ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/373522673

A Catalog of Indus Signs

Book · July 2023

CITATIONS
2

READS
2,084

1 author:

Andreas Fuls
Technische Universität Berlin

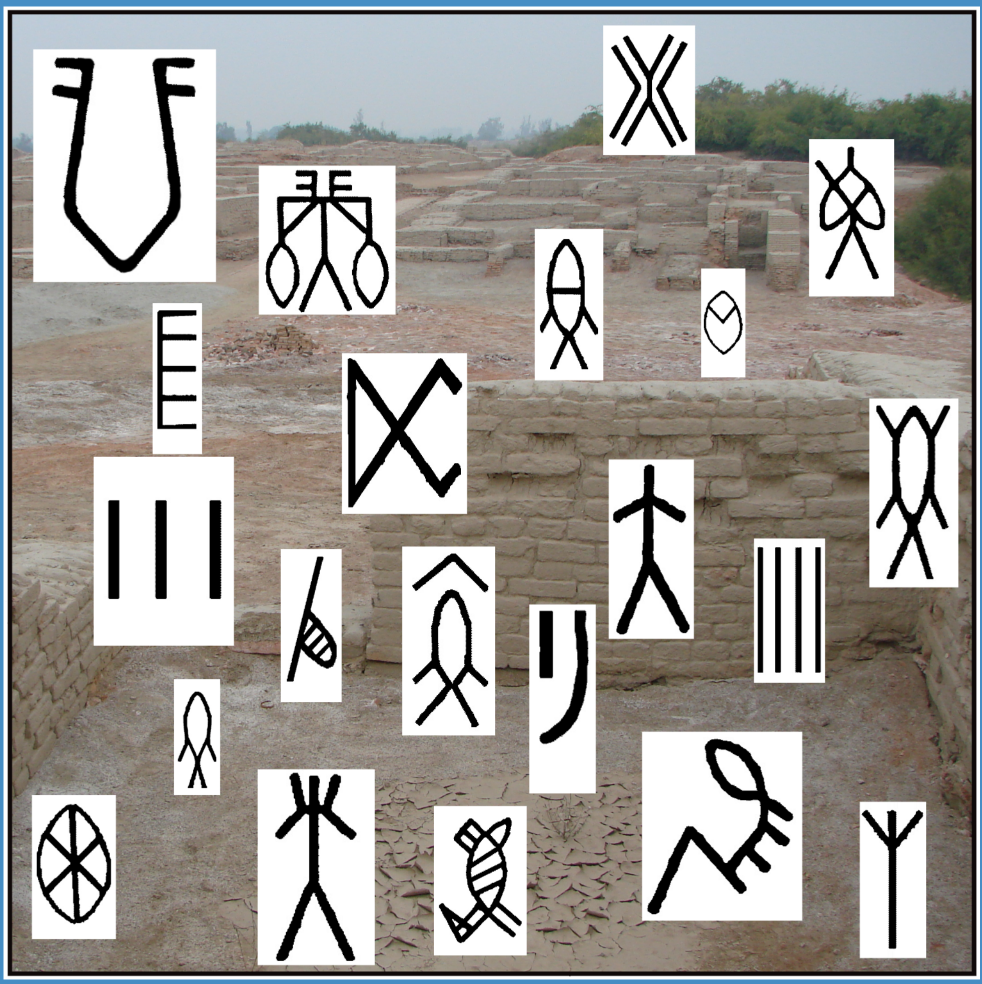
23 PUBLICATIONS 43 CITATIONS

SEE PROFILE

All content following this page was uploaded by Andreas Fuls on 31 August 2023.

The user has requested enhancement of the downloaded file.

A Catalog of Indus Signs



Andreas Fuls

A Catalog of Indus Signs

Mathematica Epigraphica

No. 4

Editor

Andreas Fuls

A Catalog of Indus Signs

Andreas Fuls

Berlin 2023

Fuls, Andreas:
A Catalog of Indus Signs/Andreas Fuls.- Berlin: Self-publishing (Mathematica Epigraphica, No. 4), 2023
ISBN: 979-8-398-42230-6
Independently published

Copyright © 2023 by Andreas Fuls, Berlin (Germany)
All rights preserved.
ISBN: 979-8-398-42230-6

Contents

List of Figures

List of Tables

Forword

This volume is part of a series analysing Indus writing. In 2008 an online epigraphic database of Indus inscriptions (ICIT) was developed in cooperation with Dr. Bryan K. Wells. At the same time new mathematical methods and tools were developed in order to analyse the statistical and positional behaviour of Indus signs, to map the spatial distributions of signs and inscribed artefacts, and to define sets of signs with similar positional behaviour.

The first volume called *Corpus of Indus Inscriptions* lists all Indus inscriptions that are currently available and presents the temporal as well as spatial distribution of inscribed artefacts. The inscriptions are indexed in the volume by their identifier in the ICIT database, the CISI number, or the excavation number.

The second volume called *A Catalog of Indus Signs* presents a description of each sign, its frequency and positional behaviour. All sign occurrences on inscribed artefacts are listed in the volume with reference to sign sequences on inscribed artefacts in the volume *Corpus of Indus Inscriptions*. Here in special sections, the type of the Indus writing system, the 'core' signs at different sites and on various artefact types, and the expected number of new signs that may be discovered in the future are discussed. The book presents a new method of positional sign analysis that accounts for different text lengths and makes it possible to classify Indus signs according to their positions within the texts.

The two volumes can be used concurrently, since the same sign numbers and text identifiers refer to each other. They represent the most up-to-date corpus of Indus inscriptions and will be updated in the future when new Indus inscriptions are found.

Acknowledgments

The present book contains a most complete collection of signs of the Indus writing system, which was first started about 30 years ago by Dr. Bryan K. Wells when he was writing his Master's thesis at the University of Calgary and later continued in the context of his doctorate at Harvard University. Since 2009, the resulting database of Indus texts has been continuously expanded in our joint cooperation and made available to the public as an online database. I am also very grateful to Dr. Wells for our discussions in the course of which a lot of details in the inscriptions were clarified and elaborated. Without his preliminary work and his invaluable suggestions, the present work would not have been possible.

All inscriptions and the encoding of the signs have been checked on the basis of photographs. I would like to express special thanks to Prof. Dr. Asko Parpola, Dr. Dennys Frenez, Prof. Dr. Farzand Masih and Dr. Hansmukh Seth for supporting this work by giving me access to the relevant materials.

With numerous inscriptions and associated metadata, it was next to impossible that errors should not creep in from time to time. Many of these inconsistencies in the database have been discovered through the careful eye of database users, especially Pallavee Gokhale and Prof. Dr. Rama Viswanathan. I would like to thank them most warmly for their suggestions and corrections.

