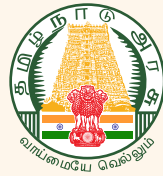


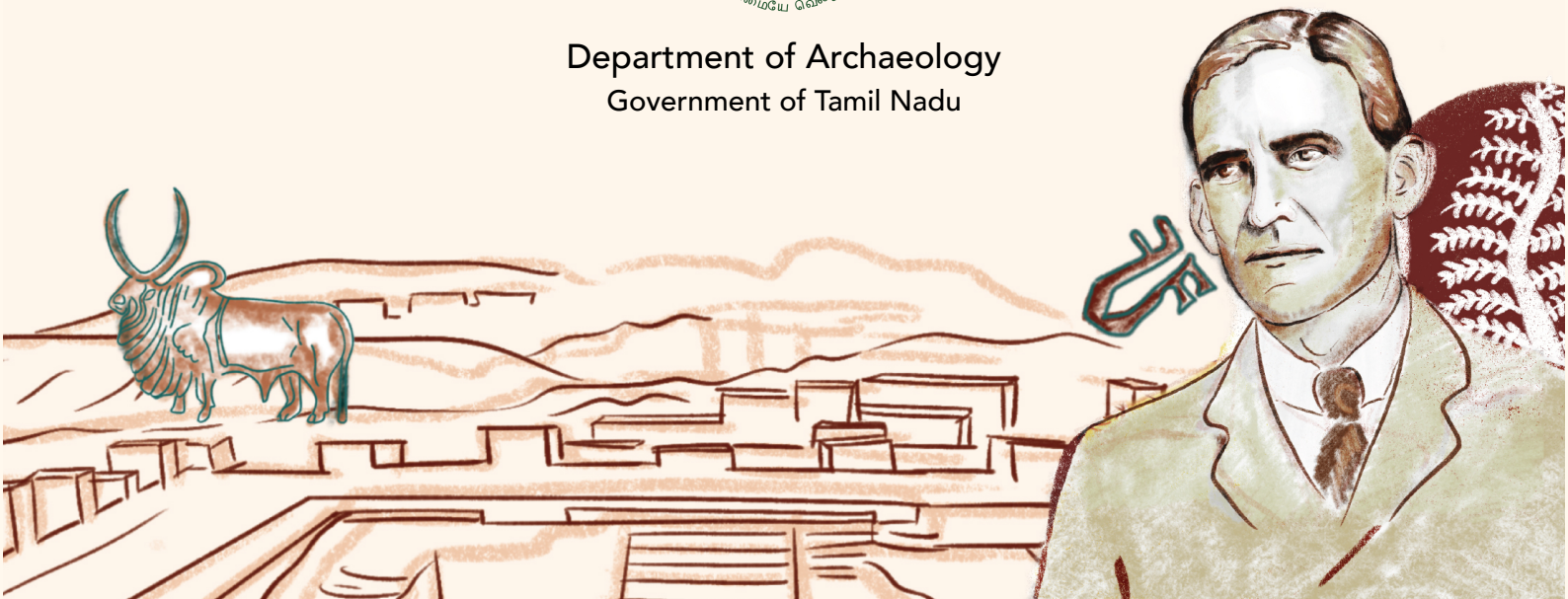
Discovery of Indus Civilization Centenary

Three-day International Conference
Commemorating the Centenary Year of the announcement
of the discovery of the Indus Valley Civilization
January 05 - 07, 2025

Conference Proceedings



Department of Archaeology
Government of Tamil Nadu



DISCOVERY OF INDUS CIVILIZATION CENTENARY

Conference Proceedings

Editors

Prof. K. Rajan
Dr. R. Sivanantham
Dr. Sundar Ganesan



**Department of Archaeology
Government of Tamil Nadu**

Contents

<i>Opening Address</i>	01
Archaeology that Changed the Picture Prof. Romila Thapar	
<i>Keynote Address</i>	07
The Indus Civilization: Looking Back at an Epochal Discovery Dr. Nayanjot Lahiri	
Harappa @ 100: A Catalogue of Photographs of Excavations at Harappa in 1924-25 Dr. Sudeshna Guha	17
New Light on Marshall: Exploring the Archaeological Methods, Approaches, and Perspectives of Sir John Marshall Through His Photographic Collection at Durham University Dr. Christopher Davis	27
Discovery of the Harappan Civilization: Some Questions and Issues that Remain 100 Years Later Mr. Tony Joseph	43
Indus Script and Graffiti Marks: A Comparative Study Prof. K. Rajan and Dr. R. Sivanantham	55
Seals of the Indus Tradition: Current Perspectives and Future Goals Dr. Gregg Jamison	111
The Bead Making: A Cultural Identity of the Harappan People Ms. Dilshad Fatima	143
Indus Valley Civilization and its Material Culture in South India Dr. R.N. Kumaran and Dr. V.P. Yathees Kumar	159
Some Cultural Vestiges of the Indus Civilisation in the Vedic Sources Dr. N. Marxia Gandhi	189

