Indian Journal of History of Science, 54.2 (2019) 125-134 DOI: 10.16943/ijhs/2019/v54i2/49656

Structure of Indus Script

Nisha Yadav*

(Received 30 January 2019; revised 09 April 2019)

Abstract

The script of the Indus valley civilization has defied decipherment. Several attempts have been made in the past to decipher the script but there is no consensus about its content. The lack of definite knowledge about its structure makes it difficult to objectively evaluate any claim of decipherment. We have tried to fill this lacuna by analyzing the structure of the script using various computational techniques including machine learning and data mining. The focus of our study is to identify patterns in the Indus writing and explore its underlying logic without making any assumptions about its content. The methods identified in the study can also be used to analyse the structure of other undeciphered scripts. In the present paper we summarize our studies of the Indus script.

Key words: Ancient scripts, Computational linguistics, Data mining, Harappan civilization, Harappan script, Indus script, Indus seals, Ligatures, Machine learning, Sign compounding, Sign design, Undeciphered scripts.

1. Introduction

Indus script is a creation of the Harappan civilization that flourished in the north-western parts of India ca. 2600–1900 BCE (Wright, 2010; Agrawal, 2007; Possehl, 2002; Kenoyer, 1998). About 4000 samples of their writing have been discovered from various sites of the Harappan civilization. These include seals, sealings, miniature tablets, copper tablets, bronze implements, ivory sticks, and other miscellaneous objects. Indus script decipherment is difficult due to brevity of the Indus texts (with an average of 5 signs per line of text), lack of information about their spoken language(s), absence of multilingual text(s), paucity of the data and other background information. In spite of several attempts in the past to decipher the Indus script, the problem of Indus script lies unresolved (Possehl, 1996).

We use techniques developed in the field of computer science that can probe specific aspects of various types of data. The objective of our study is to identify the structure and nature of a collection of written material especially when the background knowledge is not enough. We make no assumptions about the nature, content or purpose of the Indus script. The present approach aims to identify the syntactic framework of the Indus script which can be used to evaluate various claims of decipherment (see for example, Yadav *et al.*, 2012). Our studies of the Indus script include: Analyses of the structure of Indus script; Contextual studies of the Indus script; and Study of the design of the Indus signs.

In the following sections we briefly summarise our studies on these aspects of the Indus script.

Dataset: We use a digitized version of the concordance of the Indus writing created by Iravatham Mahadevan in 1977 (Mahadevan, 1977, henceforth referred to as M77). In M77, the Indus signs are indexed from 1 to 417. From M77, we

*Tata Institute of Fundamental Research, Homi Bhabha Road, Navy Nagar, Colaba, Mumbai-400005, Email: y_nisha@tifr.res.in.

removed ambiguous Indus texts and created a filtered dataset EBUDS (for details see Yadav *et al.*, 2008a). EBUDS records 1548 Indus texts and was used in most of our analyses. As a convention followed in the paper, the sign sequences depicted as strings of sign images are to be read from right to left, whereas the sign sequences given as strings of sign numbers are to be read from left to right. In Fig. 1 we have shown an image of an Indus seal from Harappa and a sample of Indus signs from the sign list of M77.

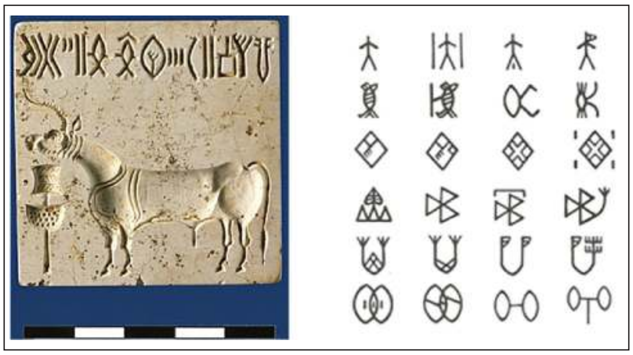


Fig. 1. A large unicorn seal from Harappa on the left and Indus signs from the sign list of M77 on the right. (Copyright Harappa Archaeological Research Project/J.M. Kenoyer, Courtesy Dept. of Archaeology and Museums, Govt. of Pakistan).

2. Analyses of the Structure of Indus Script

The structure of the Indus script was explored in the following studies of the Indus script.

