

## A FREQUENCY ANALYSIS OF THE INDUS SCRIPT

Subhash C. Kak

ADDRESS: Department of Electrical and Computer Engineering, Louisiana State University,  
Baton Rouge LA 70803 USA.

ABSTRACT: A statistical analysis of the Indus signs is presented. It is shown that the frequencies of the most common signs support the theory of the derivation of Brahmi from Indus. Furthermore, evidence is presented that suggests that the language of the Indus inscriptions was Indo-Aryan.

KEYWORDS: Decipherment of the Indus script, Indo-Aryan languages, Brahmi script, ancient writing systems.

### INTRODUCTION

In a recent article [1] it was argued why the connections between the Brahmi (earliest available record c. 300 B.C. [See Figure 1]) and the Indus (3000 B.C.-1500 B.C.) scripts should be investigated further in order to make progress in the decipherment of the latter. In that article the identification of many Indus signs was made on general conditions of shape and the assumption that the text for the words *sapta sindhu* had been correctly identified. The question of the frequency distribution of the Indus signs is now being taken up. It is shown that this validates the earlier study and makes it possible to advance the decipherment further.

A careful analysis of the structural characteristics of the Indus texts was published by G. R. Hunter in 1934.[2] Another noteworthy analysis is that of Mitchiner [3] who also provides an excellent critique of the other attempts at decipherment.

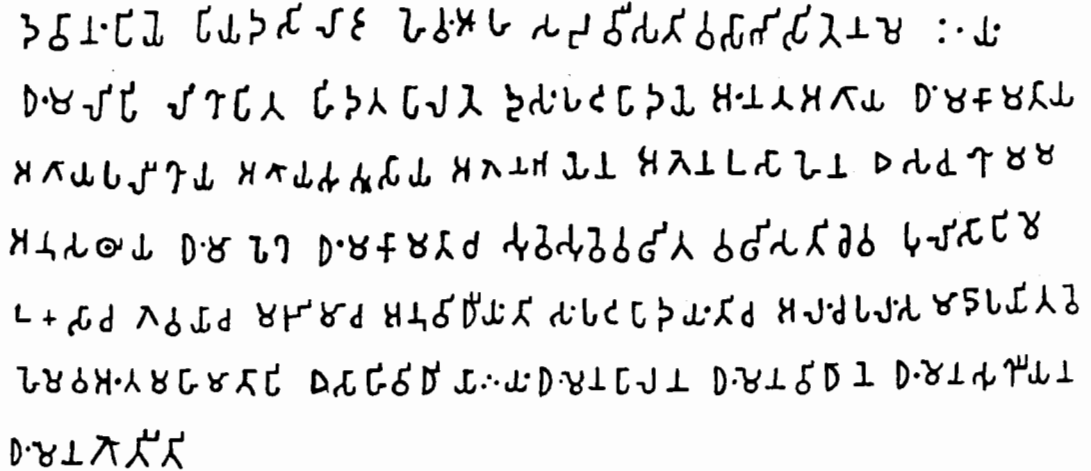
The major conclusion of Hunter was that Brahmi is derived from Indus. He demonstrated that the script is a syllabary of open and closed syllables. Since many of the signs of the script appear singly, he further concluded that the language of the script is essentially monosyllabic. This last conclusion may not be correct since single signs could represent abbreviations for longer words or be symbols with religious or business implications. That abbreviations and other palaeographical formulae are likely to have been used in Indus records is indicated by such use in later Indian inscriptions.[4] To arrive at a list of possible equivalents between Indus and Brahmi signs, Hunter also considered connections with Sumerian, Egyptian, Proto-Elamite, South Semitic, Phoenician, and Cypriot.

Hunter suggested that the similarities implied that Sabaeen, Phoenician and Cypriot were derived from the Indus script. Table 1 summarizes the morphological relationship between Indus and Brahmi given by Hunter. Note that this has several agreements with the smaller set

of equivalences given in [1]. The most notable differences are shown in brackets in the table itself.

Mitchiner's main contribution is an analysis of the case-endings of the Indus inscriptions. He concluded that these endings ruled out Elamite or Dravidian as the languages of the inscriptions. Furthermore, using Indus-Brahmi comparisons, he argued that the case-endings were in consonance with early Indo-Aryan. He also noted that the languages of Harappa and Mohenjo Daro had distinct features that could well indicate ancestry of Sanskrit and Prakrit respectively.

