

Indus-Dravidian Philology

S.K.Venkatesan
CQRL Bits, Chennai

This article establishes a close connection between Dravidian languages and the Indus script, as deciphered by the author [1]. It then further hypothesizes that the agglutinative proto-Dravidian language (Tamil or Tamir, with retroflex “r”) was the creative structural output of the Indus Valley Civilization (IVC) production system, which unified a vast area from Afghanistan to Gujarat. It has been argued from the existence of several languages now in those regions that in the past also there must have been many more languages in use in the IVC and that a single language could not have been in use in those regions. To counter such an argument, one can say with confidence that many such languages of India after colonial rule have been replaced by “English” as a bridge language in India, Africa and the West Indies. Before the British rule, during the Moghul rule, Hindustani was used as the bridge language. Even before that Sanskritization and its broad influence over Indian languages is well observed. So in essence if a broad overwhelming production system takes over an ancient society, the language of the newly emerging class becomes the language of the region, at least initially in major production activity such as trade and other transactions. In the case of IVC such a language, the Dravidian language, seems to have emerged into existence with its mild but persuasive class society.

It has been estimated independently using Dravidian etymology [2,3] that the proto-Dravidian language emerged roughly around 2500 BCE, coinciding roughly with the mature period of the IVC. In this work, by the decipherment of the Indus script and the broad connections that it establishes, we show the formation of the root-word structure of the proto-Dravidian language. Before establishing such deep connections, through the use frequency statistics we show some broad correlations.

The alphabetic system of vowel and consonants is an excellent robust system for writing spoken languages. Most alphabetic systems have around 25-35 consonants and vowels, which also roughly corresponds too the number of phonemes in each language. However, these alphabetic systems are universal systems and can in principle can fit any language, i.e., any language can be written using any alphabetic system (say to 95% accuracy). Even the phonemic systems as covered by IPA Unicode characters also have certain universality about them and they are able to cover a broad range of language families. This shows that these are not the minimal writing systems for each of those languages, i.e., they are not writing system that uses the minimal number of symbols to represent them. In a ideal system that is tightly wedded to the language, it should not only be able cover the phonemic possibilities but also connect the semantic-pun of the language from which the language was built-up from its atomic constituents. This is quite difficult as these languages have borrowed words from each other and have got mixed-up quite a bit, especially after the globalization and merger and synthesis of many cultures. The Dravidian languages have also gone through many such changes over centuries, especially the influence of Sanskrit, Arabian, Persian and recently by English.

In this article we attempt the miraculous feat of establishing that the Indus script as the minimal possible fit to proto-Dravidian. The reconstruction of Proto-Dravidian as we see today is mostly been through the work of many Dravidian scholars [2,4]. We use the Dravidian Etymology Dictionary, DEDR [4] as the main source of our study.

Consonant frequency finger print of Indus script and Dravidian

The frequency of occurrence of letters of alphabets of languages is a good starting point for creating a finger print of languages, like the most frequent letters arranged in rank order in the Wikipedia text of different

languages [5]. Too broad statistical study of languages do not have the distinguishing power to differentiate languages, instead they show the universality of the human larynx and it's capacity. Everett [6] showed that there is a global similarity in frequency of occurrence of consonants in the intralinguistic corpora. There are a few consonants that are unique to some languages and are good at identifying those languages. However, in general the frequency finger print do not tell us much about the structure and morphology of the language, especially how they were historically constructed from their primitive root-word stems.

There is another intelligent and subtle way to construct unique finger print of languages, i.e., by using frequency statistics on the basic list of words constructed by Morris Swadesh [7,8]. He used a list of 207 words translated in multiple languages as the basis for comparing languages. We will use frequency statistics on the translated words. In **Table 1** we display the statistics on starting letters on the Swadesh words for different languages.