A concordance of the inscriptions in the ICIT database with the texts in the Mahadevan corpus (IDF80) has been achieved in collaboration with Prof. Dr. Rama Viswanathan. Thus, it is now possible to compare both text corpora and to provide a concordance with Mahadevan's sign list, for which I am very grateful.

I would also like to thank Irina Poljakova for reviewing the manuscript and making linguistic corrections, so that I hope the reader will find this book a comprehensible introduction to the writing system of the Indus culture.

Chapter 1

Introduction

1.1 How many Indus signs are there?

Since the excavations in Harappa and Mohenjo-daro (Mackay (1998); Marshall (1931); Vats (1940)), Indus signs have been tabulated and distinguished according to graphic features (graphemes). As a result, there were several sign lists published by Hunter (1934), Meriggi (1934), Fairservis (1992), Rao (1982), Dani (1963), Mahadevan (1977) and Parpola (1994). Hunter (1934) lists 234 distinct signs based on 750 objects, the corpus collected by Mahadevan (1977) comprises 2911 inscribed objects with 419 distinct signs and the sign list proposed by Parpola (1994) has 386 signs. However, the question remains: which of these are distinct signs and which are allographs (signs with the same meaning)? The identification of allographs is a time-consuming task with the search for replacement patterns of signs and it requires a large text corpus. This can be performed through structural analysis (developed by Forrer (1932)) applied by Wells (2006, pp. 67–93) to create a detailed sign list with 676 signs. For those who believe that the Indus sign list should be as short as possible, Wells' sign list may look like a step backwards. Until no other evidence can be found to distinguish an allograph from other signs, this certainly is the most secure way forward, especially for low frequency signs. The future goal I set for myself is to gradually reduce the sign list, analysing the behaviour of each sign in order to identify allographs and graphic variants.

The present research is based on Wells' updated list of signs (Wells, 1998, 2006, 2011, 2015) with 713 signs (last update: May 2023, version 2.9). Inscriptions are stored in the ICIT database (Interactive Corpus of Indus Texts) with a total of 4674 inscribed artefacts. Since some artefacts are inscribed on more than one side, there are in total 5659 texts and 19869 sign occurrences. 3664 texts are complete with 13695 sign occurrences. Only complete texts can be used for the analysis of the position of signs within each text (Chapter 3).

1.2 Type of writing system

The primary task is to classify the type of the Indus writing system. The different number of signs is a first indication which allows us to distinguish between alphabets, syllabic and logographic writing systems. Different sign lists have been suggested, ranging from only 62 signs (Rao, 1982) up to 694 signs (Wells, 2015). The sign list by Mahadevan (1977) with 419 signs and by Parpola (1994) with 386 signs fall within this range. The question is which graphemes are distinct signs and which are allographs? Because of the low frequency of many signs the problem remains unsolved, since a detailed structural analysis of each sign requires an adequate sample size. Therefore, graphemes should be regarded separately as long as they cannot, with certainty, be identified as sign variants (allographs). As a result a preliminary sign list comprises about 713 signs (Fuls, 2020). Most scholars agree that the Indus script represents a mixed logographic-syllabic writing system, which can be confirmed by comparing the sign frequency distribution with known writing systems (Chapter 4). But how many syllables and logograms are there and how to identify them? This question is discussed in Chapter 4.

To estimate the percentage of syllables and logograms in a sign list, the sign frequency distribution can be compared with that of known writing systems, i.e. logographic, syllabic, logographic-syllabic systems, abugidas, abjads, and alphabets (Fuls, 2015c,d). The analysis shows that the sign list of Indus writing should include from about 10% to 20% of syllables and from 78% to 90% of logograms (Chapter 4.5). In the corpus, however, we may expect about 75% of syllables and 25% of logograms. This means that, on average, every forth sign might be a logogram with a preferred position depending on context, logographic meaning, and syntax. Furthermore, the analysis provides an average word length of 1.7 signs for Indus writing. The latter issue will be discussed in detail in a separate book devoted to the analysis of Multivariate Segmentation trees of Indus inscriptions (Fuls, 2015a).

1.3 Reading Direction

Most Indus sign sequences should be read from right to left. There are only a few exceptions when an inscription is read from left to right, top to bottom, or boustrophedon. But about 75% of all 5659 texts are read from right to left. For about 14% of the texts the reading direction is unknown, and about 4% of the texts are read from left to right. 16 texts should be read from top to bottom and only 10 texts are written boustrophedon. The reading direction of Indus sign sequences is discussed in detail in Chapter 2.2 of the book *Corpus of Indus Inscriptions* (Fuls, 2022).

1.3.1 Epigraphic indications of the reading direction

Based on the assumption that the writing direction is the same as the reading direction, the reading direction is often indicated by blank space at one end of a sign sequence. This is, for example, the case on square seal H-9 with blank space at the left end of the seal impression. In some cases, signs are squeezed on the left side of inscriptions because of the lack of space (e. g. square seal M-66 and K-15). On incised pottery we may find incised signs overlapping the previous sign located on the right side, which also shows that Indus inscriptions were written from right to left (e. g. pottery fragment K-100).

1.3.2 Statistical method to determine the reading direction

There are several methods available to determine the reading direction of undeciphered writing systems. Besides epigraphical methods, Ashraf and Sinha (2018) published a statistical method that is comparing the probability of occurrence of signs at the left and at the right ends of words. They showed that the difference in the Gini index $dG = G(L) - G(R)$ is strongly related to the reading direction of known writing systems. Reading from left to right in English, Tamil, or other languages leads to a negative value of dG. By contrast, Hebrew and Arabic reveal a positive difference in the Gini index since they must be read from right to left.

The method can also be applied to Indus writing, assuming that complete sign sequences (texts) start and end with one word. For the purpose of the analysis it is not necessary to segment longer texts into words. Based on this assumption, they are able to prove that dG is greater than zero and, therefore, most Indus sign sequences must be read from right to left (Figure 1.1).

According to the method, only complete sign sequences are used, while multiple lines on the same side of inscribed artefacts are split into independent sign sequences. It also verifies the assumption that the writing direction is the same as the reading direction.

1.4 The ICIT Database

The database is a collection of Indus inscriptions published in Volumes 1, 2, 3.1 and 3.2 of the *Corpus of Indus Seals and Inscriptions* (Joshi and Parpola, 1987; Shah and Parpola, 1991; Parpola et al., 2010, 2019) and other publications (Mackay, 1998; Marshall, 1931; Meadow and Kenoyer, 2001; Jansen and Urban, 1985; Vats, 1940; Bisht, 2015). It is called after an earlier digital collection which has been developed by Dr. Bryan Wells since 1998: Interactive Concordance of Indus Texts (ICIT).