The analysis of [5], as well as the changing understanding of the Harappan civilization [6], validates the general direction of the researches of Hunter and Mitchiner. The objective of this paper is to determine if a frequency analysis of the signs of the Indus and the Brahmi scripts can yield further insights into the nature of the Indus script and its language. The analysis should be of interest to the general cryptologist as well, since the underlying problem is to see how two alphabets, one of which is used for unread texts, can be shown to be related.



Asoka's First Pillar Edict, Lauriyā Nandangarh, c. 242 B.C.  
Transcript

- (Line 1) De-vā-ṇaṃ-pi-ye Pi-ya-da-si lā-ja he-vaṃ ā-ha sa-ḍu-vi-sa-ti-va-sā-bhi-si-te-na me i-yaṃ
- (2) dhaṃ-ma-li-pi li-khā-pi-ta. Hi-da-ta-pā-la-te du-saṃ-pa-ṭi-pā-da-ye aṃ-na-ta a-gā-ya dhaṃ-ma-kā-ma-tā-ya
- (3) a-gā-ya pa-li-khā-ya a-gā-ya su-sū-sā-ya a-ge-na bha-ye-na a-ge-na u-sā-he-na. E-sa cu kho ma-ma
- (4) a-nu-sa-thi-ya dhaṃ-mā-pe-kha dhaṃ-ma-kā-ma-tā ca su-ve su-ve va-dhi-ta va-dhi-sa-ti ce-va. Pu-li-sā pi me
- (5) u-ka-sā ca ge-va-yā ca ma-jhi-mā ca a-nu-vi-dhi-yaṃ-ti saṃ-pa-ṭi-pā-da-yaṃ-ti ca a-laṃ ca-pa-laṃ sa-mā-dā-pa-yi-ta-ve,
- (6) he-me-va aṃ-ta-ma-hā-mā-tā pi. E-sā hi vi-dhi yā i-yaṃ dhaṃ-me-na pā-la-na dhaṃ-me-na vi-dhā-ne dhaṃ-me-na su-khī-ya-na
- (7) dhaṃ-me-na go-tī ti.

Translation.

Thus speaks the King, Dear to the Gods, Priyadarśi. When I had been consecrated twenty-six years I ordered this inscription of the Law (Dharma) to be engraved. Both this world and the other are hard to reach, except by great Love of the Law, great self-examination, great obedience (to the Law), great respect (for the Law), great energy. But through my leadership respect for

the Law and love of the Law have grown and will grow from day to day. Moreover my officers, of high, low and medium grades, follow it and apply it, sufficiently to make the waverer accept it; the officers on the frontiers do likewise. For this is (my) rule: government by the Law, administration according to the Law, gratification (of my subjects) by the Law, protection by the Law.

Figure 1. Brahmi writing (Source: Basham [10]).

s in the table

s inscriptions.  
ages of the in-  
e case-endings  
Harappa and  
it and Prakrit

ivilization [6],  
jective of this  
Brahmi scripts  
The analysis  
blem is to see  
elated.

• ↓  
8+8LJ  
L 7 8 8  
P L C 8  
P S L L 8  
L 4 7 6 1

e,  
r-khl-ya-na  
  
ll grow from day  
medium grades,  
erer accept it; the  
(my) rule: gov-  
the Law, gratifi-  
the Law.

Indus	Brahmi	Devanagari	Indus	Brahmi	Devanagari
H	𑀓, H	a		𑀔	ā
	𑀓	ā	𑀕, 𑀖, 𑀗, 𑀘	𑀕	ṭa
𑀙	𑀚	i	○	○	ṭha
𑀛	𑀜	ī	𑀝, 𑀞, 𑀟	𑀟	ḍa
𑀠, 𑀡, 𑀢	L	u	𑀣	𑀤	ḍha
𑀥	E	ū	𑀦	I	ṇa
		r	𑀧, 𑀨, 𑀩, 𑀪	𑀫	ta
		r̄	○	○	tha
		!	𑀬	𑀭, C	da
		ī	D	D	cha
A, Δ	▷	e	𑀮 [L]	𑀯	na
[↑]	→	ai	U, J	𑀰	pa
𑀲, 𑀳	𑀴	o	𑀵, 𑀶, 𑀷	𑀸, 𑀹	pha
	𑀺	au	𑀻	𑀼	ba
𑀽, 𑀾	.	m̄	𑀿	𑀿, 𑀿	bha
		h̄	𑀿	𑀿	ma
𑀿, 𑀿	+	ka	𑀿, 𑀿 [Y]	𑀿	ya
𑀿, 𑀿	𑀿, 𑀿	kha	𑀿, 𑀿	𑀿	ra
𑀿, 𑀿, 𑀿	𑀿, 𑀿	qa	𑀿, 𑀿, 𑀿	𑀿, 𑀿	la
𑀿, 𑀿	𑀿	gha	𑀿	𑀿, b	va
	𑀿	ṇa	𑀿 [↑]	↑	ṣa
𑀿, 𑀿, 𑀿	𑀿	ca	𑀿	𑀿	ṣa
𑀿, 𑀿	𑀿	cha	𑀿 [Y]	𑀿, 𑀿	sa
E, E, E	E	ja	𑀿	𑀿	ha
Y, Y, Y	H	jha			

Table 1. Hunter's 'identification' of the Indus signs. The noteworthy differences in [1] have been shown in brackets.

