

Structure of Indus Writing

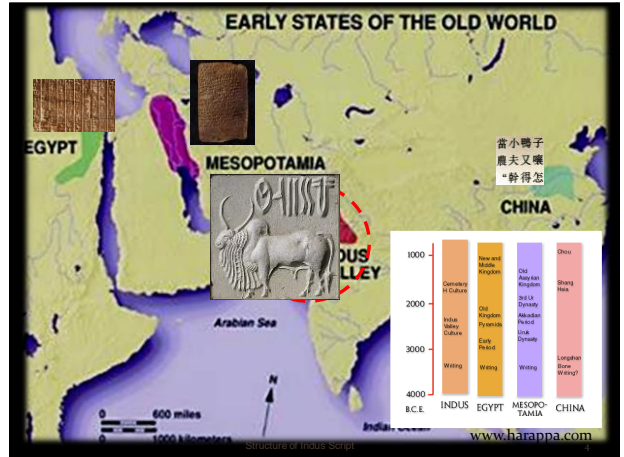
Nisha Yadav
Tata Institute of Fundamental Research, Mumbai

"The practice of writing and the development of a coherent system of signs, a script, is something which is seen only in complex societies

.... Writing is a feature of civilizations."

-- Colin Renfrew
(Archaeology and Language, 1987)

Structure of Indus Script 2



Various Objects--> Various Contexts (?)

Seals (62.16%) & Sealing (19.31%)

Miniature Tablets (9.43%)

Copper Tablet (5.37%)

Pottery Graffiti (2.10%)

Ivory sticks (0.78%)

Structure of Indus Script www.harappa.com, Parpola (1992)

Impression on Clay Tags --> Trade (?)

Fig. 2.26. A clay tag from Umma, Iraq. (a) Obverse bearing an impression of a typically Harappan square stamp seal with Indus script and the 'unicorns' bull. (b) Reverse with an impression of cloth. Department of the Ancient Near East (accession no. 143.1.12), Ashmolean Museum, Oxford.

Structure of Indus Script Parpola, 1992 6

Scale of a Typical Seal

For the most part, seals are between 2.5 to 5 square cm in size.

Photographs by Jatin Acharya

Structure of Indus Script 7

Dholavira Sign Board

Source: pubweb.cc.u-tokai.ac.jp/indus/english/2_4_03.html

Structure of Indus Script 8

Indus Valley Civilization

Source: <http://pubweb.cc.u-tokai.ac.jp/indus/english/map.html>

Structure of Indus Script 9

Components of Indus Seals

1. Script (42.30%)
2. Animal Motif (22.26%)
3. Manger (14.23%)
4. Crude (10.51%)
5. Geometry (4.55%)
6. Abstract (1.70%)
7. Scene (1.30%)
8. Human (1.07%)
9. Plant (1.01%)
10. Mythical figure (0.46%)
11. Composite animal (0.45%)
12. Multi-headed animal (0.18%)

Same object can have more than one of these basic components.

Analysis based on CISI Volumes 1 & 2, Yadav & Vahia (Submitted)

Structure of Indus Script 10

Some Interesting Components


Courtesy: Prof. John Huntington

Structure of Indus Script

CISI, www.harappa.com

Emphasis on Symmetry!


Structure of Indus Script 12



“At their very best, it would be no exaggeration to describe them as little masterpieces of controlled realism, with a monumental strength in one sense out of all proportion to their size and in another entirely related to it”


-- Wheeler, 1968

Structure of Indus Script 13




What do we know about the Indus Script?

Structure of Indus Script 14




1) Indus script is one of the few scripts that has defied decipherment!

Structure of Indus Script 15



What makes the problem challenging?


- Very short and brief texts –
 - average number of signs is 5
 - longest single line text: 14 signs
 - longest text: 26 signs running in 3 distinct lines.
- Language underneath (if any) is unknown
- Lack of bilingual texts



Longest single line text

Structure of Indus Script 16

Source: Wikipedia



Rosetta Stone


Ancient Egyptian (Hieroglyph)

Demotic

Greek

No Rosetta Stone for Indus script yet!

17



2) The sign list consists of about 400 to 700 signs which look like human, fish etc.

Structure of Indus Script 18

Indus Script Datasets

Database	Mahadevan (1977)	Parpola (1979)	Wells (2006)
No. of texts	3573	3700	3835
No. of signs	417	398	676
Total sign occurrences	13,772	~ 14,800	17,427

We use Mahadevan's concordance for analysis.