All data are stored in a relational database. The database contains each text

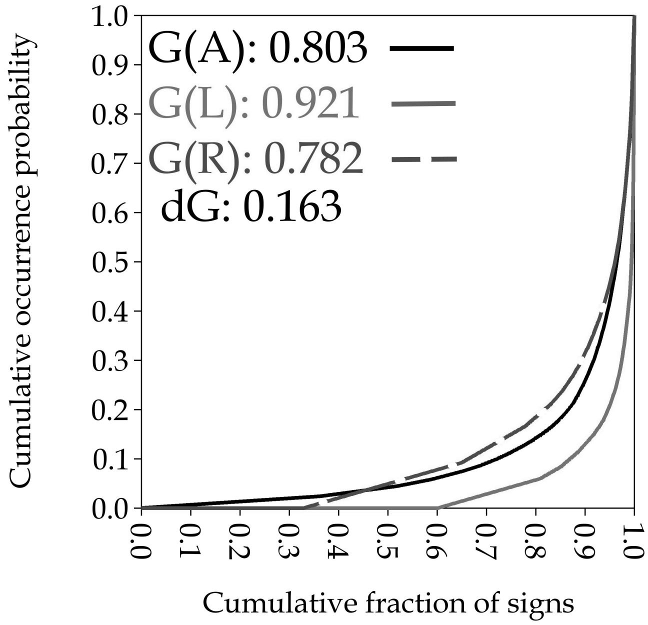


Figure 1.1: Gini indexes of Indus sign sequences indicating a reading direction from right to left, since $dG = G(L) - G(R)$ is greater than zero.

as a sequence of sign numbers referring to the sign list (Chapter 2.1). The general and archaeological information such as site, artefact type, material, size, depth, excavation number and the location where it was found is stored for each artefact. The CISI number from the *Corpus of Indus Seals and Inscriptions* is also listed, if available.

Every artefact in the database has a unique identifier (ICIT ID, also called ID). The identifier of each text starts with the ICIT ID of the artefact (multiplied by ten). The last digit of the text identifier falls between zero and nine and is used to distinguish the texts on multi-sided artefacts. All sign occurrences are given in the description of a sign (Chapter 5.1) referring to the ICIT ID and thereby to the inscription on each artefact listed in the *Corpus of Indus Inscriptions* (Fuls, 2022).

The database contains additional information for each text: the number of signs, the number of lines, the type of the text, the reading direction, and the completeness of a text (the options are 'yes', 'no', or 'unknown'). If there is at least one sign not legible or eroded, the text is not complete. But even if all the signs can be identified, the text might be incomplete on a fragmentary artefact and thus classified as 'unknown' in terms of its completeness.

The database is available online and free to use for personal and scientific

Set 01:	**1** (228)	**2** (867)	**3** (260)	**4** (99)	**5** (49)	**6** (7)	**7** (6)
12 (3)	**13** (33)	**14** (6)	**15** (8)	**16** (48)	**17** (79)	**18** (7)	**19** (6)
20 (2)	**25** (3)	**26** (1)	**27** (6)	**28** (5)	**29** (4)	**31** (251)	**32** (586)
33 (523)	**34** (185)	**35** (35)	**36** (7)	**37** (2)	**39** (2)	**41** (2)	**42** (2)
43 (1)	**44** (1)	**45** (2)	**46** (2)	**47** (1)	**48** (18)	**49** (6)	**50** (1)
51 (1)	**55** (63)	**56** (10)	**57** (1)	**58** (1)	**59** (2)		
Set 02:	**60** (220)	**61** (94)	**62** (3)	**63** (13)	**64** (14)	**65** (15)	**66** (17)
67 (8)	**68** (1)						
Set 03:	**69** (1)	**70** (40)	**71** (4)	**72** (14)	**73** (2)	**74** (1)	**75** (1)
80 (2)	**81** (3)	**82** (2)	**83** (1)	**84** (1)	**85** (1)		
Set 04:	**90** (197)	**91** (28)	**92** (2)	**93** (7)	**94** (2)	**95** (29)	**96** (4)

SP: absolute sign position

L: text length

NL: normalised text length (here $NL = 10$ signs)

MINP: minimum of relative (normalised) text position

MAXP: maximum of relative (normalised) text position

The result MINP is rounded down and MAXP is rounded up to integer values in the range [1,NL]. They represent the minimum and maximum values of the relative sign position, which is the normalised sign position of text length NL.

After normalisation the sign position of a short text might be counted more than once (like Sign 002 falling into relative sign positions 9 and 10, see Figure 3.1). To avoid an overweighing of sign positions coming from short texts each sign position is weighted linearly in relation to the text length. This means that signs of short texts get a lower weight than signs of longer texts. The weighing is calculated as follows:

$$W = L/NL \tag{3.3}$$

L: text length

NL: normalised text length

W: weight

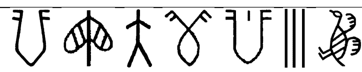
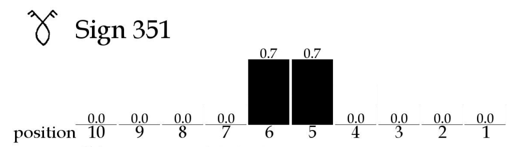
The formula implies that each sign in a text of NL signs gets a weight of 1, whereas longer texts have a higher precision of each sign position and get a higher weight. By contrast, shorter texts get a smaller weight, since their sign position is less precise. Only complete texts with a text length of at least two signs are used for the calculation; texts with only one sign do not convey any positional information and are, therefore, left aside.

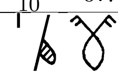
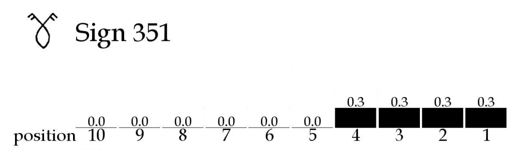
3.2.2 Histogram

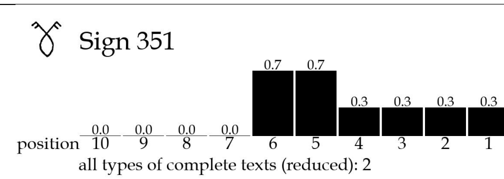
The histogram shows the frequency of sign positions for a constant number of intervals. Each interval represents a normalised sign position from 1 to NL. In the calculations given below the normalised text length NL is assigned to 10. To calculate the histogram for each sign, the weight is added to the relative sign position(s) between MINP and MAXP. For example, Sign 002 in a text with five signs (text length $L = 5$) has a weight of 0.5 and is counted twice, for relative sign positions 9 and 10, respectively. By contrast, Sign 003 in a text of length $L = 12$ has a weight of 1.2 which is added to position 1 (Figure 3.1).

Examples depicted in Tables 3.2 and 3.3 will illustrate the method step-by-step. To calculate the histograms of Signs 351 and 720, respectively, all complete, single-line texts in which the sign occurs are used.

