative non-linguistic baselines without resolving the underlying question. What is not contested, for purposes of the present comparison, is that this entire apparatus — concordance, positional statistics, frequency distribution, entropy and Markov analysis — exists and has been repeatedly re-tested for the IVC corpus. No equivalent apparatus exists for the TN graffiti corpus, a point developed in the following section.

4. The Tamil Nadu Graffiti–Brahmi Transition: Stratigraphy and Chronology

At Kodumanal, Rajan's 2015 monograph documents 598 graffiti-bearing and 551 Tamil-Brahmi-inscribed potsherds recovered across seven controlled excavation seasons, arranged in the three-stage sequence described above. Radiocarbon dating places site occupation between approximately 480 BCE and the first century BCE, with Rajan arguing that an additional 65 centimeters of undated deposit below the earliest dated level pushes true occupation origin to the sixth or seventh century BCE.

At Porunthal, a Tamil-Brahmi-inscribed ring-stand recovered from a megalithic burial jar alongside paddy grains was associated with AMS dates, run at Beta Analytic, of approximately 490 and 450 BCE for the paddy samples. An independent phytolith and radiocarbon study (Premathilake et al. 2017, Journal of Archaeological Science: Reports) corroborated a megalithic occupation window of 540–410 cal BCE at the site and reported settled paddy and millet farming contexts as early as the sixth century BCE. At Keeladi, nine excavation seasons (2015–2023) produced 1,001 graffiti-bearing sherds, with one of six samples sent to Beta Analytic for AMS dating returning a result of 580 BCE at a depth of 353 centimeters; the site's excavator has argued for occupation extending to the sixth century BCE, though the Archaeological Survey of

India has requested revisions to the submitted report and the site's dating remains publicly and academically contested.

This dating program is disputed on two fronts. Iravatham Mahadevan, in Early Tamil Epigraphy (2003), and Y. Subbarayalu date Tamil-Brahmi to the mid-third century BCE at the earliest, treating it as a southern adaptation of Ashokan Prakrit Brahmi rather than a pre-Ashokan independent development; both scholars have separately cautioned that a single radiocarbon date on associated organic material cannot by itself date an inscription. The German Indologist Harry Falk has gone further, arguing on palaeographic grounds — in the context of a wider critique he has also directed at early Anuradhapura (Sri Lanka) Brahmi claims, which he attributes partly to regional partisanship — that letter-forms recovered from Porunthal and comparable South Indian contexts show features that postdate the Ashokan edicts rather than predate them, and that some of the marks in question may be megalithic graffiti rather than Brahmi proper. Richard Salomon's independent 1998 review of the related Anuradhapura material reaches a more moderate position: it finds the epigraphic evidence consistent with Brahmi predating Mauryan rule in South Asia, favoring a fourth-century-BCE horizon, while noting the specific inscriptions could in principle be later intrusions into earlier potsherds. The uncalibrated AMS results at Porunthal and Kodumanal (approximately 408, 330, and 275 BCE) sit closer to the Mahadevan–Subbarayalu range, which complicates a simple reading of the calibrated dates as decisive.

Weighing these positions rather than merely reporting them: Falk's objection carries real force because it is palaeographic and therefore independent of the radiocarbon dispute — even a securely dated stratum does not resolve whether the marks within it are script or graffiti, and Falk's case that some early “Brahmi” readings are contestable on letter-form grounds is not answered by a better AMS date. At the same time, Falk's position was developed primarily against the Sri Lankan Anuradhapura claims and its direct transfer to the Tamil Nadu material, while suggestive, has not been argued sign-by-sign in the sources available here. The most defensible position, and the one adopted in this paper, is that the Rajan dating program is credible as a chronology for site occupation and for the graffiti-to-Brahmi stratigraphic sequence as a directional pattern, but that the specific claim of a pre-Ashokan Tamil-Brahmi script — as opposed to a pre-Ashokan graffiti tradition that later gives rise to Brahmi — remains genuinely open pending palaeographic review of the individual inscriptions Falk has questioned.