Table 1 Frequency of first letter for Swadesh word list of some languages

Language	k/g		p/b		m	a	n	t/d		e	i	v/u		r	s/c	
Elam	11	2	6	7	16	23	11	9	2	4	13	0	3	6	9	0
Sumer	13	20	11	9	22	17	20	15	21	6	10	0	17	7	19	0
Ashkun	24	7	15	8	12	23	13	10	16	0	7	14	5	0	15	7
German	13	11	1	17	9	7	10	5	18	13	3	6	2	12	51	0
Persian	19	12	18	21	13	20	15	9	22	6	5	3	1	12	33	14
Hindi	24	12	56	18	18	22	9	16	25	3	5	4	3	4	23	13
Bengali	22	11	25	34	17	14	8	12	15	5	2	1		4	19	16
Gujarati	20	12	30	22	14	18	11	21	13	3	3	9	5	5	25	11
Marathi	27	12	35	19	17	21	13	23	23	3	7	11	5	7	16	15
Telugu	34	9	40	11	28	23	25	19	3	17	11	19	4	7	9	21
Tamil	47	0	43	0	38	38	37	31	0	21	21	27	11	4	0	18
Malayalam	54	0	39	4	37	29	30	27	3	11	14	28	7	6	10	14
Kannada	38	8	4	19	22	20	23	16	4	12	7	0	9	3	8	5
Tulu	16	1	10	7	8	23	5	6	1	11	10	0	4	1	3	2
Brahui	11	6	14	16	9	17	8	6	14	2	5	1	2	3	12	8
Korean	7	33	12	26	30	15	27	9	22	10	8	0	3	0	28	7
Finish	35	0	28	0	26	4	8	24	0	5	6	14	2	5	28	0
Turkish	42	15	2	28	2	18	7	11	30	6	15	3	6	1	24	1
Mandarin	2	11	5	14	6	0	16	17	17	0	0	0	0	10	14	13
Sanskrit	24	10	38	11	18	28	15	18	17	4	3	23	2	9	39	7

We can construct multi-letter finger print (Swadesh-ID) using frequency of occurrence of initial consonants (that occur in the beginning of the word) in different languages (Table 2).

Table 2 Swadesh-ID of some languages

Language	Swadesh-ID	Occurrence frequencies
Sanskrit	PCTKVMN	(K/G:34)(P/B:49)(T/D:35)(M:18)(N:15)(V/U:27)(S/C:46)
Ashkun	KTPVNMC	(K/G:31)(P/B:23) T:26 S/C:22 M:12 N:13 A:23 E:0 I:7 V:19
Turkish	KTPCVNM	(K/G:57)(P/B:38)(T/D:41)(M:2)(N:7)(V/U:9)(S/C:25)
Finish	KPCMTVN	(K/G:35)(P/B:28)(M:26)(T/D:24)(N:8)(V/U:14)
Korean	KPCTMNV	(K/G:40)(P/B:38)(S/C:35)(T/D:31)(M:30)(N:27)(V/U:3)
Mandarin	TPKCVNM	(K/G:13)(P/B:19)(M:6)(N:16)(T/D:34)(V/U:11)(S/C:27)
Brahui	PTCKMNV	(P/B:30)(T/D: 20)(S/C:20)(K/G:17)(M:9)(N:8)(V/U:3)
Tulu	KPMTNCV	(K/G:17)(P/B:17)(M:8)(T/D:7)(N:5)(S/C:5)(V/U:4)
Kannada	KPNMTCV	(K/G:46)(P/B:23)N:23M:22(T/D:20)(S/C:13)V:0
Malayalam	KPMVNTC	(K/G:54)(P/B:39)M:37(V/U:35)N:30(T/D:30)(S/C:24)
Tamil	KPVMNTC	(K/G:47)(P/B:43)(V/U:38)M:38N:37(T/D:31)(S/C:18)
Telegu	PKCMVNT	(P/B:51)(K/G:43)(S/C:30)M:28(V/U:27)N:25(T/D:22)
Marathi	PTKCMVN	(P/B:54)(T/D:46)(K/G:39)(S/C:31)M:17(V/U:16)N:13
Gujarati	PTKCMVN	(P/B:52)(S/C:36)(T/D:34)(K/G:32)(V/U:14)M:14N:11
Bengali	PCKTMNV	(P/B:59)(S/C:35)(K/G:33)(T/D:27)M:17N:8(V/U:1)
Hindi	PCKTMNV	(P/B:74)(T/D:41)(S/C:36)(K/G:36)M:18N:9(V/U:7)
Persian	CPTKNMV	(S/C:47)(P/B:39)(T/D:31)(K/G:31)N:15M:13(V/U:4)
German	CPTKNMV	(S/C:51)(P/B:18)(T/D:23)(K/G:24)N:10M:9(V/U:8)
Sumer	TKMPNCV	(T/D:36)(K/G:33)(M:22)(P/B:20)(N:20)(S/C:19)(V/U:17)
Elam	MKPNTCV	(K/G:13)(P/B:13)(M:16)(N:11)(T/D:11)(S/C:9)(V/U:3)