Table 3.2: Process of calculating the normalised weighted sign position histogram of Sign 351 occurring altogether in two texts. The histogram has a normalised text length of $NL = 10$. Once all calculations are over, all weights W are added to the normalised positions (NP) depending on $MINP$ and $MAXP$ of each text.

<table>
<tr><th>Text and calculation of NP</th><th>Histogram</th></tr>
<tr><td>

Text lenth $L = 7$, sign position $SP = 4$
Relative sign position from
$MINP = INT(\frac{(4-1)\times 10}{7} + 1) = 5$
to $MAXP = CEIL(\frac{4\times 10}{7}) = 6$
with weight $W = \frac{7}{10} = 0.7$
</td><td>


</td></tr>
<tr><td>

Text lenth $L = 3$, sign position $SP = 1$.
Relative sign position from
$MINP = INT(\frac{(1-1)\times 10}{3} + 1) = 1$
to $MAXP = CEIL(\frac{1\times 10}{3}) = 4$.
with weight $W = \frac{3}{10} = 0.3$
</td><td>


</td></tr>
<tr><td>
Adding weight W for each normalised position:
NP 1: 0.3 NP 6: 0.7
NP 2: 0.3 NP 7: 0.0
NP 3: 0.3 NP 8: 0.0
NP 4: 0.3 NP 9: 0.0
NP 5: 0.7 NP 10: 0.0
</td><td>


</td></tr>
</table>

Some signs are very frequent and therefore have high values in the histogram. For illustration purposes, the bar of the maximum value in the histogram is set for all histograms to the same height. This makes it easier to compare the histograms of rare and frequent signs.

3.3 Classification of Indus signs

Indus signs can be classified by their positional behaviour in complete texts. If a sign mostly occurs at the beginning of the text (at the right end of a sign sequence that should be read from right to left) it is labelled as an initial sign (INS), and if it preferably appears at the end of the text (at the left end of a sign sequence), it is a terminal sign (also called a terminal marker, TMK). Some examples will illustrate typical sign sequences with initial and terminal signs (Figure 3.2).

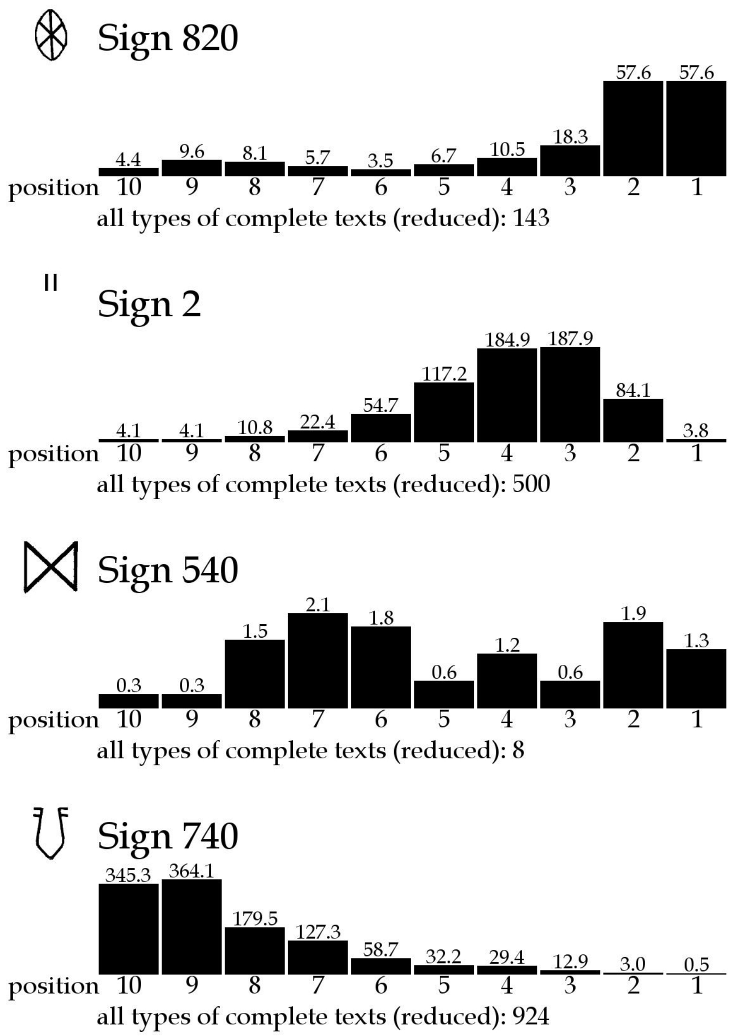


Figure 3.26: Positional histogram of signs in the text +740-540-002-820+ (Seal M-1088, see Figure 3.25). This is a typical example of a short patterned text with initial cluster 002-820 and terminal Sign 740.

Nahuatl word counts illustrating the behaviour of pure logographic writing systems (Figures 4.4b, 4.15, 4.21, 4.22). Mixed logographic-syllabic writing falls in between the two extremes (Figures 4.16, 4.17).

A high ratio of logograms with low frequencies is indicated by a sharp rise at the beginning of the curve in the diagrams of entropic redundancy. This occurs when logograms dominate in the sign list of a small corpus. In a logographic-syllabic writing system, however, as the size of the corpus increases, syllabic signs become more dominating than logograms because of their higher frequencies, and the curve becomes flatter (for a larger corpus) with greater degree of phonetization.

The relation between the curvature of sigmoid functions representing the entropic redundancy and the degree of phonetization of the writing system can be verified by comparing Ur III cuneiform and Korean writing. Both writing systems have roughly the same ratio of syllables in their sign lists and in their corpora (Table 4.1, Figure 4.7). As a result, their curves representing the entropic redundancy come very close together (Figure 4.21).

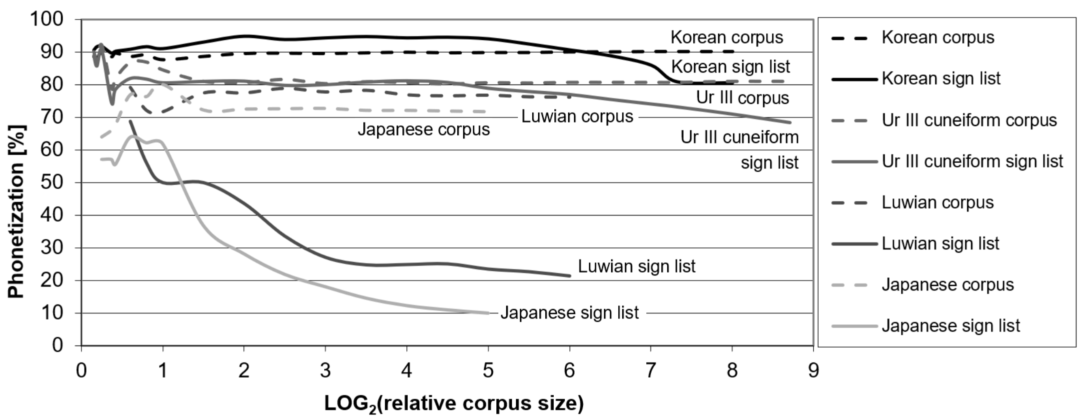


Figure 4.7: Change of phonetization in the sign list and in the corpus depending on the corpus size in logographic-syllabic writing: Korean, Ur III cuneiform, Luwian hieroglyphic and Japanese writings. The percentage of phonetic signs in the corpus is almost constant, but the number of syllables in the sign list decreases with increasing corpus size.