4002	100101	𑀮𑀭𑀢𑀺𑀢𑀺𑀢𑀺𑀢𑀺
4003	100101	𑀮𑀭𑀢𑀺𑀢𑀺𑀢𑀺𑀢𑀺𑀢𑀺
4004	100101	𑀮𑀭𑀢𑀺𑀢𑀺𑀢𑀺𑀢𑀺𑀢𑀺
4005	100101	𑀮𑀭𑀢𑀺𑀢𑀺𑀢𑀺𑀢𑀺𑀢𑀺
4006	100101	𑀮𑀭𑀢𑀺𑀢𑀺𑀢𑀺𑀢𑀺𑀢𑀺
2901	510003	𑀮𑀭𑀢𑀺𑀢𑀺𑀢𑀺𑀢𑀺𑀢𑀺
	20001	𑀮𑀭𑀢𑀺𑀢𑀺𑀢𑀺𑀢𑀺𑀢𑀺
2847	210001	𑀮𑀭𑀢𑀺𑀢𑀺𑀢𑀺𑀢𑀺𑀢𑀺
	20001	𑀮𑀭𑀢𑀺𑀢𑀺𑀢𑀺𑀢𑀺𑀢𑀺
	30001	𑀮𑀭𑀢𑀺𑀢𑀺𑀢𑀺𑀢𑀺𑀢𑀺
4012	100101	𑀮𑀭𑀢𑀺𑀢𑀺𑀢𑀺𑀢𑀺𑀢𑀺
4013	100101	𑀮𑀭𑀢𑀺𑀢𑀺𑀢𑀺𑀢𑀺𑀢𑀺
4014	100101	𑀮𑀭𑀢𑀺𑀢𑀺𑀢𑀺𑀢𑀺𑀢𑀺
4015	100101	𑀮𑀭𑀢𑀺𑀢𑀺𑀢𑀺𑀢𑀺𑀢𑀺
4016	100101	𑀮𑀭𑀢𑀺𑀢𑀺𑀢𑀺𑀢𑀺𑀢𑀺
4017	100101	𑀮𑀭𑀢𑀺𑀢𑀺𑀢𑀺𑀢𑀺𑀢𑀺
4018	100101	𑀮𑀭𑀢𑀺𑀢𑀺𑀢𑀺𑀢𑀺𑀢𑀺
4019	100101	𑀮𑀭𑀢𑀺𑀢𑀺𑀢𑀺𑀢𑀺𑀢𑀺
4020	100101	𑀮𑀭𑀢𑀺𑀢𑀺𑀢𑀺𑀢𑀺𑀢𑀺
4021	100101	𑀮𑀭𑀢𑀺𑀢𑀺𑀢𑀺𑀢𑀺𑀢𑀺
4022	100101	𑀮𑀭𑀢𑀺𑀢𑀺𑀢𑀺𑀢𑀺𑀢𑀺
4023	100101	𑀮𑀭𑀢𑀺𑀢𑀺𑀢𑀺𑀢𑀺𑀢𑀺
4024	100101	𑀮𑀭𑀢𑀺𑀢𑀺𑀢𑀺𑀢𑀺𑀢𑀺
4025	100101	𑀮𑀭𑀢𑀺𑀢𑀺𑀢𑀺𑀢𑀺𑀢𑀺
4026	100101	𑀮𑀭𑀢𑀺𑀢𑀺𑀢𑀺𑀢𑀺𑀢𑀺
4027	100101	𑀮𑀭𑀢𑀺𑀢𑀺𑀢𑀺𑀢𑀺𑀢𑀺
5256	210001	𑀮𑀭𑀢𑀺𑀢𑀺𑀢𑀺𑀢𑀺𑀢𑀺
	23690	

Figure 2. A page of Indus writing. The numbers in the left column identify the texts. Note Number 2847 is 26 characters long.