Structure of Indus Script 19

Indus Script Signs (1 to 110)

From Mahadevan (1977)

Structure of Indus Script 20

Some More Indus Signs

From Mahadevan (1977)

Indus Script Signs (1 to 110)

From Mahadevan (1977)

Structure of Indus Script 22

A Page from the Concordance of Indus Texts

Text No.	Indus Text
1001	𑀩𑀭𑀮𑀯𑀰
1002	𑀩𑀭𑀮𑀯𑀰
1003	𑀩𑀭𑀮𑀯𑀰
1004	𑀩𑀭𑀮𑀯𑀰
1005	𑀩𑀭𑀮𑀯𑀰
1006	𑀩𑀭𑀮𑀯𑀰
1007	𑀩𑀭𑀮𑀯𑀰
1008	𑀩𑀭𑀮𑀯𑀰
1010	𑀩𑀭𑀮𑀯𑀰
1011	𑀩𑀭𑀮𑀯𑀰
1012	𑀩𑀭𑀮𑀯𑀰
1013	𑀩𑀭𑀮𑀯𑀰
1014	𑀩𑀭𑀮𑀯𑀰
1015	𑀩𑀭𑀮𑀯𑀰
1016	𑀩𑀭𑀮𑀯𑀰
1017	𑀩𑀭𑀮𑀯𑀰

Indus texts are strings of the 417 distinct signs of length not more than 14 signs in a single line.

We analyze the corresponding strings of sign numbers.

→ 342-194-67-59-87-99-267

From Mahadevan (1977) 23

3) What about its direction?

- Cramping of signs towards the left end of objects
- Overflow of signs at the left end of objects
- Gap towards the right end of objects

All this is indicative of right to left direction (83% of times), though there are a few exceptions!

Mahadevan (1977), Parpola (1992)

Structure of Indus Script 24

Several Past Attempts

- ◉ For 130 years, scientists from a variety of disciplines have been attempting to read the Indus script with no clear answer.
- ◉ The sign system of Indus culture remains ambiguous, with contested claims of decipherment, but no consensus on any of them.
- ◉ Some of the various attempts so far include:
 - ◉ Mahadevan's work – First published concordance (1977)
 - ◉ Parpola et al.'s work (Finnish group) – Dravidian
 - ◉ Knorozov et al.'s work (Soviet group) – Dravidian
 - ◉ Gift Siromoney's statistical analysis
 - ◉ Subbarayappa's interpretation as pure numerals
 - ◉ S. R. Rao's interpretation as Vedic literature
 - ◉ Others (see Possehl, 1996)

Structure of Indus Script 25

“Decipherment is an art. Unfortunately, in the case of Harappan civilization, an art not yet learned.”

*- Gregory Possehl
(Indus Age: The Writing System, 1996)*

WE DECIDED TO START AFRESH...

Structure of Indus Script 26

Our Approach

- ◉ We make no assumption about its content or meaning.
- ◉ Our first emphasis is to attempt to WRITE IN THE SCRIPT RATHER THAN READ.
- ◉ We search for rules of writing without assigning meanings or interpretations.

Structure of Indus Script 27

We start with the Sign Frequencies!

Only 67 out of 417 signs account for over 80% of the writing.

Structure of Indus Script 28

Is there an order in writing?

Structure of Indus Script 29

- ◉ The first question we wanted to answer was:

“Are Indus texts just a collection of randomly ordered signs or is there any sequencing?”
- ◉ So we randomized all the written material and checked how often we get similar sequences by chance and in real Indus data.

Structure of Indus Script 30

Positional Distribution of Sign Sequences

85 % of times starts a text.

89 % of times comes in the middle of texts.

96 % of times ends a text.