Table 2 data helps distinguish between the different families of languages, especially between Dravidian and Indo-European languages. This also points to the cross-pollination between Indian languages, especially the strong influence of the Indus-Dravidian substratum on languages like Hindi and Bengali. We can also see lateral drift of Telugu to a South Central Dravidian. Tamil language has tried to avoid Sanskritization through various political and social movements, thus showing a distinct characteristics, but as can be seen at the Swadesh-ID finger print, that Kannada, Malayalam and Tamil show a clear similarity within the Southern Dravidian, with Telugu showing the different characteristics of a South Central Dravidian language.

Using the decipherment [1] we were able to report on the frequency of occurrence initial consonant of Indus Signs by means of cumulative concordance frequency of signs grouped together on the basis of initial consonants.

Table 3 Indus script frequency table

M77 sign numbers	Initial Consonant	Glyph	Concordance Frequency
59+67+70+72	k*	𑀭++	910=373+276+73+188
373+375+379+391+403	p*	𑀮++	422=61+57+17+195+92

M77 sign numbers	Initial Consonant	Glyph	Concordance Frequency
287+293+298	v*	𑀭++	226=86+135+5
53+225+222+229	t*	𑀭𑀭++	160=129+4+22+5
197+198+201+202+204	m*	𑀭𑀭++	154=56+4+9+9+76
374+400+321+402	n*	𑀭𑀭++	130=9+14+9+98
240+241+358+381	c*	𑀭𑀭++	50=3+1+32+14

We can conclude from this table that the equivalent Swadesh-ID for Indus script is KPVTMNC. This Swadesh-ID has highest similarity with the Southern Dravidian languages. If we study the Levinshtein copyediting distance between the Indus script Swadesh-ID (KPVTMNC) and the Swadesh-IDs of languages listed in Table 2, we can conclude that Tamil Swadesh-ID has the smallest Levinshtein copyedit distance (d=2) from the Indus Script Swadesh-ID. However, an improved copyedit distance metric with weights for rank-diff creates a better distance metric as shown in Table 4 below.