Let us have a look at the entropic redundancy of Indus writing in comparison with known writing systems. To avoid the so-called TAB-effect, identical duplicated texts are reduced to a single text. The calculations repeatedly performed 20 times for Indus writing give us a regression line what fits well ($R^2 = 0.992$) and has a slope of $drs = 0.598$ (Figure 4.8).

The change of the entropic redundancy of Indus writing is similar to that of Chinese, Japanese, and Luwian hieroglyphic writings (Figure 4.16). These writing systems are well known and have from about 10% to 35% phonetic signs

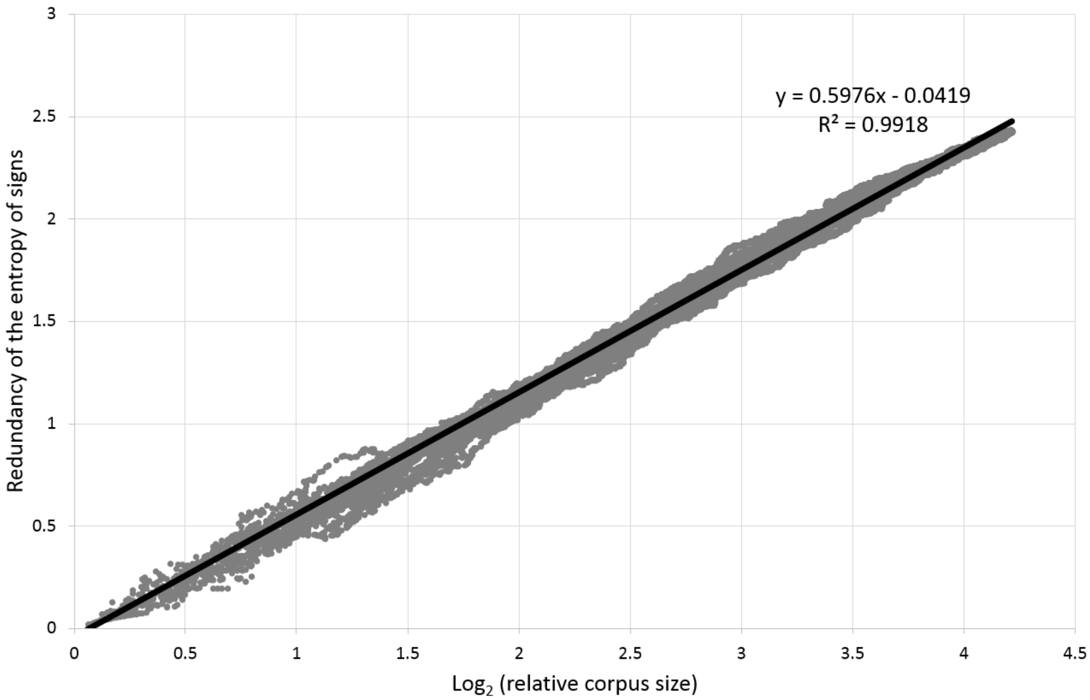


Figure 4.8: Entropic redundancy of Indus writing in relation to the relative text size of the corpus in a logarithmic-linear scale. The relative text size is the average frequency of each sign and constantly increases during the simulation of a growing corpus, which was performed 20 times, each time with a random order of adding new texts to the corpus.

in their sign lists (the value also depending on the corpus size, see Figure 4.7). Proceeding from this, we may expect a similar degree of phonetic writing in Indus inscriptions. At least for the Indus script we may exclude an alphabetic, pure syllabic and pure logographic writing system. This is in sharp contrast to what was proposed by many other scholars. Rao (1982) reduced the Indus sign list to 62 'basic' signs, since he assumed alphabetic writing during the Mature Harappan period. Ansumali Mukhopadhyay (2019), however, proposes a pure logographic writing system without syllables based on the sign list by Mahadevan (1977). Both extreme standpoints must now be abandoned.

4.5 Degree of phonetization

Most ancient writing systems are logographic-syllabic ones, i. e. they use a mixture of logograms and syllables. Every logographic-syllabic writing system has its own ratio of logograms to syllables, and spelling rules can approximate the linguistic value of a word. There is often some degree of flexibility of how a spe-

Chapter 5

Statistical data of signs

5.1 Description of signs

The chapter contains an elaborate description of each sign. It provides the grapheme of each sign and its variants and lists the total and positional frequency (solo, initial, medial, terminal) as well as the spatial frequency for Mohenjo-daro (MD), Harappa (Har.), Dholavira (Dhol.), Lothal, Kalibangan (Kali.), Chanhujo-daro (CD), and other sites. Each sign is assigned to a class according to the grapheme: compound (CMP), complex (CMX), marked (MKD), and simple signs (SIM). Set numbers of signs refer to the sign list in Chapter 2.3. The relationship between frequency and artefact type is also shown for square seals (SEAL:S), rectangular seals (SEAL:R), bas-relief tablets (TAB:B), copper tablets (TAB:C), incised tablets (TAB:I), graffiti on pottery (POT:T:g), and seal impressions (TAGs). For this reason identical inscriptions are not reduced. Below the frequencies of sites and artefact types the FHD and its sensitivity is listed for each sign with a frequency greater than one (Formula 1.1).

Class	Set	Total	Terminal	Medial	Initial	Solo
Sign class	Set number	Total and positional frequencies				
MD	Har.	Dhol.	Lothal	Kali.	CD	other
Frequency at each site						
Factor of Homogenous Distribution (FHD) and its estimated error at each site						
SEAL:S	SEAL:R	TAB:B	TAB:C	TAB:I	POT:T:g	TAGs
Frequency of each artefact type						
Factor of Homogenous Distribution (FHD) and its estimated error of each artefact type						

Every table is followed by the occurrences of the sign at each site including the frequencies. Site names are listed in alphabetical order. Sign occurrences in texts of unknown provenience are summarized and labeled ’unknown’.