Structure of Indus Script
Yadav et al. (2008 a)

Positional Analysis of Frequent Triplets

Three-sign Combination	Frequency	Solo (%)	Left (%)	Middle (%)	Right (%)		
211	89	336	34	2.94	88.24	5.88	2.94
343	123	293	25	0.00	0.00	0.00	100.00
342	162	249	24	4.17	83.33	8.33	4.17
342	169	249	20	5.00	70.00	20.00	5.00
342	8	171	19	5.26	73.68	5.26	15.79
149	130	51	19	0.00	0.00	78.95	21.05
59	87	99	16	0.00	0.00	100.00	0.00
342	87	403	16	6.25	81.25	6.25	6.25
342	149	130	16	0.00	75.00	25.00	0.00
67	99	267	14	0.00	0.00	7.14	92.86
87	99	267	14	0.00	0.00	21.43	78.57
89	336	72	14	0.00	0.00	85.71	14.29
65	99	267	12	0.00	0.00	8.33	91.67
342	244	67	12	8.33	66.67	8.33	16.67
15	389	178	11	9.09	72.73	0.00	18.18
59	171	53	10	0.00	0.00	60.00	40.00
245	245	25	10	10.00	90.00	0.00	0.00

Structure of Indus Script
Yadav et al. (2008 a)

Conclusion 2:

The frequent sign combinations tend to occur at preferred locations in Indus texts.

Structure of Indus Script
Yadav et al. (2008 a)

Can we say that longer texts have multiple units strung together?

Structure of Indus Script

Segmentation of Indus Texts

Longer text can be shown to consist of 2 or more shorter texts occurring as complete texts elsewhere indicating the boundaries.

4254 [Signs]

2371 [Signs]

2015 [Signs]

2605 [Signs]

Mahadevan (1986), Yadav et al. (2008 b)

Structure of Indus Script

Segmentation of Indus Texts

A few highly frequent signs form stable combinations with other signs. Then these sign-combinations can be treated as separate segments.

2476 [Signs]

2168 [Signs]

Mahadevan (1986), Yadav et al. (2008 b)

Structure of Indus Script

Segmentation of Indus Texts

Comparison of frequencies of successive adjacent pairs reveals boundaries at the 'weakest junctions'.

Pairwise Frequencies (1010)

Segmented Text

Mahadevan (1986), Yadav et al. (2008 b)

Structure of Indus Script 43

Indus Dataset Before and After Segmentation

Dataset before Segmentation

Dataset after Segmentation

Text lengths 1 to 14 No segment > 9 signs

Yadav et al. (2008 b)

Structure of Indus Script 44

Examples of Segmentation

Object No.	Segments of Text						
1232:	P148	P86					
	1279	4441					
4254:	P63	T148	P116	PMS	389		
	2371	2015	1226				
2537:	P41	PM14	67	PMS	389	344	PB1
	8001		1093			4385	
2451:	T94	326	87	P131	178		
	1437	2873	4560	2682			

Yadav et al. (2008 b)

Structure of Indus Script 45

Conclusion 3:

It is possible to segment 88% of Indus texts of length 5 and above into smaller segments of length 4 or less.

Yadav et al. (2008 b)

Structure of Indus Script 46

Analysis of Indus Texts by Machine Learning and Data Mining: N-grams & Markov Models

Structure of Indus Script 47

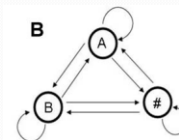
- ### Applying Machine Learning and Data Mining
- Research in machine learning and data mining has led to new techniques for:
 - Learning statistical models of sequences
 - Grammar discovery
 - Pattern recognition
 - Pattern completion
 - We apply these new techniques to the Indus script problem.
- Structure of Indus Script 48

N-gram & Markov Models

- Are probabilistic models which provide a very useful method of modelling different types of sequences.
- These models are not sensitive to the semantic content of the sequences but, reveal the syntax, if any, that the sequences follow.
- The order of the Markov model decides the length of correlation.

Yadav et al. (2009, Submitted), Rao et al. (2009, PNAS)
Structure of Indus Script 49

Markov Model for 3 States A, B and



A	0.2
B	0.8
#	0


A	0.3	0.7	0
B	0.7	0	0.3
#	0	0	1
	A	B	#

Some example Sequences with this Markov model:
BAAB, ABAB, B, etc. (the terminal sign # is not shown).

Sequences not seen with above Markov model:
All texts with repetition of BB or all texts which end in A.

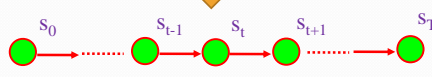
Structure of Indus Script 50

Indus Texts are mapped to sequences



M-355 A

Indus Text



Yadav et al. (2009, Submitted), Rao et al. (2009, PNAS)
Structure of Indus Script 51

How do we model Indus Script?