Table 4 Languages and their Swadesh distance with Indus Script

Language	Swadesh-ID	Edit distance (Levenshtein)	Edit distance by rank difference
Indus script	KPVTMNC	0	$7 1-1 +6 2-2 +5 3-3 +4 4-4 +3 5-5 +2 6-6 + 7-7 = 0$
Brahui	PTCKMNV	3	$7 1-4 +6 2-1 +5 3-1 +4 4-2 +3 5-5 +2 6-6 + 7-3 = 49$
Korean	KPCTMNV	2	$7 1-1 +6 2-2 +5 3-7 +4 4-4 +3 5-5 +2 6-6 + 7-3 = 24$
Finish	KPCMTVN	4	$7 1-1 +6 2-2 +5 3-6 +4 4-5 +3 5-4 +2 6-7 + 7-3 = 28$
Turkish	KTPCVNM	5	$7 1-1 +6 2-3 +5 3-5 +4 4-2 +3 5-7 +2 6-6 + 7-4 = 39$
Sanskrit	PCTKVMN	5	$7 1-4 +6 2-1 +5 3-5 +4 4-3 +3 5-6 +2 6-7 + 7-2 = 61$
Ashkun	KTPVNMC	3	$7 1-1 +6 2-3 +5 3-4 +4 4-2 +3 5-6 +2 6-5 + 7-7 = 24$
Mandarin	TPKCVNM	5	$7 1-3 +6 2-2 +5 3-5 +4 4-1 +3 5-7 +2 6-6 + 7-4 = 45$
Tulu	KPMTNCV	4	$7 1-1 +6 2-2 +5 3-7 +4 4-4 +3 5-3 +2 6-5 + 7-6 = 29$
Kannada	KPNMTCV	4	$7 1-1 +6 2-2 +5 3-7 +4 4-3 +3 5-4 +2 6-3 + 7-6 = 34$
Malayalam	KPMVNTC	4	$7 1-1 +6 2-2 +5 3-4 +4 4-6 +3 5-3 +2 6-5 + 7-7 = 21$
Tamil	KPVMNTC	2	$7 1-1 +6 2-2 +5 3-3 +4 4-6 +3 5-4 +2 6-5 + 7-7 = 13$
Telugu	PKCMVNT	5	$7 1-2 +6 2-1 +5 3-5 +4 4-7 +3 5-4 +2 6-6 + 7-3 = 42$
Marathi	PTKCMVN	6	$7 1-3 +6 2-1 +5 3-6 +4 4-2 +3 5-5 +2 6-7 + 7-4 = 48$
Gujarati	PTKCMVN	6	$7 1-3 +6 2-1 +5 3-6 +4 4-2 +3 5-5 +2 6-7 + 7-4 = 48$
Bengali	PCKTMNV	4	$7 1-3 +6 2-1 +5 3-7 +4 4-4 +3 5-5 +3 6-6 + 7-2 = 45$
Hindi	PCKTMNV	4	$7 1-3 +6 2-1 +5 3-7 +4 4-4 +3 5-5 +3 6-6 + 7-2 = 45$
Persian	CPTKNMV	5	$7 1-4 +6 2-2 +5 3-7 +4 4-3 +3 5-6 +2 6-5 + 7-1 = 56$
German	CPTKNMV	5	$7 1-4 +6 2-2 +5 3-7 +4 4-3 +3 5-6 +2 6-5 + 7-1 = 56$
Sumer	TKMPNCV	6	$7 1-2 +6 2-4 +5 3-7 +4 4-1 +3 5-3 +2 6-5 + 7-6 = 60$
Elam	MKPNTCV	6	$7 1-2 +6 2-3 +5 3-7 +4 4-5 +3 5-1 +2 6-4 + 7-3 = 57$

These results conclusively show the close affinity between Dravidian languages and the Indus script as deciphered by the author [1].

Origins of proto-Dravidian language as rooted in the Indus script

In this short section we will describe some of the root words of Dravidian language and it's relationship to the Indus Script. For more details please refer to the actual decipherment [1].

The numbers as syllabic affixes

(1a) | – -al, mutal (syllable, first)

- (1b) 𒌦 – -ir, iru (syllable, vast)
- (1c) 𒌦𒌦 – irū (two)
- (1d) 𒌦𒌦 – mū (mature)
- (1e) 𒌦𒌦𒌦 – mūvū (three)
- (1f) 𒌦𒌦 – na! (good)
- (1g) 𒌦𒌦𒌦 – nālū (four)
- (1h) 𒌦𒌦𒌦 – in-ai (join)
- (1i) 𒌦𒌦𒌦 – aitū (five)
- (1j) 𒌦𒌦𒌦 – ār (river)
- (1k) 𒌦𒌦𒌦 – ārū (six)

The human affix

- (2a) 𒀭 – -an: avan (person, human)
- (2b) 𒀭𒀭 – annan (brother)
- (2c) 𒀭𒀭 – kāvalan (guardian)
- (2d) 𒀭 – mutalvan (chief)
- (2e) 𒀭, 𒀭 – vēlan, vēṭan (hunter)

Ūr and it's various manifestations

The original role model of a town, the Sumerian town:

- (3a) 𒌦 – ūr – town/village – origin – farmer's town

It has deep connections to the Dravidian language. All towns first start as farming villages where they melt the soil to increase fertility. The earthworm also helps in melting the soil, especially in the black soil:

- (3b) 𒌦𒌦𒌦 nālam-nilam - fertile land, ricefield (DEDR 2913, 3676)
- (3c) 𒌦 – plough
- (3d) 𒌦 (DEDR 661) – melt the soil
- (3e) 𒌦 – ūṛavan – farmer

Such a fertile farmer's town creates a great cogglomeration of near relatives:

- (3f) 𒌦, 𒌦 (DEDR 710)– to be close, nearness, relation
- (3g) 𒌦 – u! – inside
- (3h) 𒌦 – ū – flesh, body, thing

The sun in other ancient scripts

The sun ☼ is quite central to life in early civilizations. In the Indus script, we associate the oval sign ○ with the syllable “pakal” (daylight) and so in essence it is the syllable “pa”. The “pa” has many bright-white

associations such as “pal” (teeth), “pala” (many teeth), “pāl” (milk), “pallā” (tusk, elephant), “pantu” (ball), “pattu” (ten – many). In the ancient Chinese characters also the sun is also denoted by the oval sign ☉ which then regularized later on to the rectangular form 日, with a combination of sun (日) and the moon (月): 日月 being the common abstraction, the brightness. Just like the Dravidian mild-male association of “appan”, “appu”, and “appa”, in the Egyptian hieroglyph also it leads to a patriarchal bright sun the “ra” (king-god). In the Indo-European, the etymological root of the “sun” is also the feminine “*sāwel/sunne” similar to the feminine Greek goddesses “Sol” (sun) and “Luna” (moon). The “*sawelyo” eventually morphed into masculine “helios”.

The sun in Indus: the “pa” root syllable

- (4a) ○ – pa: pakal (sun, daylight)
- (4b) ① – pal, pāl, pala (teeth, milk, many)
- (4c) ② – pallā, pillir (tusker, elephant trumpet)
- (4d) ③ – pammu (twine, to drum)
- (4e) ④ – pū (flower, insect)
- (4f) ⑤ – pol (bright)
- (4g) ⑥ – par, parai (drum)
- (4h) ⑦ – pan, pan, paṇ, pān (palm tree, agriculture land, work, song)

The bird

- (5a) ⑧ – porī, purī, purā, purā, pūr (courage, pride, chicken, pigeon, dove, quail, pheasant)
- (5b) ⑨ – irakkam, irappu, irakkam, irappu (mercy, death, loss)

The frog-pot: the “ta” root syllable

- (6a) ⑩ – ta: tavaḷa (frog, small pot)
- (6b) ⑪ – tattū, tatti (leap, jump)
- (6c) ⑫ = ⑩ + ⑬ – taḷu, tadu (push, block)
- (6d) ⑭ – mū-taṛi (burial pot)
- (6e) ⑮ – toṭu (touch)
- (6f) ⑯ – tai (weave)
- (6g) ⑰ – taṭṭi, taṭṭu, taṭṭāṇ (mat, strike, carpenter, bronze smith)

The fish-boat: the “ka” root syllable

- (7a) ⑱ – kanni (fish, boat, vehicle, trap)
- (7b) ⑲, ⑳ – kāl, kōl, kol (vehicle, sail-boat, accept)
- (7c) ㉑ – konde (bullock)
- (7d) ㉒ – kō (cattle herder chief)
- (7e) ㉓ – kōṭa (fort)

(7f) 𑌕 – kōṭṭam (fort, congregation)

(7g) 𑌕 – kōṭṭai (fortress, united fort)

The net: the “va” root syllable

(8a) 𑌕 – vala, valai, vaḷai, vāḷai (right, net, bent, a fish)

(8b) 𑌕, 𑌕, 𑌕 – valaiyan (fisherman, trapper)

(8c) 𑌕 – iṭa, iṭai (gap, left)

(8d) 𑌕 – iṭaiyan (herdsman)

Tiny zero: the “na” root syllable

(9a) 𑌕 – no, noccu, noyya (small, tiny, minute)