## STATISTICAL CONSIDERATIONS

The concordances of Parpola [7] or Mahadevan [8] may be consulted to give frequencies of various Indus signs. It should be noted, however, that since the Indus inscriptions are not representative of the language, being mostly proper names, these frequencies cannot be directly related to the letter frequencies of the alphabet of the underlying language. As to the data, only around 3000 inscriptions (many of which are multiple copies) have yet been found. Most of the inscriptions are very brief, the average length being 5 characters (Figure 2). There are no bilinguals and Brahmi, the other ancient script from the region, is not attested from epochs before 300 B.C.

There is evidence that Brahmi had a long history in India before the Mauryan times. The vowel/consonant classification system of Brahmi is a part of *Aṣṭādhyāyī*, the great grammar of Pāṇini (fifth or sixth century B.C.) who himself was at the end of a long tradition of grammarians. Written characters (letters) are mentioned in *Chāndogya* and *Taittirīya Upaniṣad*, and *Aitareya Āraṇyaka* refers to the distinctions between the various consonant classes. The fact that *Chāndogya Upaniṣad* lists the basic alphabet and at the same time analyzes words by letter counts of syllables suggests that writing at the time must have been syllabic as in Brahmi. This indicates that the antecedents of Aśokan Brahmi can be seen *at least* several centuries before Pāṇini. The interval between Indus writing and early Brahmi writing is thus likely to be much narrower than the inscriptional evidence implies.

A hypothesis of Indus-Brahmi relationship suggests that the frequencies of the different sounds of the Sanskrit alphabet may be a guide in identifying the Indus alphabet. Table 2 is from [9] and was obtained from 10,000 sounds of continuous text, in ten different passages, of 1,000 sounds each, selected from different epochs of the literature. Note that since Brahmi is really a syllabary where each consonant subsumes *a*, and where the other vowels shown in the first column could be represented by ligatures on the consonants, the frequencies of these vowels as separate letters in written text will be much lower. Since our objective, in this study, is to determine what letters the most likely Indus signs could represent, and then validate this on morphological grounds, we only indicate the most likely 10 consonants in decreasing frequency:

*t, r, v, n, m, y, s, d, p, k.*

The frequency of each one of these consonants is greater than 1.99 percent. The next most frequent sounds are ś (palatal sh) and ṣ (lingual sh) with frequencies of 1.57 and 1.45 respectively. It is conceivable that like Prakrit, early Indo-Aryan could have often substituted *s* for ś and ṣ. If this were to have been the case the total frequency of *s* could be close to that of *t*.

Assuming that the system of ligatures used for the Indus alphabet had not reached the perfection it shows in Brahmi, the vowels that may need to be taken into consideration are *ā, i, u, e*. With the possible exception of *ā*, the frequencies of these vowels in individual appearance is likely to be lower than that of the ten consonants listed before, however.

Next we list the ten most common Indus signs (Table 3) from the concordances [7] or [8]. The counts include signs when they appear as components of 'compound' signs. The specific frequencies have not been indicated since they can be misleading because the concordances list multiple inscriptions and because the texts can hardly be representative of the Indus language.

Sound Freq.	Sound Freq.	Sound Freq.
a 19.78	k 1.99	th 0.58
ā 8.19	kh 0.13	d 2.85
i 4.85	g 0.82	dh 0.83
ī 1.19	gh 0.15	n 4.81
u 2.61	ñ 0.22	p 2.46
ū 0.73	c 1.26	ph 0.03
r 0.74	ch 0.17	b 0.46
ṛ 0.01	j 0.94	bh 1.27
ṛ̣ 0.01	jh 0.01	m 4.34
ṛ̣̣ -	ṇ̃ 0.35	y 4.25
e 2.84	ṭ 0.26	r 5.05
āi 0.51	tḥ 0.06	l 0.69
o 1.88	ḍ 0.21	v 4.99
āu 0.18	dḥ 0.03	ś 1.57
ṁ 0.63	ṇ 1.03	ṣ 1.45
ḥ 1.31	ṭ 6.65	ṣ̣ 3.56
		h 1.07

Table 2. Frequencies of various sounds of the Sanskrit alphabet (Source: Whitney [9]).

### ANALYSIS OF DATA

Table 3 presents the ten most frequent Indus symbols so that even though the Indus texts cannot be expected to yield reliable statistics, one would be reasonably sure to have included many of the ten most likely symbols of the language. An examination of this Table shows a clear morphological connection between the Indus signs and the ten most common consonants of Brahmi. Further note that Lal has observed Chalcolithic-Megalithic intermediary forms of  $\text{𑀘}$  (the most common Indus sign) such as  $\text{𑀙}$  and, therefore, it is comparable to the Brahmi sign for *sa*. [3, p.73] The other similar sign is the 'fish' sign of Indus and *ma* of Brahmi, and Indus has signs that look identical to the *ra* and *va* of Brahmi. Another case of similarity are the Brahmi *ta* and the Indus 'man' signs.