$S = \{ \text{𑀓}, \text{𑀔}, \text{𑀕}, \dots, \text{𑀖}, \dots, \text{𑀗}, \dots, \text{𑀘} \}$ → 417 States

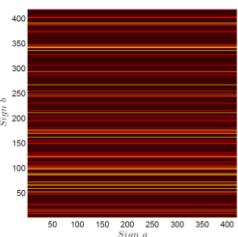
sign	𑀓	𑀔	...	𑀘
𑀓	$P_{1,1}$	$P_{1,2}$...	$P_{1,417}$
𑀔	$P_{2,1}$	$P_{2,2}$...	$P_{2,417}$
...
𑀘	$P_{417,1}$	$P_{417,2}$...	$P_{417,417}$

→ Transition matrix
(417 × 417)

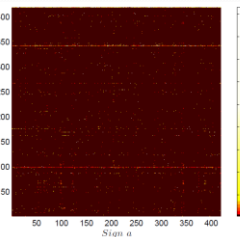
$\Pi = \{ p(X_0 = \text{𑀓}), p(X_0 = \text{𑀔}), \dots, p(X_0 = \text{𑀘}) \}$ → Initial state prob. vector

Structure of Indus Script 52

Comparison of Bigram Matrices



No correlation



Indus Script

Yadav et al. (2009, Submitted)
Structure of Indus Script 53

Conclusion 4:

- a. The difference in the bigram plots indicate presence of correlations between signs.
- b. A simple Markov model (bigram) of Indus texts provides insights into most probable (and improbable) subsequences and the preliminary results are consistent with results discussed earlier.
- c. This bigram model has several interesting applications.

Structure of Indus Script 54

Markov Model of Indus Texts

We use this Markov Model of Indus texts for

- filling-in damaged or illegible Indus texts
- generating Indus like texts
- finding the likelihood of a string to the learned model - how closely the statistical properties of a string matches the texts used for learning the model?





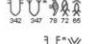







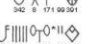

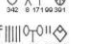











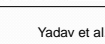

Yadav et al. (2009, Submitted), Rao et al. (2009, PNAS)

Structure of Indus Script 55

a) Filling in damaged & illegible texts

Structure of Indus Script 56







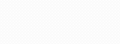

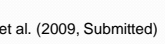
Restoration of Doubtfully Read Texts

Text No.	Text	Incomplete Text	Most Probable Restoration	Probable Restored Sign
8302				
5317				
1193				
1407				
2179				
3396				
8101				

Yadav et al. (2009, Submitted)

Structure of Indus Script 57

Most Probable Texts: Generated by Model

Text Length	Blank Text	Predicted Text	Text No.	Closest matching text from M77
4			1232	
5			2476	
6			1322	






Yadav et al. (2009, Submitted)

Structure of Indus Script 58

b) Generation of Indus like texts

Structure of Indus Script 59

Generated Texts from Model

Yadav et al. (2009, Submitted)

Structure of Indus Script 60

Structure of Indus Script

c) Finding likelihood of a string

61

Structure of Indus Script

Strange Sequences on West Asian Seals



Impression of a round stamp seal from West Asia

Sequence $\text{U} \text{U} \text{U}$

never occurs in Indus corpus of about 4000 texts from Harappan sites!

Use of Indus script to write West Asian content?

62

Structure of Indus Script

Comparing West Asian vs. Indus Area Inscriptions: Difference in structure?

Example	Signs	Likelihood
Synthetic example:		
Indus text	$\text{U} \text{U} \text{U} \text{U} \text{U} \text{U}$ 342 51 336 99 267	2.8×10^{-5}
Altered text	$\text{U} \text{U} \text{U} \text{U} \text{U} \text{U}$ 336 51 99 267 342	~ 0
Actual example:		
Indus seal	$\text{U} \text{U} \text{U} \text{U} \text{U} \text{U}$ 342 53 345 297 150	1.4×10^{-6}
WestAsian seal	$\text{U} \text{U} \text{U} \text{U} \text{U} \text{U}$ 342 342 97 53 178	5.6×10^{-11}

West Asian seal text is approximately 100,000 times as unlikely to be generated by the learned model as other Indus area texts.

Rao et al. (2009, PNAS)

63

Structure of Indus Script

Conclusion 5:

- The Markov model of Indus texts can restore missing signs in damaged or illegible inscriptions.
- It allows stochastic generation of new Indus texts which could help unravel sequential sub-structure in Indus inscriptions.
- It provides a quantitative measure for judging whether an Indus string has structure similar to that of the texts in the training corpus.
- This performance of the model can be further improved by extending it to higher order (till $n=4$).