(9b) 𑌕 – ñāṇ (thread)

(9c) 𑌕 – nuḷḷāṇ, nuṇan (mosquito, tailor)

The chip: the “ca” root syllable

(10a) 𑌕 – cī, cīy, cīpu (cleanse, chisel, wooden peg)

(10b) 𑌕 – cil, calli (broken piece, splinter)

(10c) 𑌕 – ciṛa (eminent, illustrious)

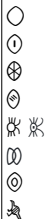

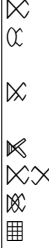



The meluha or the raised floor: the “ma” syllable

(11a) 𑌕 – mē, mēl, mēṭu, mettai (high, superior, above, raised)

(11b) 𑌕, 𑌕 – māṭi, māṭu (upper storey, bull)

(11c) 𑌕 – mēṭṭu, mēḍu, metta (height, eminence, hillock)

K-P-V-T-M-N-C starting consonants	Glyphs	Dravidian	English Description (proposed new Indus Swadesh list of words)
ka	𑌕 𑌕 𑌕 𑌕 𑌕 𑌕 𑌕 𑌕	kanni kāḷ, kaḷe, kalu, kalam kōḷ, koḷ konde kō, kōṭa kōṭṭam, kōṭṭai	boat, fish, trap vehicle raft, boat bullock (cart) hearer, chief, fort fortress

pa		pakal, pagaṭi, hāgalu, pagelda pal, pallu, hallu, dantālu pāl, pālu, hāllu, pala, cālā, halavu, valare pallā, piḷḷir pammu, pammal, hammu pū, pū, pūvu, pāi, poḷ paṛa, dolu, paṛai paṇ, pan porī, poru, puri, purā, paṛa, puṛa, paṛa pūr	daylight teeth milk many tusker, elephant trumpet twine flower, to flower drum, drumming, message work chicken, strength, courage, pride, love, pigeon, quail
va/u		val, vala, valai, vaḷai, vāḷai valaiyan ū, ūla uḷ, uḷe ūr	right, net, trap, scabbard fish trapper, fisherman flesh, body inside, in town
ta		tavaḷa, tavaḷa tattū, tatti tala, tale, talai taḷu tadu taṛi toṭu tai taṭṭi, taṭṭu, taṭṭān, taṭṭaṇ	frog, small pot leap head push block burial pot touch stitch mesh, tinker, smith
ma		mē, mēṭu, mettai māṭi, māḷikai, māṭam mēṭṭu, mēḍu, metta	above upstairs terrace
na		no nān, nān nullan	tiny, null thread musquito, tailor
ca		cipu cil ciṛa ciḷu-ciḷ-eṇal	cleanse, chisel, chip wooden peg fine, excellent shining, jingly, lively

Indus script - short Swadesh list

Syllable	English Swadesh list (short version)	Morphemes	Dravidian variations and expansions
mutal	first, prime	-al, -il, -ol, -ul, -l	mutal, mutalai (that which came first), modaṭi, modalu
iru	vast, exist, stable, person, two	-ar, -ir, -or, -ur, -r	iru (exist), iruvai. irukku, iruppu, iruutu, iruttam, irutti, irumai (greatness)
mū	mature, ripe, three	mū-, -mū	mū, mutu, mutāri, mūppaṇ, mutiruka
nal	good, four	nal-	nal, nalla, nalatu, nallavai, nallatōr, nalam, nallavar, nalavu, nalku, nallē, nalmē, nelā
inai	join, five	-ai	inai, iṇa, eniyu
avan	him, human	-an	avan, avvōṇ
avvōṇannan	elder brother	-anan, -aṇṇal	aṇ, aṇṇaṇ, aṇṇe, aṇṇa, anna, aṇṇal
kāvalaṇ	guardian	kā-	kāvalaṇ, kākkai, kaval, gavana, kavanam
vēlan	lancer	vē-l-an	vēlan, vēlāṇ, vēli
vēṭan	hunter	vē-ṭ-an	vēṭan, vēṭṭuvaṇ, vēṭṭai, vēṭṭa
ūr	town, village	ū-r	ūr, ūru, uṛu
ñālam	black soil	nal-mū	ñālam
nilam	land	nil-mū	nilam
uṛu	melt	ū-r-ū	uṛu
urukku	melt	ū-r-ū-ku	urukku
uṛavan	tiller	ū-r-an	uṛavan