If it is accepted that four signs have survived into Brahmi from Indus, the question of what is the probability that this signifies a real connection arises. Similarity between signs could be due to one of the three following reasons:

Sanskrit consonants (decreasing freq)	t, r, v, n, m, y, s, d, p, k
Brahmi	𑀓, 𑀔, 𑀕, 𑀖, 𑀗, 𑀘, 𑀙, 𑀚, 𑀛, 𑀜
Indus (Hunter [2])	𑀓, 𑀔, 𑀕, 𑀖, 𑀗, 𑀘, 𑀙, 𑀚, 𑀛, 𑀜
Indus (Kak [1])	𑀓, 𑀔, 𑀕, 𑀖, 𑀗, 𑀘, 𑀙, 𑀚, 𑀛, 𑀜
Indus (decreasing freq) [8]	𑀓, 𑀔, 𑀕, 𑀖, 𑀗, 𑀘, 𑀙, 𑀚, 𑀛, 𑀜

Table 3. The ten most common consonants in decreasing frequency.

1. The Brahmi letters have no connection whatsoever with the Indus signs, and any perceived similarities are purely random arising owing to the simplicity of the shapes.
2. The Brahmi letters were devised by those who had seen the Indus inscriptions but did not know how to read them. This would explain why the letters of the two scripts would look greatly similar; but the sound values of the letters would have no connection.
3. The Brahmi letters are derived from the Indus signs.

We investigate each of the cases now.

Cases 1 and 2

The problem here is to determine how two scripts can be found to be related. Note that we have found that special subsets - the most frequent letters - of the two alphabets are morphologically connected. Assuming that four letters are unchanged (for which different evidence will be presented in the next sections), the probability of this happening by chance for two scripts of size 33 (Brahmi consonants) and about 300 (Indus signs with obvious variants and transformed signs not counted) is

$$\frac{29! \times 296!}{33! \times 300!} \approx 0.1 \times 10^{-12}$$

This probability is so small that the correspondence of Table 3 cannot be taken to be accidental.

Indus	Brahmi	Devanagari	Indus	Brahmi	Devanagari
𑀓	𑀠, 𑀡	a		𑀣	ā
	𑀢	ā		𑀤	ṭa
	𑀣	i		𑀥	ṭha
	𑀤	ī		𑀦	ḍa
𑀧	𑀧	u		𑀨	ḍha
	𑀩	ū		𑀪	ṇa
		r	𑀫	𑀬	ṭa
		ṛ		𑀭	ṭha
		ḷ	𑀮	𑀯, 𑀰	ḍa
		ṛ		𑀱	ḍha
	𑀲	e	𑀳, 𑀴	𑀵	ṇa
𑀶	𑀷	ai	𑀸	𑀹	pa
𑀺	𑀻	o		𑀼, 𑀽	pha
	𑀾	au	𑀿	𑀽	ba
	·	m		𑀿, 𑀽	bha
		ḥ	𑀽	𑀽	ma
	+	ka	𑀽	𑀽	ya
	𑀽, 𑀽	kha	𑀽, 𑀽	𑀽	ra
𑀽	𑀽, 𑀽	ga		𑀽, 𑀽	ḷa
	𑀽	gha	𑀽	𑀽, 𑀽	va
	𑀽	ṇa	𑀽	𑀽	ṣa
𑀽	𑀽	ca		𑀽	ṣa
	𑀽	cha	𑀽	𑀽, 𑀽	sa
𑀽, 𑀽	𑀽	ja		𑀽	ha
	𑀽	jha			

Table 4. Tentative identification of the Indus characters.

## Case 3

The conclusion that there is no alternative to Case 3 is strengthened by the connection between the Indus signs of Hunter, obtained using structural analysis and comparisons with other



Devanagari
na
ta
tha
da
dha
na
ta
tha
da
dha
na
pa
pha
ba
bha
ma
ya
ra
la
va
śa
ṣa
sa
ha

scripts, and Indus for the ten most frequent signs. There is an exact correspondence for four signs. The connection between the Indus identification of [1] and the ten most frequent signs of the script is even more striking, with five correspondences.