64

Structure of Indus Script

Comparison with other linguistic & non-linguistic systems

65

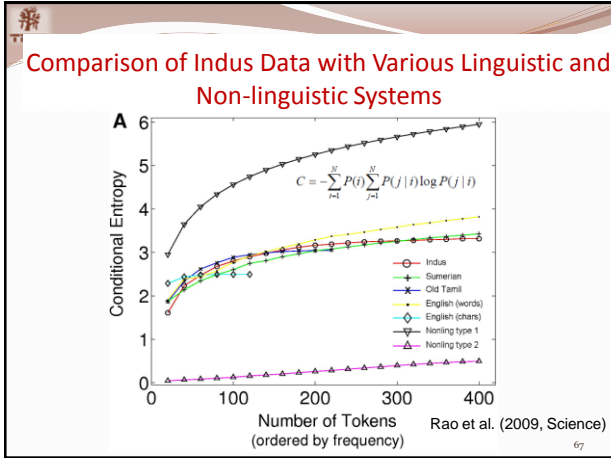
Structure of Indus Script

Comparison with Other Sign Systems

- We have compared the conditional entropy of Indus sign system with other linguistic and non-linguistic sign systems.
- The study *indicates* that the flexibility in choosing a sign given a preceding sign in Indus script is relatively close to linguistic systems than non-linguistic sign systems.

Rao et al. (2009, Science)

66



What does all this imply?

All this shows that

- The script is found on a variety of objects at several sites spanning over a million square kilometers for 700 years.
- The signs were designed carefully and their style was maintained.
- The signs vary in complexity.
- The order of signs is important.
- The signs are written at *specific relative* location within the texts.

Structure of Indus Script

- The sequencing in West Asian seals suggests that the script was versatile enough to be used for their purpose.
- The bigram model can be used to restore damaged or illegible Indus texts.
- Put together, these results indicate that the script has a rich syntax with an underlying logic in its structure.
- We can't read Indus script but (we think) we can WRITE it with some degree of confidence.

Structure of Indus Script

Collaborators

TIFR, Mumbai

- Mayank Vahia

IRC, Chennai

- Iravatham Mahadevan

Univ. of Washington, Seattle

- Rajesh Rao

Hyderabad

- Hrishikesh Joglekar

IMSc, Chennai

- Ronojoy Adhikari
- Sitabhra Sinha
- Raj K Pan

Acknowledgement

<http://www.harappa.com>

Our published work is available at:

<http://www.harappa.com>
<http://www.indusresearch.wikidot.com/script>

Structure of Indus Script

Our Papers

1. N. Yadav, M. N. Vahia, I. Mahadevan, H. Joglekar, *A statistical approach for pattern search in the Indus writing*, International Journal of Dravidian Linguistics, 37, 39 (2008).
2. N. Yadav, M. N. Vahia, I. Mahadevan, H. Joglekar, *Segmentation of Indus texts*, International Journal of Dravidian Linguistics, 37, 53 (2008)
3. M. N. Vahia, N. Yadav, *Harappan Weights, Puratattva*, Vol. 32 (2008)
4. R. P. N. Rao, N. Yadav, M. N. Vahia, H. Joglekar, R. Adhikari, I. Mahadevan, *Entropic evidence for linguistic structure in the Indus script*, Science, 324, 1165 (2009).
5. R. P. N. Rao, N. Yadav, M. N. Vahia, H. Joglekar, R. Adhikari, I. Mahadevan, *A Markov model for the Indus script*, Publications of the National Academy of Sciences, Early Edition (2009)
6. N. Yadav, H. Joglekar, R. P. N. Rao, M. N. Vahia, I. Mahadevan, R. Adhikari, *Statistical analysis of the Indus script using n-grams*, Submitted, 2009
7. S. Sinha, R. K. Pan, N. Yadav, M. N. Vahia and I. Mahadevan, *Network analysis reveals structure indicative of syntax in the corpus of undeciphered Indus civilization inscriptions*, ACL-IJCNLP 2009, pages 5-13
8. N. Yadav and M. N. Vahia, *Classification of patterns on Indus objects*, Submitted, 2009
9. M. N. Vahia and N. Yadav, *Harappan Geometry and Symmetry: A study of geometrical patterns on Indus Objects*, Submitted, 2009

Structure of Indus Script

Thank you!

Structure of Indus Script