The positional histogram is given for signs with a total frequency greater than 9. For the calculation of positional histograms (Chapter 3) only complete, one-sided texts are used. Identical inscriptions from the same site and on the same type of artifact are reduced to one text. This accounts for the so-called TAB-effect, when identical inscriptions (mostly tablets) are often found at the same site and in the same mold.

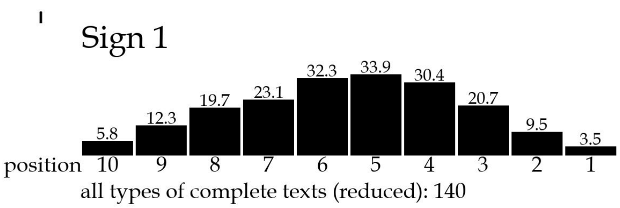
All ICIT ID's are listed at the end of the sign description. They refer to the identifier of the artefacts in the ICIT database and in the Volume *Corpus of Indus Inscriptions* where all texts are listed (Fuls, 2022). Please note that the ICIT ID is not the same as the text ID, since some artefacts are inscribed on more than one side.

Sign 1

Class	Set	Total	Terminal	Medial	Initial	Solo
SIM	01	228	24	189	16	4
MD	Har.	Dhol.	Lothal	Kali.	CD	other
115	50	17	22	2	0	22
1.16 ±0.01	0.55 ±0.01	2.16 ±0.13	2.55 ±0.12	0.31 ±0.16	- -	
SEAL:S	SEAL:R	TAB:B	TAB:C	TAB:I	POT:T:g	TAGs
89	54	11	15	7	15	17
0.95 ±0.01	2.59 ±0.05	0.26 ±0.02	1.23 ±0.08	0.27 ±0.04	1.31 ±0.09	1.87 ±0.11

Sign 1 occurs at Banawali (2), Chanhu-daro (3), Desalpur (1), Dholavira (17), Failaka (1), Harappa (50), Hissam-dheri (2), Kalibangan (2), Karzakan (1), Khirsara (2), Kot-Diji (1), Lothal (22), Mohenjo-daro (115), Nausharo (2), Saar (1), Susa (1), Unknown (3), Ur (2).

Positional Histogram



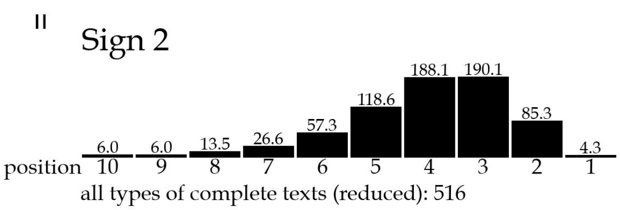
ICIT ID: 26, 79, 99, 111, 123, 144, 148, 164, 228, 229, 262, 278, 319, 339, 340, 371, 374, 383, 400, 434, 649, 718, 787, 814, 822, 835, 897, 898, 932, 988, 1046, 1067, 1068, 1069, 1086, 1087, 1099, 1120, 1164, 1184, 1186, 1195, 1196, 1199, 1227, 1266, 1415, 1422, 1465, 1504, 1545, 1552, 1557, 1563, 1621, 1710, 1711, 1872, 1873, 1882, 1944, 1968, 1972, 1985, 1992, 2009, 2034, 2042, 2047, 2059, 2068, 2069, 2070, 2071, 2072, 2084, 2112, 2113, 2115, 2117, 2125, 2135, 2152, 2157, 2161, 2212, 2213, 2218, 2221, 2227, 2233, 2236, 2280, 2313, 2321, 2332, 2362, 2368, 2370, 2377, 2414, 2419, 2443, 2488, 2493, 2531, 2549, 2551, 2567, 2573, 2583, 2593, 2611, 2638, 2683, 2697, 2730, 2757, 2775, 2784, 2789, 2791, 2819, 2829, 2849, 2853, 2854, 2868, 2874, 2876, 2882, 2890, 2896, 2898, 2912, 2994, 2995, 3000, 3001, 3002, 3003, 3004, 3043, 3044, 3073, 3084, 3101, 3109, 3181, 3196, 3219, 3252, 3276, 3301, 3325, 3369, 3374, 3385, 3395, 3405, 3427, 3431, 3457, 3461, 3474, 3487, 3515, 3524, 3548, 3557, 3562, 3569, 3571, 3574, 3583, 3589, 3591, 3594, 3597, 3605, 3613, 3615, 3619, 3634, 3639, 3644, 3645, 3680, 3712, 3735, 3738, 3739, 3741, 3882, 3889, 3898, 3908, 3993, 3995, 4059, 4161, 4162, 4164, 4177, 4194, 4197, 4201, 4208, 4212, 4229, 4247, 4265, 4274, 4280, 4284, 4349, 4895, 5223, 5224, 5225, 5228, 5229, 5371, 5384, 5388, 5441, 5509.

Sign 2

Class	Set	Total	Terminal	Medial	Initial	Solo
SIM	01	867	52	767	32	13
MD	Har.	Dhol.	Lothal	Kali.	CD	other
430	222	41	61	42	0	71
1.14 ±0.00	0.64 ±0.00	1.37 ±0.03	1.86 ±0.03	1.72 ±0.04	- -	
SEAL:S	SEAL:R	TAB:B	TAB:C	TAB:I	POT:T:g	TAGs
514	85	88	5	14	40	74
1.44 ±0.00	1.07 ±0.01	0.55 ±0.01	0.11 ±0.02	0.14 ±0.01	0.92 ±0.02	2.14 ±0.03

Sign 2 occurs at Allahdino (1), Bakkar Buthi (1), Bala-kot (2), Banawali (7), Baror (1), Bhirrana (1), Chanhu-daro (21), Desalpur (2), Dholavira (41), Gonur Depe (1), Harappa (222), Jhukar (1), Kalibangan (42), Kanmer (1), Kot-Diji (1), Lakhanjo-daro (1), Lohumjo-daro (1), Lothal (61), Mohenjo-daro (430), Nausharo (11), Rakhigarhi (3), Salut (1), Shikarpur (2), Surkotada (1), Susa (1), Tell Umma (1), Tepe Yahya (1), Unknown (7), Ur (1).