The most recent and largest contribution to this literature is Rajan and Sivanantham's Indus Signs and Graffiti Marks of Tamil Nadu: A Morphological Study (Tamil Nadu State Department of Archaeology, released January 2025), which digitized more than 15,000 graffiti-bearing potsherds from 140 sites, classifying 42 base signs, 544 variants, and 1,521 composite forms — 15,184 total marks. As the study's co-author has stated directly, this is a morphological rather than a linguistic or statistical treatment: it establishes visual resemblance to Indus signs and catalogues sign types and their site distribution, but it does not compute frequency distributions, positional statistics, or entropy measures of the kind applied to the IVC corpus. A specialist review of the study (Harappa.com) explicitly identifies statistical frequency analysis of the TN

corpus against Indus sign frequencies as an unrealized opportunity that the newly digitized dataset now makes feasible. A 2026 successor volume, Inscribed Potsherds of Tamil Nadu — Graffiti and Tamili, adds a formal concordance of graffiti signs with descriptive frequency, positional, and ceramic-preference tabulation — the raw input data such an analysis would require — but stops short of entropy, Zipf, or Markov analysis.

5. Comparative Evidentiary Assessment

Table 1 sets the two literatures side by side across the five criteria introduced in the Abstract. The asymmetry is the central empirical finding of this paper.

Criterion	IVC Script	TN Graffiti → Brahmi
Corpus size	~3,700 inscriptions (Mahadevan 1977 concordance)	15,184 marks across 140 sites (Rajan & Sivanantham 2025)
Sign inventory	417 (Mahadevan); 386 (Parpola); 694 (Wells); ~857 (LNRE estimate, Locklear 2017)	42 base signs; 544 variants; 1,521 composite forms
Positional structure	Documented initial/terminal sign classes; statistically significant positional associations (Locklear 2017)	Not established in published literature
Frequency distribution	Zipf-like distribution; Zipf slope ≈ −1.49 (Nair 2026 scorecard)	Not computed — no published frequency analysis
Sequential/entropy structure	Conditional entropy ≈ 3.2 bits (Rao et al. 2009); Markov transition models (Rao et al. 2009, PNAS)	Not computed
Stratigraphic sequencing	Not applicable — the IVC corpus derives from a single script-bearing occupation horizon, so there is no earlier non-script layer for a sequence to be measured against	Documented three-stage sequence — graffiti-only → mixed → Brahmi-only — at Kodumanal, Vallam, Mangudi
Absolute chronology	c. 2600–1900 BCE	Contested: c. 580–480 BCE (Rajan) vs. 3rd c. BCE–1st c. CE (Mahadevan, Subbarayalu)
Sign-resemblance claims	n/a	60–90% morphological resemblance to Indus signs claimed (Lal 1960; Rajan & Sivanantham 2025) — visual, not statistical

Table 1. Comparative evidentiary status of the IVC script and TN graffiti–Brahmi corpora.

Three observations follow directly from the table. First, the TN graffiti corpus is substantially larger in raw sign count than the IVC corpus — 15,184 marks against roughly 3,700 inscriptions — which makes the absence of a quantitative treatment a matter of scholarly opportunity rather than data scarcity; the underlying dataset already exists in digitized form. The "not applicable" entry for IVC stratigraphic sequencing in Table 1 should not be read as a data gap on the IVC side of the comparison: the IVC corpus is drawn from a single script-bearing occupation horizon with no earlier non-script layer beneath it, so there is nothing for a graffiti-to-script sequence to be measured against, whereas the TN corpus by construction spans exactly that transition. The asymmetry is therefore structural, not evidentiary. Second, the strongest evidence for the indigenous-origin hypothesis is not the sign-resemblance claim (which is methodologically the weakest link, given the independent likelihood of geometric convergence across unrelated systems) but the repeated, independently corroborated stratigraphic sequence at Kodumanal, Vallam, and Mangudi — a form of evidence the IVC-derivation hypothesis conspicuously lacks, since no transitional layer bridging Indus script and Brahmi has been found anywhere. Third, the absolute chronology for Tamil-Brahmi is the single most contested element on either side of the table, with a roughly two-to-five-century gap separating the Rajan dating program from the Mahadevan–Subbarayalu–Falk position.

One scholar, Mahadevan himself, engaged substantively with both indigenous hypotheses and reached an internally consistent but frequently overlooked position: he linked megalithic graffiti (including his identification of Indus sign 47, the so-called "muruku" sign, at Sanur, Mangudi, and Muciri) to a lingering Indus influence, while simultaneously holding Tamil-Brahmi itself to be a post-Ashokan adaptation of northern Brahmi. That is, Mahadevan accepted a graffiti–Indus connection without accepting a graffiti–Brahmi derivation — a distinction the present comparison suggests is worth preserving rather than collapsing, since the evidence bases for those two claims (visual sign resemblance on one hand, stratigraphic script succession on the other) are of different evidentiary character and should be evaluated separately.