2.1 Sign Frequency Distribution

We studied the sign frequency distribution of the Indus script and found that it follows the Zipf-Mandelbrot law, an empirical law generally followed by various ordered systems (Yadav *et al.*, 2010). Zipf-Mandelbrot law states that

$$\log f_r = a - b \log(r + c)$$

where r is the rank of the sign based on its frequency f_r and a, b, and c are the coefficients of fit. The rank-ordered frequency distribution plot of Indus signs and the Zipf-Mandelbrot fit to the data is shown in Fig. 2. Thus, a small number of signs account for the bulk of the Indus writing and a large number of signs occur rarely in the Indus texts.

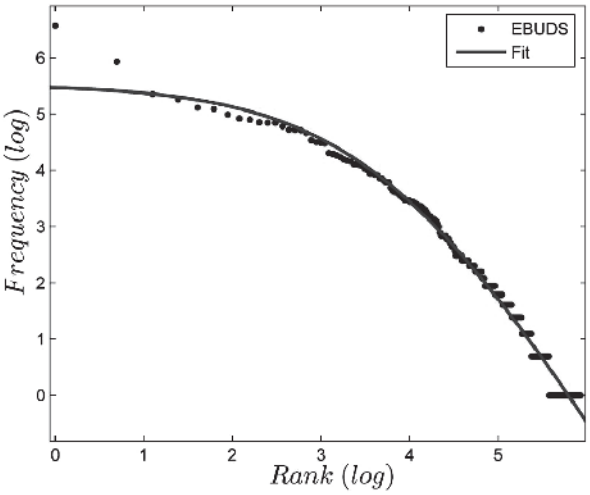


Fig. 2. Rank-ordered frequency distribution of Indus signs f_r plotted against the rank r (Yadav *et al.*, 2010). The sign frequency distribution follows the Zipf-Mandelbrot law, log $f_r = a - b\log(r+c)$, where a, b, and c are the coefficients of fit. Our fitted values are $a = 15.39$, $b = 2.59$ and $c = 44.47$. For English (Brown Corpus), $a = 12.43$, $b = 1.15$ and $c = 100$ (Manning and Schütze, 1999).

2.2 Beginner Ender Asymmetry

The pattern of occurrence of the Indus signs at the beginner and ender positions in the Indus texts was studied using the cumulative frequency distribution plots of text beginners, text enders and all signs (Fig. 3).

We found that there exists an asymmetry in the usage of the signs at the beginner and ender positions in the Indus texts. While just 23 signs account for about 80% of all text enders, around 82 signs account for about 80% of all text beginners. Thus, there is more flexibility in the occurrence of a sign at the text beginner position in the Indus texts than at the text ender position. This is indicative of the presence of syntax in the Indus writing.

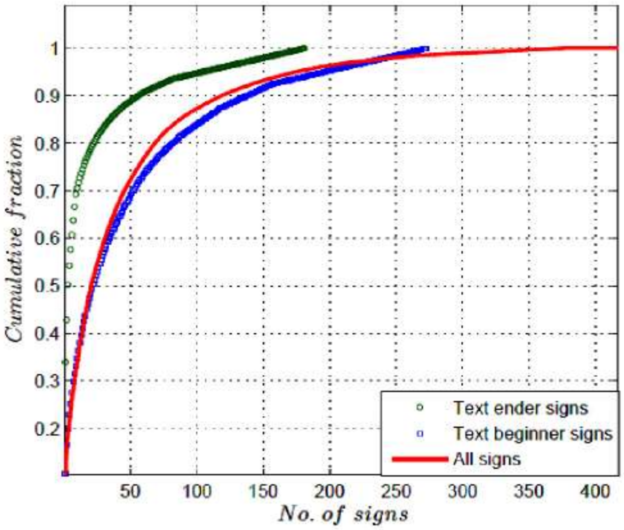


Fig. 3. Cumulative frequency plot for all signs, text beginners and text enders (Yadav et al., 2010). As the direction of the Indus script is from right to left, the signs occurring at the rightmost extreme in the seal impressions are the text beginners and the signs occurring at the leftmost extreme are the text enders.

2.3 Comparison with Randomly Sequenced Dataset

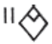
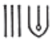
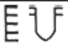
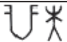
In order to ascertain if the sequencing of the signs in the Indus texts is significant, we compared the sequencing of signs in the Indus script dataset with randomly sequenced datasets (Yadav *et al.*, 2008a). We created ten randomized datasets by randomizing the sequences of signs in the Indus texts and compared the frequency of occurrence of the sign sequences of 2, 3 and 4 signs in the Indus script dataset with the randomized datasets.

We found that sign sequences of 2, 3 and 4 signs (sign pairs, sign triplets and sign quadruplets occurred far more frequently in the Indus script dataset than what is expected by chance. The study affirmed the presence of correlations between signs in the Indus texts.