Positional Histogram



ICIT ID: 13, 18, 19, 20, 27, 30, 31, 36, 37, 38, 41, 52, 61, 62, 69, 71, 74, 76, 78, 81, 82, 84, 87, 89, 91, 92, 93, 94, 95, 98, 106, 111, 123, 125, 130, 132, 136, 143, 145, 146, 216, 223, 246, 256, 259, 280, 284, 293, 296, 303, 315, 323, 341, 342, 349, 350, 356, 370, 388, 401, 402, 409, 411, 412, 448, 466, 476, 488, 507, 513, 518, 529, 536, 537, 543, 594, 599, 604, 617, 621, 623, 626, 634, 665, 682, 729, 731, 744, 771, 787, 789, 805, 835, 937, 943, 951, 955, 960, 967, 974, 975, 989, 996, 1004, 1007, 1020, 1024, 1030, 1047, 1060, 1064, 1069, 1073, 1074, 1076, 1078, 1081, 1082, 1091, 1100, 1101, 1103, 1104, 1110, 1114, 1119, 1127, 1132, 1135, 1139, 1141, 1148, 1155, 1164, 1170, 1171, 1172, 1175, 1176, 1178, 1181, 1182, 1190, 1195, 1204, 1211, 1258, 1269, 1277, 1278, 1279, 1280, 1281, 1282, 1283, 1284, 1285, 1286, 1287, 1288, 1289, 1290, 1293, 1296, 1297, 1298, 1299, 1307, 1314, 1315, 1394, 1398, 1399, 1401, 1402, 1410, 1412, 1413, 1414, 1420, 1426, 1434, 1435, 1436, 1445, 1446, 1447, 1449, 1450, 1452, 1456, 1459, 1463, 1466, 1468, 1472, 1474, 1479, 1484, 1485, 1491, 1515, 1517, 1526, 1537, 1540, 1541, 1543, 1549, 1551, 1552, 1554, 1555, 1564, 1566, 1568, 1573, 1578, 1594, 1602, 1606, 1607, 1611, 1622, 1665, 1695, 1696, 1697, 1700, 1717, 1719, 1724, 1742, 1743, 1744, 1745, 1746, 1747, 1748, 1749, 1750, 1751, 1752, 1753, 1769, 1786, 1791, 1792, 1875, 1878, 1879, 1885, 1887, 1893, 1897, 1898, 1899, 1901, 1902, 1906, 1908, 1910, 1917, 1927, 1928,

5.2 Relation of sign frequencies to artefact types

The frequency of each sign selected according to the artefact type is listed in the table below. To reduce the size of the table, similar types of artefacts are summarized as follows:

Seals: SEAL, SEAL:C, SEALCY, SEAL:L, SEAL:O, SEAL:R, SEAL:S.

Tablets: TAB:B, TAB:C, TAB:I.

Pottery: POT:T:g, POT:T:p, POT:T:s.

Tags: TAG, TAG:B, TAG:C, TAG:L, TAG:O, TAG:P, TAG:R, TAG:S, TAG:W.

other: BEAD, BNGL, IMPL, MDLN, MISC, ROD, Oth.

For a detailed list of sign frequencies in relation to the artefact types SEAL:R, SEAL:S, TAB:B, TAB:C, TAB:I, POT:T:g, and TAGs check the tables for each sign in Chapter 5.1.

For a frequency greater than three the FHD is given in brackets (Chapter 1.7, Formula 1.1). The FHD indicates over- and under-representation of signs in relation to what can be expected for a perfect homogeneous frequency distribution. A FHD greater than 1.8 is highlighted and indicates a high factor, which means that the sign is overrepresented. A FHD smaller than about 0.5 indicates the underrepresentation of a sign for that type of artefact.

Sign	Freq.	Seals	Tablets	Pottery	Tags	other
all	17991	9400	6365	1062	719	445
1	228	156 (1.3)	33 (0.4)	15 (1.1)	**17 (1.9)**	7 (1.2)
2	867	620 (1.4)	107 (0.3)	47 (0.9)	**74 (2.1)**	19 (0.9)
3	260	130 (1.0)	91 (1.0)	27 (1.8)	6 (0.6)	6 (0.9)
4	99	47 (0.9)	30 (0.9)	**14 (2.4)**	3	5 (2.0)
5	49	33 (1.3)	7 (0.4)	5 (1.7)		4 (3.3)
6	7	4 (1.1)		1		2
7	6	4 (1.3)		1		1
12	3	2		1		
13	33	18 (1.0)	9 (0.8)	**4 (2.0)**		2
14	6	3	2	1		
15	8	5 (1.2)	2		1	
16	48	34 (1.4)	8 (0.5)	3		3
17	79	48 (1.2)	23 (0.8)	5 (1.1)	2	1
18	7	4 (1.1)	2			1
19	6	5 (1.6)	1			
20	2		1	1		
25	3		3			
26	1	1				
27	6		**6 (2.8)**			
28	5	2	3			
29	4	2		2		
31	251	146 (1.1)	42 (0.5)	**41 (2.8)**	7 (0.7)	15 (2.4)
32	586	223 (0.7)	251 (1.2)	**71 (2.0)**	30 (1.3)	11 (0.8)
33	523	151 (0.6)	287 (1.6)	**68 (2.2)**	9 (0.4)	8 (0.6)
34	185	6 (0.1)	**145 (2.2)**	**30 (2.8)**	1	3
35	35	12 (0.7)	7 (0.6)	**13 (6.3)**	1	2
36	7	2	2	3		
37	2			1		1
39	2			2		
41	2	1				1
42	2	2				
43	1	1				
44	1	1				
45	2	1	1			
46	2	1		1		
47	1	1				
48	18	4 (0.4)	4 (0.6)		**10 (13.9)**	
49	6	3	1	1	1	
50	1			1		

5.3 Relation of sign frequencies to regions

The frequency of each sign selected according to the region is listed in the table below. Additionally, the Factor of Homogeneous Distribution (FHD) is given in brackets (Formula 1.1). Values are highlighted in case of a FHD greater than 1.8 to indicate an over-proportional frequency. Each region includes the following sites (for each site the total number of legible signs of all texts is given in brackets):

Eastern Headwaters: Alamgirpur (6), Banawali (62), Bhirrana (11), Chandigarh (8), Farmana (16), Hulas (3), Rakhigarhi (47), Rupar (15).

Gujarat/Kutch: Desalpur (17), Dholavira (622), Gola Dhoro (Bagasra) (28), Juna Khatiya (2), Kanmer (18), Khirsara (27), Lothal (681), Pabumath (5), Rodji (7), Shikarpur (15), Surkotada (15).

Iranian Plateau: Luristan (4), Miri Qalat (4), Nindowari-damb (12), Ornach area (2), Sibri (1), Susa (13), Tepe Yahya (2).

Inter-Riverine: Baror (3), Derawar Ther (1), Ganweriwala (10), Guddal A (1), Harappa (7165), Kalibangan (508), Karanpura (5), Rajanpur (7), Rappwala Ther (1), Tarkhanewala-dera (6), Tigrana (2), Wattoowala (2).

Lower Indus: Allahdino (40), Amri (7), Bakkar Buthi (6), Bala-kot (25), Chanhujo-daro (326), Gharo Bhiro (3), Jhukar (5), Kot-Diji (21), Lakhanjo-daro (34), Lohumjo-daro (10), Mohenjo-daro (7806), Naru-Waro-dharo (4), Nausharo (129), Nuhato (7), Pirak (5).