Consider now the letter frequencies for short Brahmi records in Sanskrit. If these records were to represent ownership or epithets of kings, or sacred formulae, they would generally terminate with the genitive case and sometimes with the dative case. This was true of the Brahmi seals. The letter frequencies would now change somewhat with *s*, *m*, *ai* becoming more probable since *s* and *m* are associated with the genitive singular and plural respectively; and *ai* is associated with the dative case-endings. This implies that *s* and *m* may be the most likely consonants in such brief Brahmi records. It is noteworthy to see that these are precisely the two most likely Indus signs. Not only does this indicate an Indus-Brahmi continuity, but also a connection in the underlying languages.

### FURTHER SIGN IDENTIFICATION

Given that the frequency of all the *s* signs of Brahmi is next to that of *t*, the six most likely Brahmi consonants then become *t*, *s*, *r*, *v*, *n*, *m*, and we observe that four of these correspond to the signs amongst the six most frequent Indus signs. This increases our confidence in the identification of *s*, *v*, *m*, and suggests that the 'man' sign could be *t*. Also the sign for *y* may be the "arrow with the triangle head" that had been identified as *ai* in [1]. This is because Indo-Aryan languages often substitute *ai* for *y* and also because the 'triangle/arrow' sign is amongst the most frequent in the Indus script. Of course, all sign identifications will remain tentative till the script is fully deciphered. Nonetheless, as the analysis proceeds, one needs to weigh evidence from different lines of attack to constantly refine ones assumptions. At present we will keep both *y* and *ai* as possible readings of the 'triangle/arrow' sign. Also the Brahmi *p* and the Indus U look related. Perhaps the Indus *p* is U and U represents a common ligature such as *pu* or *pra*.

Table 4 presents Indus signs whose relationship with the corresponding Brahmi signs is supported by Sumerian parallels or by frequency considerations. One may thus place some confidence in this identification at this stage of decipherment.

### CASE-ENDINGS

The case-endings of the Indus inscriptions have been analyzed by Hunter, Parpola, and Mitchiner. Table 5 presents a list of these case endings, that we have grouped under two categories. Figure 3 shows how these case-endings are defined.

The terminal sign 1 of Table 5 was shown by Hunter to represent an open syllable, and that it is an affix or suffix rather than a determinative. Most scholars take it to be the ending of the genitive case. Mitchiner argues that it should represent the genitive singular case-ending. The optional form U U was shown by Hunter to represent the 'spelling-out' of the sign U.

The second main terminal sign ^ performs a task different to that of U, as can be seen by the mutually exclusive nature of sign groups that precede them. It was suggested by Parpola that this sign could represent a form of dative case-suffix and this suggestion has been widely accepted.

ction between  
ns with other



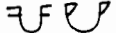



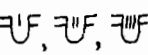

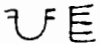



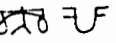
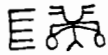

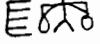

Genitive case-endings		Dative case-endings	
1.	 -sa	15.	 -ya(-āi)
2.	 -s.sa	16.	 -ya.ja(-āi.ja)
3.	 -sa.ja	17.	 -jaya(-ja.āi)
4.	 -vowel variants of sa		
5.	 -s.sa.ja		
6.	 -ja.sa		
7.	 -sam		
8.	 -sat(-sa.am)		
9.	 -ām		
10.	 -sa.ām		
11.	 -sām.ja		
12.	 -ta.ja(-am.ja)		
13.	 -am.ja		
14.	 -yam (-āi,ām)		

Table 5. Case-endings of Indus.

Table 5 presents hypothetical values of the case-endings. Except for the case-ending number 7 and related signs, the values are supported by the relation with Brahmi signs. Note that these are essentially the same as proposed by Mitchiner. Wherever there is a difference, Mitchiner's values are shown in brackets. These case-endings are in consonance with the Indus language being an early Indo-Aryan. For a quick reference of the case-endings of Sanskrit, one may consult Whitney.[9]

To repeat Mitchiner's argument [3] as to why the case-endings rule out other languages: The indicator of genitive singular in Sumerian is *-ak*, in Elamite *-na*, in Dravidian *-a* or *-in*.

That the case-endings appear appropriate for an Indo-Aryan language further validates the sound values assigned to the Indus signs in the endings, especially because these signs look strikingly similar to the corresponding Brahmi letters. The probability of this happening by chance would be even lower than the number in Section 3. Also note that the archaeological evidence now supports the hypothesis of the presence of the Aryan people in the Indus area during the period of the Indus civilization.[6]

Furthermore, the charge of circularity that could have been leveled at Mitchiner's original study of case-endings is no longer valid. It cannot be now said that arbitrary values to case-endings have been assigned to support the Indo-Aryan theory. The assignment has been guided by comparing the frequencies of the most common signs.