2.4 Positional Distribution of Sign Sequences

We investigated the positional distribution of the frequent sign sequences of 2, 3 and 4 signs in the Indus texts and found that they have a preferred location in the Indus texts (Yadav *et al.*, 2008a). For instance, 85% of the total occurrences of the most frequent sign pair (267, 99) are at the beginning of the Indus texts and 96% of the total occurrences of the frequent sign pair (342, 176) are at the end of the Indus texts (Table 1).

Table 1. Positional distribution of frequent sign pairs (Yadav *et al.*, 2008a).

Sign pair	Frequency	Solo (%)	Initial (%)	Medial (%)	Final (%)
 99 267	168	0.60	85.71	11.90	1.79
 89 336	75	0.00	10.67	89.33	0.00
 176 342	59	0.00	0.00	3.39	96.61
 342 8	58	1.72	0.00	25.86	72.41
 99 391	56	0.00	91.07	8.93	0.00

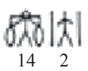
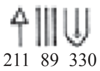
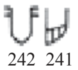
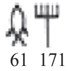
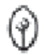
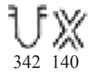
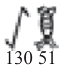
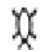
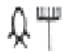
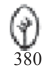
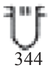
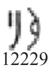
2.5 Segmentation of Indus Texts

In order to explore the possibility of segmenting the longer Indus texts into smaller units, we devised a segmentation scheme which incorporated various approaches of segmentation such as comparing near identical texts, using frequent sign sequences, comparing adjacent sign pair frequencies and using text beginners or text enders (Yadav *et al.*, 2008b). The length of an Indus text in a single line varies from 1 to 14 signs and hence we applied the devised segmentation scheme on all the Indus texts of length ≥ 5 signs. We found that about 88% of Indus texts of length ≥ 5 signs can be segmented into smaller units of 2, 3 or 4 signs (Table 2).

2.6 Statistical Model of the Indus Script

Advances in the fields of computational linguistics, machine learning and data mining have led to techniques such as *n*–gram modelling for studying statistical properties of sequences, pattern recognition and pattern completion. *n*–gram models have found wide use in several fields

Table 2. Examples of segmented Indus texts (Yadav *et al.*, 2008b). The four–digit numbers are the text identification numbers from M77. The alphanumeric sequences above the segments are the markers used for identification of these segments.

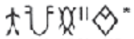
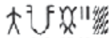
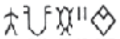
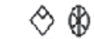
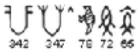
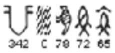
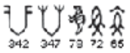
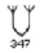
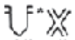
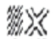
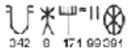
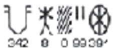
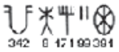
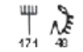
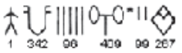
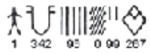
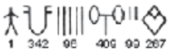
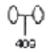
Text number	Text segments						
4254	P53	T148	P116	PM9	389		
	 14 2	 211 89 330	 242 241	 61 171	 380		
	2371	2015	1226				
2537	P41	PM14	67	PM9	389	344	PB1
	 342 140	 130 51	 67	 50 171	 380	 344	 122290
	8001		1093			4305	

where sequences are to be analyzed such as computational linguistics and bioinformatics (Jurafsky and Martin 2008; Manning and Schütze 1999).

An *n*-gram model can identify the correlations that exist between signs $s_{1,} \ldots, s_N$ in a sequence S_N of *N* signs. In a general *n*–gram model, all correlations beyond the (*n–1*) preceding signs are discarded. Conditional probabilities form the core of an *n*-gram model. For a sequence $S_N = s_1 s_2 \ldots s_N$ the *n*–gram model is a specification of conditional probabilities of the form $P(s_N | s_1 s_2 \ldots s_{N-1})$, quantifying the probability that the previous *N–1* signs of the sub string $S_{N-1} = s_1 s_2 \ldots s_{N-1}$ is followed by the sign s_N.

We created an *n*–gram model (bigram model) of the Indus script (Yadav *et al.*, 2010). The bigram model of the Indus script can be used for the restoration of damaged Indus texts, comparison of Indus texts from distinct sites of discovery (or distinct types of objects) and for generation of artificial Indus texts conforming to the structural patterns followed in the Indus writing. We found that the bigram model of the Indus script can predict signs in the damaged or illegible Indus texts with about 75% accuracy (Table 3).