Mesopotamia: Girsu (4), Kish (9), Nippur (1), Tell Umma (5), Tello (4), Ur (18).

Persian Gulf: Failaka (7), Hajar (2), Janabiyah (5), Kalba (1), Karzakan (10), Qala'at al-Bahrain (5), Ra's al-Junayz (8), Saar (6), Salut (6).

South India: Daimabad (2).

Upper Indus: Gumla (2), Hissam-dheri (4), Rehman-deri (1).

other: Altin Tepe (2), Gonur Depe (8), Shortughai (2), and unknown provenience (93).

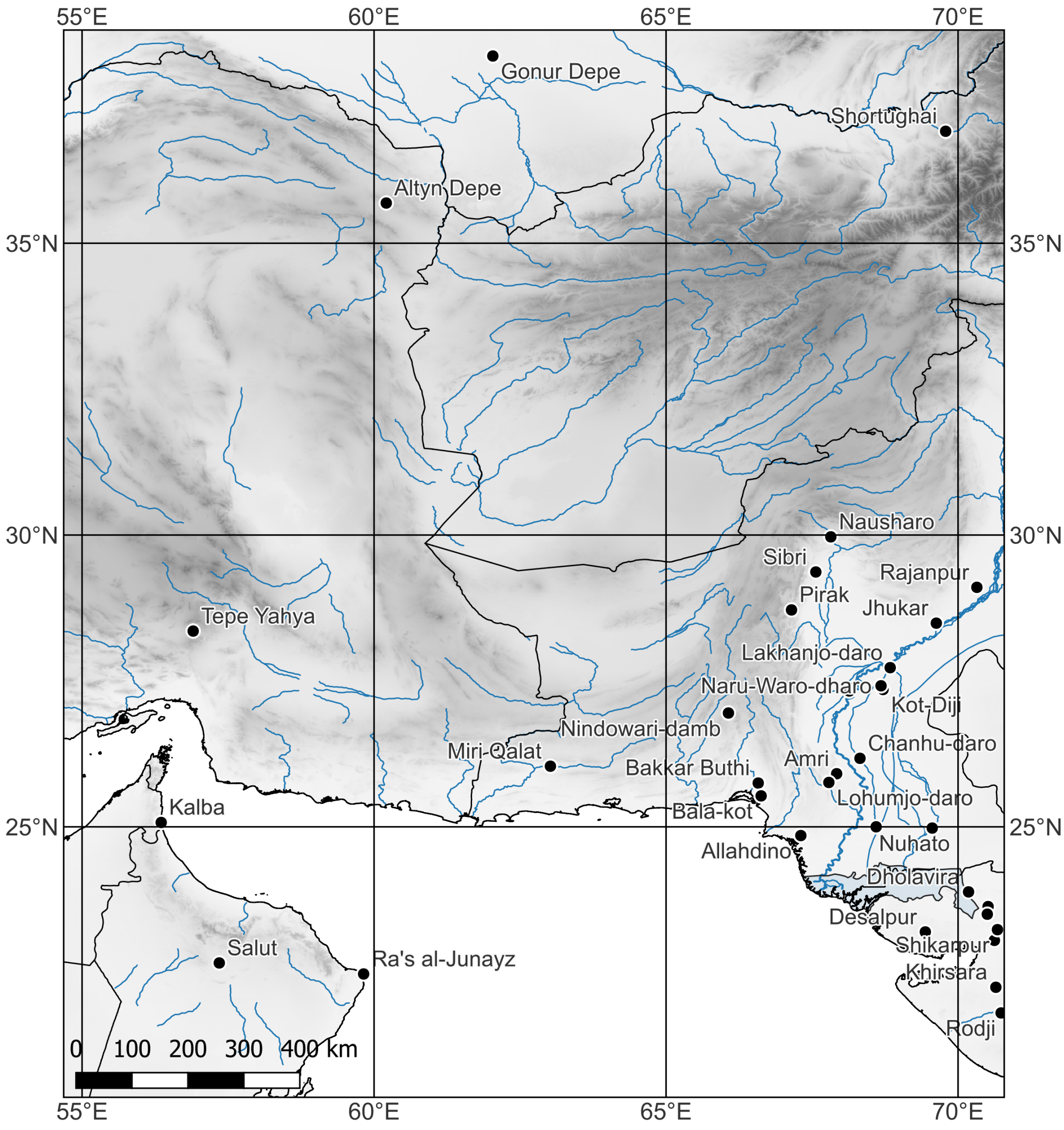
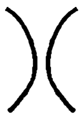
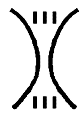
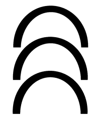
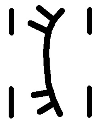
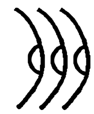
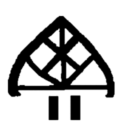
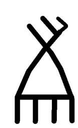
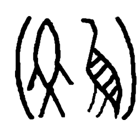


Figure 5.4: Map of Iranian Plateau.

Sign	Freq.	Lower Indus	Upper Indus	Gujarat / Kutch	Inter-Riverine	other
747	7	**6 (2.0)**			1	
748	3				2	1
749	6	**6 (2.3)**				
750	2	2				
751	1	1				
752	53	39 (1.7)			14 (0.7)	
753	2	2				
759	2	1			1	
760	134	71 (1.2)		14 (1.7)	46 (0.9)	3
761	1	1				
762	1				1	
763	1	1				
764	1	1				
765	1	1				
766	1	1				
767	2	1			1	
768	1	1				
770	2	1			1	
772	10	**8 (1.8)**			2	
773	9	**8 (2.0)**			1	
775	4				**4 (2.6)**	
776	11	2		2	7 (1.7)	
777	7	**6 (2.0)**			1	
778	1			1		
780	2	2				
781	2	2				
782	3	3				
783	2				2	
784	1				1	
785	2				2	
786	3	3				
790	56	29 (1.2)		5 (1.4)	22 (1.0)	
791	6	3			3	
792	13	6 (1.1)			7 (1.4)	
793	1	1				
794	13	8 (1.4)		2	3	
795	2	1			1	
796	1	1				
797	4	1			2	1

Index

 **906** (16)	 **907** (1)	 **908** (3)	 **909** (3)	 **910** (4)
 **911** (1)	 **912** (4)	 **913** (2)	 **914** (1)	 **918** (1)
 **919** (2)	 **920** (150)	 **921** (16)	 **922** (4)	 **923** (45)
 **924** (10)	 **925** (1)	 **926** (1)	 **927** (11)	 **928** (2)
 **930** (8)	 **931** (2)	 **932** (1)	 **933** (2)	 **940** (2)

View publication stats