### THE COMMON COMBINATIONS

The most frequent pairwise combinations of the Indus signs are shown in Figure 3. The data has been taken from the Mahadevan Concordance.[8] The frequency of a pair has been shown at the corresponding link. Figure 3a shows all pairs that occur more than 40 times, and Figure 3b shows all pairs that occur more than 20 times. These figures allow us to see the common case endings, as well as the most frequently occurring sign combinations graphically. Note that we can examine combinations that are more than 2 signs long.

An examination of Figure 3 reveals that the most frequent links are between what are clearly basic signs. As the frequencies decrease the signs that come into play are increasingly compound and some are (perhaps) polysyllabic words. This implies that syllabic representation in Indus is not as systematic as in Brahmi.

As to the Indus language, it is clear that words are constructed using a system of prefixes and suffixes, a process that occurs in Indo-Aryan languages as well.

It might be said that the graphs of Figure 2 have excessive directional structure for the signs to represent a syllabary. Two arguments may be made against this view. First, as new graphs are made where the number of pairs is much less than 20, the directionality reduces greatly. Secondly, the structural characteristics of graphs such as in Figure 3 cannot represent the structure of the underlying language since the length of the inscriptions is so short. Note also that many compound signs of the Indus script are words. Also, as true for later Brahmi records, *many signs are abbreviations* for commonly occurring words. The evidence does not rule out a core group of Indus signs representing a syllabary.