Table 3. Restoration of doubtfully read signs in the Indus texts of M77 (Yadav *et al*., 2010). The signs with asterisk are the doubtfully read signs restored using the bigram model.

Text No.	Text	Incomplete text (one sign removed)	Most probable restoration	Most likely choices for restored sign
8302	 1 342 67 99 287	 1 342 67 99 0	 1 342 67 99 287	 287 391
5317	 342 347 78 72 65	 342 0 78 72 65	 342 347 78 72 65	 347
1193	 342 149	 0 149	342 149	342
1407	 342 8 171 99 391	 342 8 0 99 391	 342 8 171 99 391	 171 48 67
2179	 1 342 98 409 99 287	 1 342 98 0 99 287	 1 342 98 409 99 287	 409

2.7 Comparison of Conditional Entropy with other Sign Systems

The flexibility in the usage of signs given a preceding sign can be quantified using an information theoretic measure called the conditional entropy. The conditional entropy $H(J|I)$ of any token j following any token i is defined as

$$H(J|I) = -\sum_{i=1}^{N} P(i) \sum_{j=1}^{N} P(j|i) \log P(j|i)$$

We compared the flexibility in the usage of signs in the Indus script with sequences from various linguistic and non-linguistic systems including English, Sanskrit, Old Tamil, Sumerian, DNA, Protein, and Fortran (Rao *et al*., 2009a). We found that the conditional entropy of the Indus script falls within the range of linguistic systems included in the study (Fig. 4).

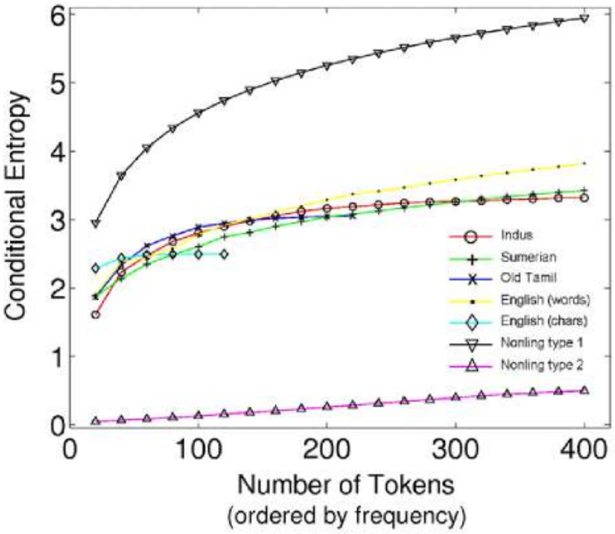


Fig. 4. Comparison of conditional entropy of Indus texts with other linguistic and non-linguistic systems (Rao *et al*., 2009).

2.8 Clustering Indus Texts

Unsupervised machine learning techniques such as clustering (Jain and Dubes, 1988) can be used to explore an undeciphered script for the presence of substructures. The technique of clustering is used in several fields including machine learning, pattern recognition, information retrieval and image analysis. Using a clustering technique (K-means) we explored the non-contiguous associations between the signs in the Indus texts and clustered the Indus texts such that the texts in each of these clusters were more similar amongst themselves than to texts belonging to other clusters (Yadav *et al*., 2017).

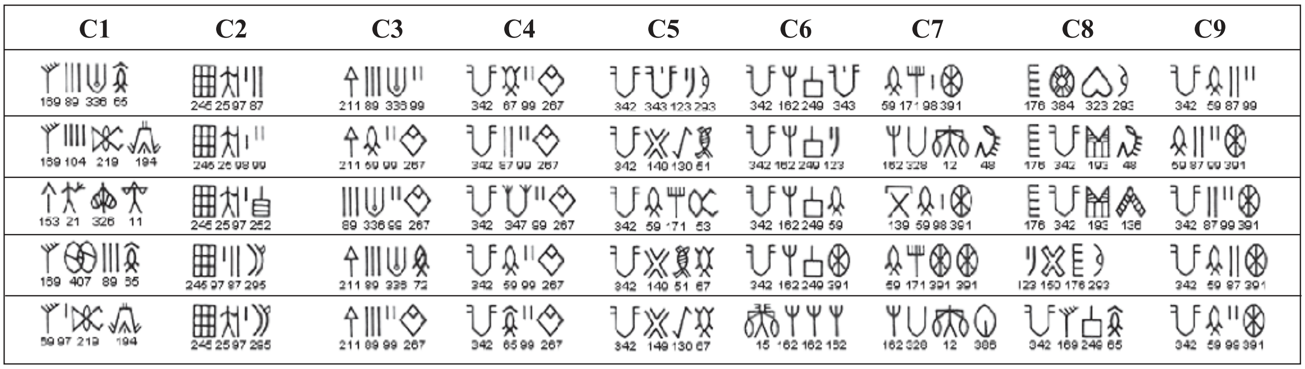
In order to cluster the Indus texts, we created a term-document matrix for the Indus texts. Here, terms corresponded to the individual signs and documents to the distinct Indus texts in the Indus script dataset. Each element in this matrix indicated the frequency with which a sign occurs in the text. This term-document matrix was subjected to K-means clustering with *cosine* as a measure of similarity. The K-means clustering routine permits creation of an arbitrary number of clusters (K). We analyzed the clusters by varying the K-value and evaluated the contents of the clusters in each case. We found with $K = 9$, the boundary separating each cluster was distinct with respect to its constituent texts. We therefore extracted the nine clusters (C1 to C9) of Indus texts. These nine clusters were found to have their signature set of signs and sign sequences (Table 4).

