The Indus Script: What We Know and What We Do Not

Corpus, structure, language hypotheses and standards of proof

IndusValley.org Research Desk 5 September 2026

Direct answer. The Indus signs form a highly structured body of short inscriptions, used predominantly right-to-left. No accepted bilingual exists, most texts are extremely short, sign variants remain disputed and the language is unknown. Statistical structure supports organised communication but does not supply phonetic values or a translation.

Contents

1 What survives 2

2 Corpus choices shape the result 2

3 How many signs? 2

4 Direction and positional structure 2

5 Brevity is the central limitation 2

6 Does the system encode language? 2

7 Language hypotheses 3

8 What computation can contribute 3

9 Reproducibility checklist 3

10 Evidence assessment 3

11 Conclusion 3

1 What survives

The evidence is a large but difficult corpus of short sign sequences on stamp seals, sealings, tablets, pottery, copper objects and other artefacts. The inscriptions belong to archaeological objects and contexts; photographs, provenience and object function are therefore as important as a typed line of sign numbers.

2 Corpus choices shape the result

The Corpus of Indus Seals and Inscriptions provides photographic documentation, while Mahadevan's concordance and later digital corpora encode sign sequences for comparison. They are not interchangeable. Scholars differ over damaged signs, allographs, segmentation and orientation, so a computational result must state exactly which corpus and normalisation rules it uses.

3 How many signs?

Mahadevan distinguished 417 signs and Parpola about 386; later inventories can be larger. The difference reflects decisions about whether visually similar forms are separate signs or variants. A high sign count does not by itself establish a writing-system type. Any decipherment must show that its results remain stable under reasonable alternative classifications.

4 Direction and positional structure

Crowding, overlap and seal impressions support right-to-left as the predominant direction, although left-to-right, radial and ambiguous cases exist. Markov, network and positional studies identify recurring initial, medial and terminal behaviour. These patterns demonstrate a constrained communication system, but they do not reveal words, grammar or sound values.

5 Brevity is the central limitation

Most surviving inscriptions contain only a few signs. Shortness sharply reduces the amount of grammar and repetition available for testing and makes accidental matches easier. The corpus also lacks an accepted bilingual inscription. A longer text or bilingual would change the evidential position far more than another visually appealing one-sign comparison.

6 Does the system encode language?

Rao and colleagues found conditional-entropy behaviour closer to linguistic sequences than to several non-linguistic controls. Farmer, Sproat and Witzel argue that extreme brevity, sign proliferation and the absence of long texts are consistent with a non-linguistic emblem system. Later work confirms strong sequential organisation, but statistics alone have not settled whether every sequence encodes spoken language.

7 Language hypotheses

Proto-Dravidian, Indo-Aryan, Munda and other proposals remain hypotheses. Rebus methods and later-language homophones can suggest readings, but isolated resemblance has little discriminatory power. A credible language identification must explain a large pre-declared sample, sign order, variation and archaeological context more economically than rival models.

8 What computation can contribute

Computation can measure ordering, test allograph proposals, map regional and chronological variation, digitise images and expose cherry-picking. It cannot manufacture phonetic anchors. A model trained only against a preferred later language risks reproducing its starting assumption. Held-out tests and comparisons with rival languages and non-linguistic systems are essential.

9 Reproducibility checklist

1. Publish the corpus, artefact identifiers and usable images.
2. State sign, allograph, damage and direction rules.
3. Pre-declare readings and test them on held-out inscriptions.
4. Require linguistic coherence with few ad hoc substitutions.
5. Check chronology, object type and archaeological context.
6. Make predictions independent researchers can falsify.
7. Compare against rival languages and non-linguistic models.

10 Evidence assessment

Strong	Structured sign use; predominantly right-to-left direction; recurring positional constraints.
Probable	Administrative, economic, ownership or identity functions that varied with object and context.
Contested	Whether every sequence encodes spoken language.
Unresolved	Language, phonetic values and translations.

11 Conclusion

The Indus signs form a highly structured, predominantly right-to-left system. Their use in administrative, economic, ownership or identity practices is plausible and may have varied by object. The underlying language, phonetic values and translations remain unresolved, and no proposed decipherment has achieved broad scholarly acceptance.

References

1. Corpus of Indus Seals and Inscriptions, ASI/IGNCA-hosted volume.
2. Rao et al., Entropic evidence for linguistic structure, Science (2009).
3. Yadav et al., A Markov model of the Indus script, PNAS (2010).
4. Rao et al., Entropy, the Indus Script, and Language (2010).
5. Farmer, Sproat and Witzel, The collapse of the Indus-script thesis (2004).
6. Recchia and Louwerse, Archaeology Through Computational Linguistics (2016).
7. Das et al., Identifying allographs in undeciphered scripts (2021).
8. Pal, Interrogating Indus inscriptions (2019).

This paper is an evidence review for public scholarship.

Corrections: info@indusvalley.org.