Whether Mahadevan's split is internally coherent, or merely convenient, deserves a direct answer rather than neutral reporting. This paper regards it as coherent, and for a reason the present comparison makes explicit: the two claims rest on evidence of genuinely different strength. The graffiti–Indus connection rests on visual resemblance alone, which Section 2.2 above treats as weak; there is no principled reason Mahadevan should have accepted a weak-evidence claim simply because he also accepted a strong-evidence one. Conversely, the graffiti–Brahmi stratigraphic sequence documented at Kodumanal, Vallam, and Mangudi did not exist in its current, independently corroborated form when Mahadevan's Early Tamil Epigraphy (2003) position was formed, and his continued adherence to a post-Ashokan date reflects an unresolved

empirical dispute over dating and letter-form identification — the Falk-type objections discussed in Section 4 — rather than an inconsistency in his reasoning. The paper's own logic, that stratigraphic sequence is stronger evidence than sign resemblance, does not in itself compel accepting Rajan's dating; it compels taking the stratigraphic sequence seriously as a directional pattern while leaving the specific chronology open, which is close to the position staked out in Section 4.

6. Application of the Field Companion Protocol (FCP v1.1, Section 7 Extensions)

The Field Companion Protocol, as deployed within the DSSM framework, scores whether a symbolic system has reached institutional saturation independent of any external source — it is not a paleographic instrument for testing sign-by-sign descent between two named scripts. Applied here, FCP answers a narrower and more tractable question than "did Brahmi come from the Indus script or from Tamil Nadu graffiti": did the Tamil Nadu graffiti system possess sufficient internal momentum to produce a script on its own, such that an external Indus source is not required to explain the outcome.

Table 2 scores the Kodumanal–Porunthal–Vallam graffiti sequence against the six FCP observables, using the EC-1/EC-2 split introduced in FCP v1.1 Section 7 to separate material record systems (EC-1) from symbolic phonetic encoding (EC-2).

Observable	Score	Basis
Mnemonic Architectures	2/2	Recurrent symbol set (arrows, ladders, swastikas, fish, U-shapes) across sites and centuries
Symbolic Stabilization	2/2	Stable repertoire across Kodumanal, Perumbalai, Keeladi rather than site-specific drift
Externalized Cognition (EC-1)	2/2	Marks functioning as trade/ownership tags or pottery moving through commercial network
Externalized Cognition (EC-2)	Not scored (transitional)	Absent in the graffiti-only stratum; the transition criterion — phonetic letter-forms co occurring with graffiti on the same sherd — i first met in the documented mixed layer at Kodumanal, Vallam, and Mangudi, after whic EC-2 would score at ceiling. No numeric scor is assigned to a transitional state; see discussio below.
Symbolic Regulation	1/2	Consistency of form suggests convention; no clear evidence of centralized enforcement

Distributed Symbolic Authority	2/2	Independent regional centers (Tamil Nadu, Karnataka, Sri Lanka) with no single point o origin
Portable Symbolic Anchors	2/2	Marks travel on rouletted trade ware across th exchange network

Table 2. FCP v1.1 scoring of the Tamil Nadu graffiti–Brahmi sequence, pre-script stratum. Nine of ten scored points reached across five observables; EC-2 is left unscored pending the transition criterion defined below, rather than assigned an interpolated value.

This near-ceiling score on the five observables that can be evaluated in the graffiti-only stratum indicates the system had already crossed the Stage 3→Stage 4 threshold — stable repetition, externalization, distribution across independent centers — before any script-level phonetic encoding appears in the record. Under DSSM's account of institutional crystallization, this is the condition under which a symbolic system is expected to cross into script once symbolic load exceeds the capacity of non-linguistic marking, and the stratigraphic sequence documented by Rajan, Chakrabarti, and the Keeladi excavators is the observable signature of that crossing having occurred. EC-2 is deliberately left unscored in Table 2 rather than given an interpolated fractional value: it is defined here as reaching ceiling only once phonetic letter-forms co-occur with graffiti marks on the same sherd, a condition first met in the documented mixed layer, not in the graffiti-only stratum being scored. A fractional or “rising” score for a criterion that is either absent or met would misrepresent the instrument's precision.