The study suggested that the Indus writing had distinct styles or contents. The text clusters were not found to have any significant correlation to the sites of discovery or object types.

3. Contextual Studies of the Indus Script

The Indus script occurs on distinct types of objects discovered at various sites of the Indus Valley civilisation. The script is often associated with distinctive motifs or geometrical patterns. We performed the classification and analysis of these individualistic patterns on the Indus seals and other inscribed material in Yadav and Vahia (2011a). A detailed analysis of some of their geometric and symmetric patterns is included in Vahia and Yadav (2010) and Sinha *et al.* (2011). We studied the

Table 4. Frequent sign quadruplets (contiguous/non-contiguous) in clusters C1 to C9 (Yadav et al. 2017).



C1	C2	C3	C4	C5	C6	C7	C8	C9
169 89 336 65	245 25 97 87	211 89 336 99	342 67 99 267	342 343 123 293	342 162 249 343	59 171 98 391	176 384 323 293	342 59 87 99
169 104 219 194	245 25 98 99	211 59 99 267	342 87 99 267	342 140 130 51	342 162 249 123	162 328 12 48	176 342 193 48	59 87 99 391
153 21 326 11	245 25 97 252	89 336 99 267	342 347 99 267	342 59 171 53	342 162 249 59	139 59 98 391	176 342 193 136	342 87 99 391
169 407 89 55	245 97 87 295	211 89 336 72	342 59 99 267	342 140 51 67	342 162 249 391	59 171 391 391	123 150 176 293	342 59 87 391
59 97 219 194	245 25 97 295	211 89 99 267	342 65 99 267	342 149 130 67	15 162 162 152	162 328 12 386	342 169 249 65	342 59 99 391

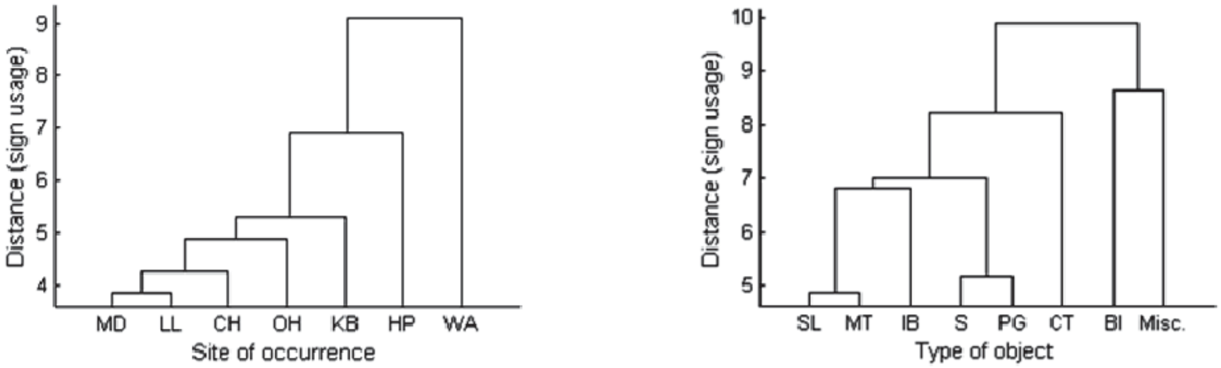


Fig. 5. On the left, similarity between sites based on the usage of signs on inscribed material discovered from the sites (MD: Mohenjo-Daro, HP: Harappa, LL: Lothal, CH: Chanhu-Daro, KB: Kalibangan, OH: Other Harappan sites, WA: West Asian sites). On the right, similarity between different types of objects based on the usage of signs on inscribed material of different types (S: Seal, SL: Sealing, CT: Copper Tablets, MT: Miniature Tablets, PG: Pottery Graffiti, BI: Bronze Implements, IB: Ivory or bone rods).

variation in the Indus writing on objects found at various sites and also across different types of objects in Yadav 2013. In this study, we compared the usage pattern of the Indus signs across various sites or types of objects (Fig. 5).