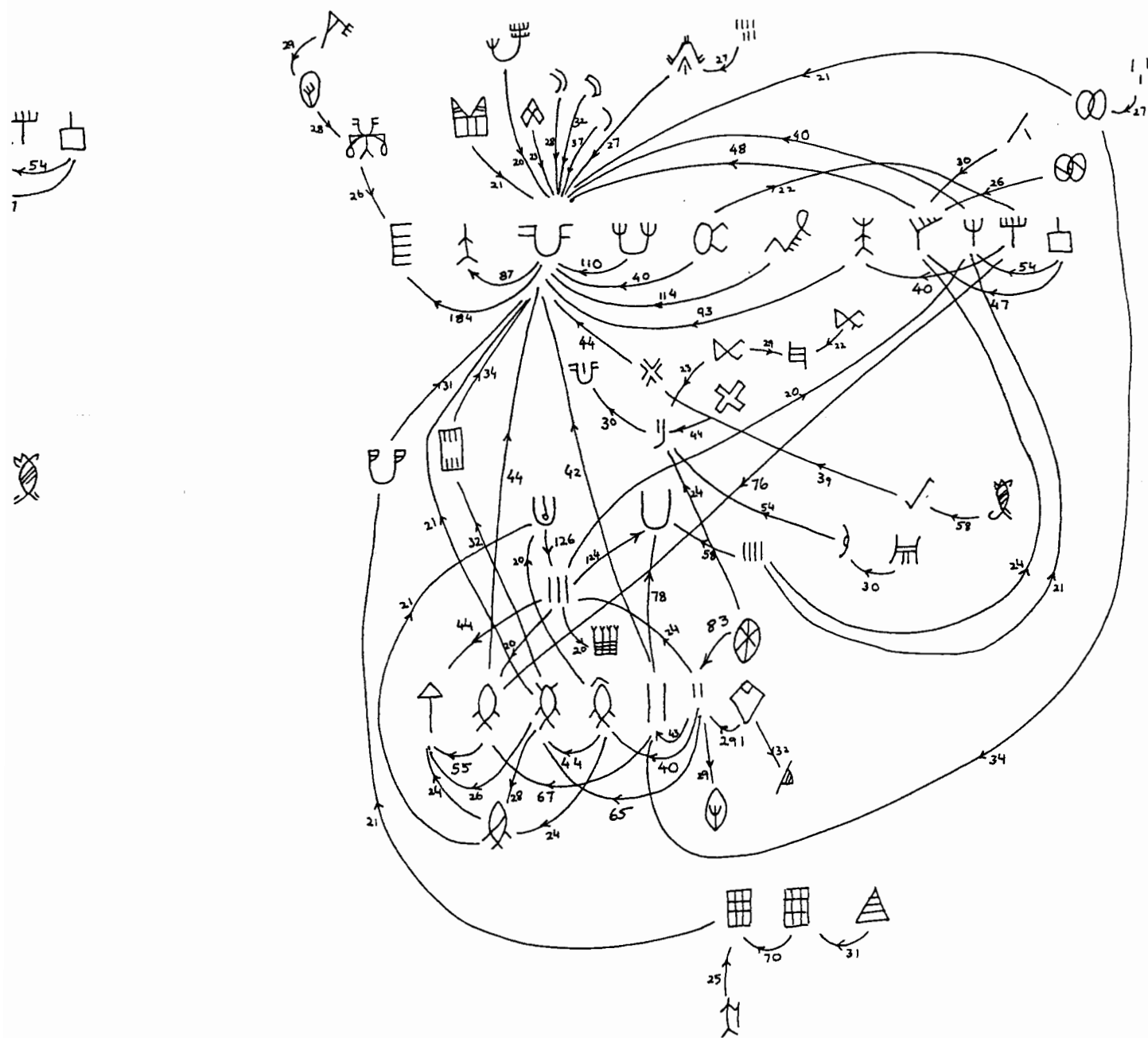


Figure 3b. Indus signs with more than 20 pairwise appearances.

## NUMERALS

Consider numerals now. Frequency considerations suggest that U should be the symbol for the number 5. This is seen in Figure 3b which shows that there is large probability that U will appear together with symbols for 2, 3, and 4. The frequency of U with 1 is 7 and with itself is 10 which do not, therefore, show up in Figure 3b. Presumably, the lower frequency for UI is because 6 is also written as six vertical strokes. It also appears that 10 is represented both as UU as well as one of the other signs. The identification of U as 5 means that Mitchiner's assumption that it might stand for 100 is invalid.

It is noteworthy that the later Nagari sign for 5 is this U with a stylized tail added to it. Also Brahmi pa is U, which looks very close to this sign. Note further that the symbol for 5 in Brahmi comes from the first syllable of *panca*. The fact that the same symbol was used by the Harappans indicates that their word for 5 started with *pa* as well. This is further evidence against the theory of Dravidian origin of the Indus language since 5 in Tamil is *aindu*, in Telegu *aidu*. It reinforces our identification of the Indus language as being Indo-Aryan.

## CONCLUSIONS

The frequency analysis of the most common Brahmi and Indus signs confirms the hypothesis that the two scripts are related. The case-ending evidence suggests that the language of the inscriptions is Indo-Aryan. The inference that the language is Indo-Aryan is strengthened by the observation that the words that follow the formulae "◇", "⊗", which Hunter has argued should be proper names, indeed read as plausible Indo-Aryan names at several places.

But an analysis of the case-endings alone has its limitations. It cannot, by itself, establish conclusively that the language is Indo-Aryan. That will have to await a full decipherment of the Indus texts. In any event, the demonstration that Brahmi is derived from Indus, and the indubitable relationship between Brahmi and the Phoenician script indicates that the theories of the rise of early writing systems require a complete revision.

## REFERENCES

1. Kak, S. C. 1987. The study of the Indus script - general considerations. *Cryptologia*. 11:182-191.
2. Hunter, G. R. 1934. *The Script of Harappa and Mohenjodaro and Its Connection with Other Scripts*. London: Kegan Paul, Trench, Trubner & Co.
3. Mitchiner, J. E. 1978. *Studies in the Indus Valley Inscriptions*. New Delhi: Oxford and IBH.
4. Pandey, R. B. 1952. *Indian Palaeography*. Banaras: Motilal Banarsidass.
5. Kak, S. C. 1987. On the decipherment of the Indus script - a preliminary study of its connections with Brahmi. *Technical Report, L.S.U.* Also in *Ind. J. of History of Science*. 22.
6. Dyson, Jr., R. H. 1982. Paradigm changes in the study of the Indus civilization. *Harappan Civilization - A Contemporary Perspective*. Edited by G. L. Possehl. Warminster U.K.: Aris & Phillips.
7. Koskeniemi, S., A. Parpola, and S. Parpola. 1973. *Materials for the Study of the Indus Script*. Helsinki: Suomalainen Tiedekatemia.

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

9. Whitney, W. D. 1888. *Sanskrit Grammar*. Reprint, 1983. Delhi: Motilal Banarsidass.

10. Basham, A. L. 1963. *The Wonder That Was India*. New York: Hawthorn.

l be the symbol  
probability that  
U with 1 is 7  
nably, the lower  
appears that 10 is  
U as 5 means

ed tail added to  
that the symbol  
me symbol was  
This is further  
Tamil is *aindu*,  
Indo-Aryan.

the hypothesis  
language of the  
trengthened by  
ich Hunter has  
several places.  
itself, establish  
ecipherment of  
Indus, and the  
at the theories

s. *Cryptologia*.

onnection with

ii: Oxford and

ry study of its  
of *Science*. 22.  
ation. *Harap-*  
minster U.K.:

'y of the *Indus*