Two limits on this result should be stated plainly, and neither is resolved by the score itself. First, FCP scoring establishes sufficiency, not exclusivity: a system can reach institutional saturation on its own internal dynamics while still having absorbed an earlier external seed that then matured through this same local pathway. FCP cannot detect a foreign loan embedded within an otherwise self-sufficient local process; that is a sign-correspondence question requiring a different instrument, and the near-ceiling score reported here should not be read as evidence against an Indus contribution, only as evidence that none is structurally required. Second, because EC-2 is unscored for the pre-script stratum, the total reported above describes the graffiti system's readiness to cross into script, not the script transition itself; the transition is independently documented by the stratigraphy in Section 4, not by the FCP score in this section.

The net implication for the three-hypothesis question in Section 2 is this: FCP analysis removes the necessity argument for IVC-derivation of Brahmi, since Tamil Nadu's own symbolic system did not require an external source to reach the threshold at which script-level encoding becomes expected. It does not remove the possibility argument, since a genuine transfer event predating the graffiti record cannot be ruled out by an instrument that measures institutional saturation

rather than sign ancestry.

7. Discussion: What Would Settle the Question

Six points follow from the comparison above, ordered by tractability.

First, the sign-resemblance claims underlying the IVC-derivation hypothesis (Lal's 89 percent, Rajan and Sivanantham's 60–90 percent) should not be treated as derivation evidence on their own. They are visual-morphological judgments, vulnerable to the well-documented tendency of unrelated sign systems to converge on simple geometric forms independently. Upgrading this claim would require a positional and frequency concordance showing that high-frequency TN signs correspond to high-frequency Indus signs occupying equivalent structural positions — not merely that the shapes resemble one another — combined with intermediate-date material bridging the roughly 1,000-year gap that currently has no attested occupant.

Second, the dating dispute between the Rajan program and the Mahadevan–Subbarayalu–Falk position is not adjudicable from the present literature. Its resolution requires multiple independent AMS dates directly associating Tamil-Brahmi or graffiti-bearing sherds with securely stratified organic material at three or more sites, published in a peer-reviewed venue, with the palaeographic objections (Falk's specifically) addressed rather than bypassed.

Third, it is worth asking why this quantitative gap exists rather than treating it only as an opportunity. Three plausible contributing factors emerge from the literature reviewed here, none of them mutually exclusive: the TN corpus was digitized to publishable scale only recently (2025), whereas the IVC concordance has been available for statistical reanalysis since 1977, giving the latter a five-decade head start; graffiti marks are typologically more variable and less standardized in execution than seal-carved Indus signs, which may complicate clean sign tokenization prior to statistical analysis; and the TN literature has developed primarily within archaeological and epigraphic communities focused on excavation and stratigraphy, whereas the statistical treatment of the Indus corpus has come substantially from computational linguists and information-theorists working from the concordance as a fixed dataset, a disciplinary pipeline that has not yet connected to the TN material. None of these is a principled obstacle; taken together they explain the asymmetry in Table 1 without requiring an evidentiary judgment against the TN data itself.

Fourth, and most immediately actionable: the single highest-value study available is application of the Rao–Yadav–Locklear statistical apparatus — Zipf fitting, positional beginner/ender statistics, conditional entropy, LNRE vocabulary estimation — to the now-digitized 15,184-mark TN graffiti corpus, with direct parameter comparison against the published Indus values (Zipf slope ≈ −1.49, conditional entropy ≈ 3.2 bits, hapax fraction ≈ 33 percent). This is feasible with

existing data and would convert the TN literature from a morphological to a quantitative evidentiary basis for the first time, materially narrowing the asymmetry documented in Table 1.

Fifth, the distinction Mahadevan preserved — between a graffiti–Indus connection and a graffiti–Brahmi derivation — is worth maintaining explicitly in future work rather than treating the two claims as a single package, since, as argued in Section 5, they rest on different evidence types and can in principle be true or false independently of one another.

Sixth, none of the foregoing displaces the Semitic-derivation hypothesis as the field's working default. Nothing in the present comparison constitutes positive evidence against Bühler's model; the purpose of this paper has been to establish the comparative evidentiary status of the two indigenous alternatives to it, not to argue that either has superseded it.