We found that Mohenjo-Daro and Lothal share a high level of similarity based on the pattern of usage of Indus signs (Fig. 5). The usage of Indus signs at Harappa and West Asian sites was found to be quite distinct from other sites. Seals were found to share a high level of similarity with pottery graffiti while sealings and miniature tablets were found to be similar based on the usage of signs (Fig. 5). We emphasize the need to understand the level of non-uniformity in the Indus script against conditions where uniformity seems to be the norm. For instance, signs belonging to the set of the 67 most frequent signs which account for 80% of occurrences in M77, account for almost similar percentage of data for most of the sites and object types. However, the relative contribution of each of these 67 signs fluctuates across various sites and types of objects. It is therefore important to study these fluctuations at different sites and on distinct object.

Mohenjo-Daro has largest percentage of inscribed objects in the form of seals, while Harappa has comparable percentage of seals, sealings and miniature tablets. It is normally assumed that sealings (being impressions of seals) were created using seals and hence we must find the seal for every sealing. However, while it has been known that not many seals corresponding to the sealings have been found, the statistical study showed that in fact, in terms of usage of signs, seals and sealings are quite different. This implies that the seals used to make the sealings are not present in the set of objects that have survived and discovered from various sites.

The contextual study of the Indus script suggests that while there is a common thread of rules and grammatical structures that are well obeyed in the Indus writing, usage of signs on distinct types of objects at different sites do provide individualistic clues to their content. Further studies on the use of various other motifs on the inscribed objects, the changes in the designs of signs and stratigraphical studies of the inscribed objects will add more clarity to this problem.

4. Study of the Design of Indus Signs

Indus signs vary in the complexity of their design. In a study analysing the design of the Indus signs (Yadav and Vahia, 2011b), we identified three types of design elements of the Indus signs: basic signs (154 in number), provisional basic signs (10 in number) and modifiers (21 in number). Provisional basic signs do not have an independent occurrence in the sign list. They only appear in the designs of certain Indus signs where they are compounded with other design elements.

Based on the complexity of their design, we classified the Indus signs into two categories: Basic signs and Composite signs (Fig. 6). Composite signs were further classified into: Compound signs (composite of basic signs) and Modified signs (signs modified by modifier).

In order to check if the compound signs are a compact version of their constituent basic signs, we compared the pattern of usage of the compound signs and their respective constituent sign sequences. We found that the compound signs are not a compact version of their constituent sign sequences. They seem to have some other function in the Indus script.

While the visual form of some of the Indus signs can be associated with familiar natural or artificial entities, a large number of Indus signs are characterized by a high level of abstraction in their design. It has been suggested that the design configuration across various types of visual signs are deeply influenced by the shapes that the designers of the signs encounter in their daily lives rather than the ease of writing (Changizi *et al.*, 2006). This can also be seen in the case of Indus script suggesting that the Indus writers were keen on conveying their ideas or information unambiguously rather than placing emphasis on the ease of writing.

The designs of the Indus signs imply logic and creativity in their structure. The designers of the Indus signs placed a special emphasis on symmetry and there seems to be an underlying effort to retain the overall aesthetic sense of the

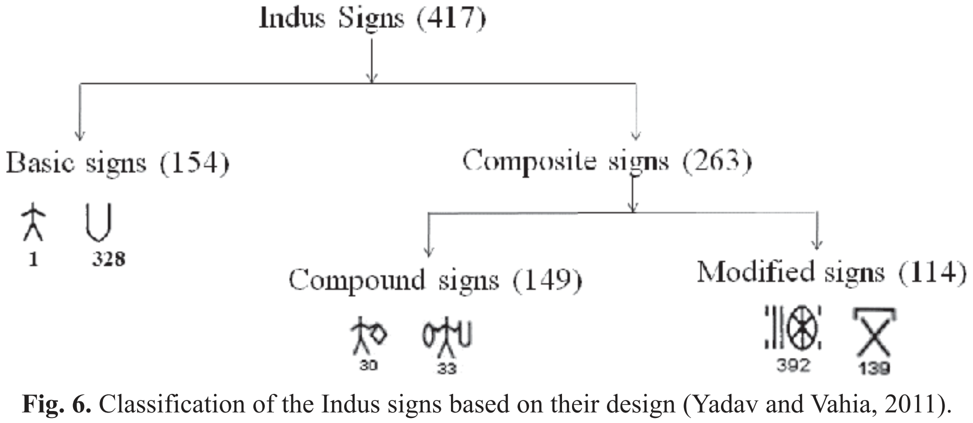


Fig. 6. Classification of the Indus signs based on their design (Yadav and Vahia, 2011).