Syllable	English Swadesh list (short version)	Morphemes	Dravidian variations and expansions
mutalvan	first one	mutal-an	mutalvan
uṛu	relate	ū-r-ū	uṛu
uṛavu	relation	ū-r-ū	uṛavu
ul	inside	ū-l	uḷ, uḷḷu
ū	meat, body	ū-, -ū	ū
pakal	daylight	pa-, -pa	pakal, pagaṭi, hagalū, pagelda
pal	teeth	pa-l	pal, pallu, hallu, dantālu
pāl	milk	pa-l	pāl, pālu, hāllu,
pala	many	pa-l	pala, cālā, halavu, valare
pallā	tusker	pa-l-āru	pallā
piḷḷir	elephant trumpet	pa-l-āru	piḷḷir
pammu	twine	pa-mū	pammu, pammal, hammu
pū	flower	pa-ū	pū, pū, pūvu, pāi
poḷ	to flower	pa-ū-l	poḷ
paṛa	drum	pa-ra	paṛa, dolu
paṛai	drumming, message	pa-r-ai	paṛai
paṇ	work	pa-no	paṇ, pan
pān	song	pa-no	pān, pāṇ, pāṇu, pāṇṇu, pāṇan
pori	chicken, trap,	paṛa-	pori, poru
puri	strength, courage, pride, love	-puri	puri
puṛā	pigeon	paṛa-	puṛā, paṛa
puṛa	pigeon	paṛa-	puṛa, paṛa
pūṛ	quail	pūṛ-	pūṛ
iṛakkam	pity	iṛa-	iṛakkam
iṛappu	death	iṛa-	iṛappu, iṛappu, iṛakkam
tavaḷa	frog, small pot	ta-	tavaḷa, tavalā
tattu	leap	-tattu	tattu, tatti
tala	head	ta-l	tala, tale, talai
taḷu	push	ta-l-ū	taḷu
tadu	block	ta-d-ū	tadu
taṛi	burial pot	taṛ-	taṛi
toṭu	touch	ta-ṭu	toṭu
tai	stitch	ta-inai	tai
taṭṭi	mesh	-taṭṭi	taṭṭi
taṭṭu	tinker	-taṭṭu	taṭṭu
taṭṭāṇ	smith	taṭṭi-an	taṭṭāṇ, taṭṭaṇ
kanni	trapper	ka-	kanni
kāl	vehicle	ka-l	kāl, kaḷe, kalu, kalam
kōl	boat	ka-l	kōl
koḷ	accept	ka-l	koḷ
konde	bull	-konde	konde
kōṭa	fort	kō-ṭa	kōṭa
kōṭṭam	fortress	kōṭṭa-mū	kōṭṭam
kōṭṭai	fortress	kōṭṭa-ai	kōṭṭai
vala	right	vala-	vala
valai	net, bend, scabbard fish	vala-ai	valai, vaḷai, vālai
valaiyan	netter	valai-an	valaiyan
iṭa	gap, left	iṭa-	iṭa, iṭai
iṭaiyan	herder	iṭa-an	iṭaiyan
no	tiny	no-, na-,	no, noccu, noyya
ñāṇ	thread	no-no	ñāṇ
nuḷḷāṇ	mosquito, tailor	no-l-an	nuḷḷāṇ, nuṛan

The minimal parametric description of proto-Dravidian

In this section we describe proto-Dravidian using very minimal syllabic spare parts, i.e., roughly around 16 syllables. These syllables are root-word syllables and, in principle, rest of the words of the language could be constructed by an aglutinative process that allows modification/insertion of vowels or glides (glides are soft vowels that act as bridge between hard consonants). We shall here provide such a description.