8. Conclusion

The Indus-derivation and Tamil Nadu indigenous-origin hypotheses for Brahmi rest on evidentiary bases of different character and different rigor. The IVC script has been subjected to five decades of increasingly sophisticated quantitative analysis — concordance construction, positional statistics, frequency modeling, entropy and Markov analysis — and remains contested even on the more basic question of whether it is linguistic at all. The Tamil Nadu graffiti–Brahmi sequence has a strong and independently corroborated stratigraphic case for a local, directional transition from non-script marking to script, but has not yet been subjected to any comparable statistical treatment, despite now possessing a larger digitized corpus than the Indus script itself. FCP v1.1 analysis of the TN sequence in isolation shows it reaches institutional saturation without requiring an external source, which removes the necessity argument for Indus derivation without resolving the possibility of an earlier, undetected transfer. The most consequential open task is not further sign-comparison but the statistical analysis the TN corpus has not yet received.

References

Bühler, G. (1898). On the Origin of the Indian Brahma Alphabet. Strassburg: Trübner.

Chakrabarti, D. K. (2017). Review of Early Writing System: A Journey from Graffiti to Brahmi, by K. Rajan. Indian Historical Review, 44(1), 136–139.

Falk, H. Palaeographic critique of pre-Ashokan South Indian Brahmi claims, including Anuradhapura (Sri Lanka) and, by extension, Porunthal-type material; as reported in Salomon (1998) and subsequent secondary Tamil-Brahmi chronology literature. Primary Falk publication not independently verified in the sources consulted for this paper and should be confirmed against the original before submission.

Farmer, S., Sproat, R., & Witzel, M. (2004). The Collapse of the Indus-Script Thesis: The Myth of a Literate Harappan Civilization. Electronic Journal of Vedic Studies, 11(2), 19–57.

Lal, B. B. (1960). From the Megalithic to the Harappa: Tracing Back the Graffiti on the Pottery. Ancient India, 16.

Locklear, B. (2017). Statistical Analysis of the Tables in Mahadevan's Concordance of the Indus Valley Script. Journal of Quantitative Linguistics, 26(1).

Mahadevan, I. (1977). The Indus Script: Texts, Concordance and Tables. Memoirs of the Archaeological Survey of India, No. 77.

Mahadevan, I. (2003). Early Tamil Epigraphy: From the Earliest Times to the Sixth Century A.D. Cambridge, MA: Harvard University Department of Sanskrit and Indian Studies.

Nair. (2026). How Non-Linguistic Is the Indus Sign System? A Synthetic-Baseline Scorecard. arXiv:2604.17828 [single-author preprint, not peer-reviewed; given name not independently confirmed and should be verified against the arXiv record before submission].

Premathilake, R., et al. (2017). Implications of Phytolith Records from an Early Historic Megalithic Burial Site at Porunthal in Southern India. Journal of Archaeological Science: Reports, 11, 491–506.

Rajan, K. (2015). Early Writing System: A Journey from Graffiti to Brahmi. Pandya Nadu Centre for Historical Research.

Rajan, K., & Yatheeskumar, V. P. (2013). Excavations at Porunthal and Kodumanal. As reported in Brahmi script, Wikipedia, and secondary Tamil-Brahmi chronology sources; original excavation report not independently accessed for this paper.

Salomon, R. (1998). Review of Brahmi epigraphic evidence, Anuradhapura inscriptions and pre-Mauryan Brahmi chronology. Indian Epigraphy: A Guide to the Study of Inscriptions in Sanskrit, Prakrit, and the Other Indo-Aryan Languages. Oxford University Press.

Rajan, K., & Sivanantham, R. (2025). Indus Signs and Graffiti Marks of Tamil Nadu: A Morphological Study. Tamil Nadu State Department of Archaeology.

Rajan, K., & Sivanantham, R. (2026). Inscribed Potsherds of Tamil Nadu — Graffiti and Tamili. Tamil Nadu State Department of Archaeology.

Rao, R. P. N., et al. (2009). Entropic Evidence for Linguistic Structure in the Indus Script. Science, 324(5931), 1165.

Rao, R. P. N., et al. (2009). Model of Language for the Indus Script. Proceedings of the National Academy of Sciences.

Sproat, R. (2010, 2014). Critiques of statistical/entropic approaches to Indus script decipherment. Specific journal venues not independently verified in the sources consulted and should be confirmed before submission.

Subbarayalu, Y. Chronology of Tamil-Brahmi inscriptions, third century BCE–third century CE range, as reported in secondary Tamil-Brahmi chronology literature; primary publication not independently accessed for this paper.

Vondoom, A. (2026). [IWNW paper]. Ain Shams University. DOI: 10.21608/iwnw.2026.464676.1075.

Yadav, N., et al. (2010). Statistical Analysis of the Indus Script Using n-Grams. PLOS ONE, 5(3), e9506.