Indus signs. About 60% of the signs conform to either vertical or horizontal symmetry (Yadav and Vahia, 2011). The signs of the Indus script seem to incorporate various techniques in their design that were used in several ancient writing systems to maximize the usage of a limited number of signs. These include sign compounding as in Chinese writing (Bottéro, 2004), conflation of signs as in Mayan glyphs (Coe, 1992) and having signs serving as determinatives as was the practice in Egyptian Hieroglyphs (Baines, 2004). In later scripts in India, merger of signs was used to combine vowels into consonants. However, we do not wish to hazard a guess on the connection or otherwise of the Indus script with other writing systems.

A lot of thought, planning and utility issues have been taken into consideration while designing the Indus signs. The Indus civilization was spread over an area of about a million square kilometers and yet, the sign list over the entire civilization was identical. This indicates that the signs, their meaning and their usage were agreed upon by people spread over a large area. This arrangement worked satisfactorily for about 700 years. Hence, the understanding of the Indus signs and their meaning must have been robust, yet versatile and easy to use.

5. Discussion and Conclusions

We have employed a series of computational methods and statistical tests on the dataset of the Indus script. Our studies suggest that the Indus writing is highly structured. The sign frequency distribution of the Indus script follows *Zipf-Mandelbrot* law, an empirical law followed by various ordered systems. There exists an asymmetry in the pattern of usage of text beginners and text enders in the Indus texts. A few signs constitute the text enders while relatively large number of signs occur as text beginners. The signs in the Indus texts, while following the standard pattern of usage as in any ordered system, have some significant characteristics such as (i) the most frequent sign in the Indus writing is a text ender (sign number 342), (ii) the second most frequent sign (sign number 99) generally follows text beginner signs and (iii) the third most frequent sign is a text beginner (sign number 267).

Indus sign sequences of 2, 3 or 4 signs occur with far higher frequency than what is expected by chance and they have a preferred location in the Indus texts. Structural analysis shows that the Indus texts tend to have three primary constituent units: Beginner units, Middle units and Ender units. Each of these units may have one or more signs. While a large number of signs are allowed to begin the Indus texts, the beginner unit most often has no more than 2 signs while the ender unit may have as many as 3 signs suggesting that the completion of information was reinforced with additional information. The middle unit has maximum flexibility in its sign usage and it seems to carry a large variety of information as inferred by the number of signs appearing in this unit. It is possible to identify pairs of Indus signs that occur together in the longer Indus texts but in general do not have affinity to each other. Using this insight, it is possible to revisit the entire corpus of Indus script and we found that the longer Indus texts can be segmented into smaller units (Yadav *et al*., 2008b).

A bigram model of the Indus script based on nearest neighbor associations can restore signs in illegible Indus texts with about 75% accuracy. The Indus script is versatile enough to permit writing of differently coded information as can be seen from the texts on the Indus seals found at West Asian sites having a distinct pattern of sign usage. Comparison of the flexibility in sign usage suggests that the usage of signs in the Indus writing is as flexible as for natural linguistic systems and is more flexible than artificial linguistic systems (computer languages). However, the usage of signs in the Indus writing is less flexible in comparison to the systems in which abstractions are conveyed (music) or the manner in which biological information is stored (DNA or Protein).

Unsupervised clustering of the Indus texts provides nine robust clusters, each with a characteristic set of signs and sign groups. While a variety of writing material is used for Indus writing, most frequent inscribed objects are seals followed by sealings. For reasons that are not clear, the original seals of most sealings (seal impressions) have not been found. Similarly, sealings of the recovered seals are also rare.

Based on their design, the Indus signs can be classified into two major categories: Basic signs and Composite signs. Composite signs can be further classified into: Compound signs and Modified signs. Statistical analysis shows that the compound signs are not a 'short hand' or space saving device since the environment (in terms of the signs preceding or following them) in which the compound signs occur in the Indus texts is unlike that of its constituents in any combination.