The root-word consonant syllables

Syllable	ka	pa	va	ta	ma	na	ca
Glyph	◇, 𐑇	○)	⋈	𐑃, 𐑄	◌̣	𐑆, 𐑇
M77 no.	261, 57	373	287	216	197, 202	374	240

The post/pre-position suffixes

Syllable	-al	-ir	mu-, -m	nal-	-ai	-ār	ēr-, ēṛ	ū-, -ū	-an, an-
Glyph								U	𐑉
M77 no.	86	87	89	95	96	108	110	328	1

As you can observe that these suffixes occur both as post and pre-position sufffixes. In the case of “iru” (two) they occur as multiple copies also:

⋈ (pa+iru = paṛa, drum)

◊ (cil+iru = ciṛa, illustrious)

In the case “nal” (four), they occur in many different ways, as four sticks surrounding as in say:

𐑉𐑆𐑇 (nal-aniyal-pala, good-jewels-many)

⋈ (nāl-vaṛi, four-cross-way-place).

In the case “āru” (six), they occur as six sticks in an axis:

⊗ (pala-ār = pallā, elephant forehead plate, the elephant)

In any case, the important observation is that Proto-Dravidian can be constructed from the above seven consonant syllables and nine suffixes, making them a total of 16 parameters (building-blocks) to fit the Proto-Dravidian language. It is hypothesized here that these are the minimal lego pieces of the logo-syllabic script that were used to construct the language. It can be noticed that these are much smaller than the 25-30 alphabets that are used to construct modern languages.

Dravidian aglutination process

The Dravidian aglutination of words and suffixes are an interesting process of welding and smoothing of the edges (to use the bronze worker analogy). The vowels are modified/deleted/inserted and similarly glides are modified/deleted/inserted to smoothen the joins. Certain ideographic clues also provide extra artistic effects not discernable by strict linguistic rules. These require certain subtle understanding of the vowels and soft-consonant flows and contours.

Conclusion

In this short note we briefly outlined some of the ancient root words of Dravidian that lie at the heart of the logo-syllabic Indus script, a script that is fundamental to the conception and the genesis of a simple proto-Dravidian language, the Indus-Dravidian, a language that germinated from the barter trade transactions of the Indus river transport system. In order to establish this further, a Indus-Dravidian Swadesh list is being proposed, which can further the cause of this decipherment. A similar proposal was published by the author [9], inspired the Chinese characters and an artificial language Lojban, twenty years back. These proposals are much easier to study now, especially with the arrival Large Language Models that are based on the semantic word-embedding layers created by the process of tokenization [10].

References

- [1] Venkatesan, SK, (2025) Decipherment of Indus Valley Script, <https://github.com/Sukii/decipher-ivc>
- [2] Krishnamurti, B, (2003) *The Dravidian languages*. Cambridge, UK: Cambridge University Press.
- [3] Burrow, T., Emeneau, M.B., Dravidian Etymological Dictionary, Oxford Clarendon Press, 1961, (see also: <https://dsal.uchicago.edu/dictionaries/burrow/>)
- [4] Kolipakam V, Jordan FM, Dunn M, Greenhill SJ, Bouckaert R, Gray RD, Verkerk A. 2018 A Bayesian phylogenetic study of the Dravidian language family. *R. Soc. Open Sci.*, **5**, [171504](https://doi.org/10.1098/rsos.171504)
- [5] Vrandečić, D (2012) Letter frequency, <http://simia.net/letters/>
- [6] Everett, C (2018) The similar rates of occurrence of consonants across the world's languages: A quantitative analysis of phonetically transcribed word lists, *Language sciences*, **69**, 125-135
- [7] Swadesh, M (1971). *The origin and diversification of language*. Chicago: Aldine
- [8] Swadesh lists by language, https://en.wiktionary.org/wiki/Category:Swadesh_lists_by_language
- [9] Venkatesan, S.K. (2005), Moving from bytes to words to semantics, *TUGboat*, **26**, 165-168
- [10] Balachandran, A. (2023) Tamil-Llama: A New Tamil Language Model Based on Llama 2, <https://arxiv.org/pdf/2311.05845>