Any proposed interpretation of the Indus script should be able to explain these characteristics. A successful decipherment of the Indus script will provide us a unique window to understand this intricate and ingenious creation of the Harappan people.

BIBLIOGRAPHY

Agrawal, D P. *The Indus Civilization: An Interdisciplinary Perspective*, Aryan Books International, New Delhi, 2007.

Baines, J. The Earliest Egyptian Writing: Development, Context, Purpose, in *The First Writing: Script Invention as History and Process,* S. D. Houston, Ed., Cambridge University Press, Cambridge, 2004, pp. 150–189.

Bottéro, F. Writing on Shell and Bone in Shang China, in *The First Writing: Script Invention as History and Process*, S. D. Houston, Ed., Cambridge University Press, Cambridge, 2004, pp. 250–261.

Changizi, M A; Zhang, Q; Ye, H; Shinsuke, S. The Structures of Letters and Symbols Throughout Human History are Selected to Match Those Found in Objects in Natural Scenes, *The American Naturalist*, 167 (2006).

Coe, M D. *Breaking the Maya Code*, Thames and Hudson, New York 1992.

Jain, A K and Dubes, R C. *Algorithms for Clustering Data*, Upper Saddle River, NJ, Prentice-Hall, Inc. USA, 1988.

Jurafsky, D and Martin, J H. *Speech and Language Processing: An Introduction to Speech Recognition, Computational Linguistics and Natural Language Processing*, 2nd ed. Pearson Prentice Hall, New Jersey, 2008.

Kenoyer, J M. *Ancient Cities of the Indus Valley Civilization*, Oxford University Press, Oxford, 1998.

Mahadevan, I. *The Indus Script: Texts, Concordance and Tables*, Archaeological Survey of India, New Delhi, 1977.

Manning, C and Schütze, H. *Foundations of Statistical Natural Language Processing*, MIT Press, Cambridge, 1999.

Possehl, G L. *Indus Age: The Writing System*, Oxford and IBH Publishing Co. Pvt. Ltd., New Delhi, 1996.

Possehl, G L. *The Indus Civilization: A Contemporary Perspective*, Vistaar Publications, New Delhi, 2002.

Rao, R P N; Yadav N; Vahia, M N; Joglekar, H; Adhikari, R and Mahadevan, I.. Entropic Evidence for Linguistic Structure in the Indus Script, *Science*, 324 (2009):1165.

Rao, R P N; Yadav, N; Vahia, M N; Joglekar, H; Adhikari, R and Mahadevan, I. Entropy, the Indus Script and Language: A Reply to R Sproat, *Computational Linguistics*, 36 (2010):795–805.

Sinha, S; Yadav, N and Vahia, M N. In Square Circle: Geometric Knowledge of the Indus Civilization, in *Math Unlimited: Essays in Mathematics*, R Sujatha; H N Ramaswamy and C S Yogananda, Eds., Science Publishers, Enfield, 2011, pp. 451–462.

Vahia, M N and Yadav, N. Harappan Geometry and Symmetry: A Study of Geometrical Patterns on Indus Objects. *Indian Journal of History of Science*, 45.3 (2010):343–368.

Wright, R P. *The Ancient Indus – Urbanism, Economy and Society*, Cambridge University Press, New York, 2010.

Yadav, N; Vahia, M N; Mahadevan, I and Joglekar, H. A Statistical Approach for Pattern Search in Indus

Writing, *International Journal of Dravidian Linguistics*, XXXVII (2008a):39-52.

Yadav, N; Vahia, M N; Mahadevan, I and Joglekar, H. Segmentation of Indus Texts. *International Journal of Dravidian Linguistics*, XXXVII (2008b):53–72.

Yadav, N; Joglekar, H; Rao, R P N; Vahia, M N; Adhikari, R and Mahadevan, I. Statistical Analysis of the Indus Script Using *n*-grams. *PLoS ONE*. 5 (2010).

Yadav, N and Vahia, M N. Classification of Patterns on Indus Objects, *International Journal of Dravidian Linguistics*, 40 (2011a): 89–114.

Yadav, N and Vahia, M N. Indus Script: A Study of its Sign Design, *Scripta*, 3 (2011b):133–172.

Yadav, N; Rao, R P N. and Vahia, M N. Indus Script, *Current Science*, 103.11(2012b):1265–1266.

Yadav, N. Sensitivity of Indus Script to Site and Type of Object, *Scripta*, 5 (2013):67–103.

Yadav, N; Salgaonkar, A and Vahia, M N. Clustering Indus Texts using *K*-means, *International Journal of Computer Applications*, 162 (2017):16–21.