What the Seven Pleiad Sisters Can Tell Us: By Their Presence in Mesopotamia, on IVC Seals, and in Kongu Nadu Today	<i>195</i>
Dr. Brenda E. F. Beck	
Hampi Tunga - Bhadrar River Basin Cave Rock Paintings Resemblance to Indus Seals	<i>233</i>
Dr. K.M. Metry	
Understanding Indus Riddles in the Absence of a Bilingual Script	<i>241</i>
Mr. R. Balakrishnan	
சிந்துவெளிப் பண்பாட்டின் திராவிட விழுமியங்களும் அதன் தொடர்ச்சியும்	<i>271</i>
Dr. கி. அமர்நாத் ராமகிருஷ்ணா	
Pleonastic Compounding, Cults and Dynastic Titles: A Few Clues to the Indus Script	<i>289</i>
Mr. Anish Bharadhwaj Lakshmanan	
IVC Deep Astronomy Harappan Celestial Omens List	<i>361</i>
Mr. Sukumar Rajagopal	
Data and Decipherment: Computational Approaches to Indus Writing	<i>411</i>
Dr. Sitabhra Sinha and Dr. Md Izhar Ashraf	
Indus Script's Gemstone and Precious Shiny Commodity Related Fish-Signs, and Indus Gemstone-Word "Maṇi": Ancestral Dravidian Symbolism in Indus Logograms and Language(s)?	<i>423</i>
Ms. Bahata Ansumali Mukhopadhyay	

Data and Decipherment: Computational Approaches to Indus Writing

*Dr. Sitabhra Sinha and Dr. Md Izhar Ashraf
Computational Epigraphy Lab (iCEL)
The Institute of Mathematical Sciences, Chennai*



“A race of hyperintelligent pan-dimensional beings once built themselves a gigantic supercomputer called Deep Thought to calculate once and for all the Answer to the Ultimate Question of Life, the Universe, and Everything.

For seven and a half million years, Deep Thought computed and calculated, and in the end announced that the answer was in fact Forty-two and so another, even bigger, computer had to be built to find out what the actual question was.”

-Douglas Adams, The Hitchhiker’s Guide to the Galaxy series

Among the remarkable changes that the world has seen over the (little more than a) century following the announcement in 1924 of the discovery of the Indus Valley Civilization [Marshall 1924], probably the most relevant from the perspective of the authors is the Information Revolution. This phenomenon, which is still ongoing, came about as a result of the development – and mass adoption – of the digital computer, and subsequently, the ability to rapidly communicate and share information worldwide made possible by the internet. Even more remarkably, the recent advances in creating Large Language Models (LLMs) has transformed and challenged the traditional models of education and research, even if one accepts that general purpose machine intelligence could be some way off. Given their impact on society in general, it’s no surprise that such software have been touted as a “magic bullet” for solving the Indus Puzzle, viz., deciphering the inscriptions that have been excavated in the sites associated with the mature phase of the Indus Valley Civilization (IVC, 2600-1900 BCE). So, is the solution of the puzzle just around the corner?

The lack of any knowledge about both the nature of the writing system, as well as, the possible language or languages that could have been expressed using the symbols, have foiled the hundreds of attempts that have been made to decipher the inscriptions ever since Marshall’s announcement of the existence of a “pictographic language” on the Indus artefacts. The large

number of distinct symbols – between four hundred to eight hundred, depending on how one identifies allographs – have suggested to many serious scholars that it was likely to have been a logo-syllabic script. An example of such a system is the cuneiform script used for writing Sumerian in the city-states of Mesopotamia. Note that the people of Indus civilization had trade connections with the Sumerians as attested by archaeological discoveries in Sumerian sites of artefacts characteristic of the Indus Valley Civilization, such as seals, beads and pottery. Nevertheless, it has not prevented assertions that the inscriptions are written using an alphabetic system. Alphabets typically use about thirty symbols (the Glagolitic script, the oldest known Slavic alphabet, at one time had 41 distinct letters, which is possibly the largest among true alphabets – i.e., alphabets that explicitly represent both vowels and consonants using distinct symbols, in contrast to consonantal scripts or abjads where letters represent consonants with vowel sounds implied, and alpha-syllabaries or abugidas where consonant-vowel combinations are represented using distinct graphemes with the vowel modifier typically indicated by diacritics).

Thus, such claims often go to absurd lengths to demonstrate that the hundreds of distinct graphemes in the Indus inscriptions are somehow reducible to combinations of just a few dozen symbols. While, in principle, it's possible that elements of the Indus signary could have been used to write phonetically, as is the case for Egyptian hieroglyphic writing, especially for rendering names that are foreign to the primary users of the script and hence cannot be rendered using logograms – very often such attempts culminate in reading the inscriptions in languages that were present only after the end of the IVC. One of the authors has had the dubious pleasure of being asked to review for a scholarly journal a manuscript that seriously proposed a reading for Indus inscriptions that jumped between four different Southern Indian languages of the present day, while others have claimed to have read the inscriptions as Sanskrit (the earliest form of which, Vedic Sanskrit, can be dated only as far back as 1500 BCE).

Returning to the theme with which we began, the past few years have already seen several attempts to use “Artificial Intelligence” (AI) to achieve success in an endeavour that humans have so far failed in. Unfortunately, the combination of algorithmic routines that are highly prone to hallucinations (repeatedly giving outputs of made-up “facts”) and the natural imbecility of humans who arrogantly believe that they can solve a problem without having the slightest knowledge of the origin and evolution of different types of writing systems, has only resulted in an accelerated pace with which new claims of decipherments of the Indus inscriptions are being made. The great confidence with which some have asserted that AI is a game-changer for solving the Indus puzzle is somewhat reminiscent of an episode from Douglas Adams' *The Hitchhiker's Guide to the Galaxy* series (see the opening quote of this article): the most powerful computer ever to be built is asked the question of “Life, the Universe and Everything” and after processing information for an extremely long time, gives the answer “42”. The absurdity of the answer of course, arises from the fact that the answer does not make sense without a good understanding of the context of the question that it is the answer to. We would like to argue that much of the effort in applying computational methods to try deciphering the Indus inscriptions may give rise to similar non-sensical results, if these exercises are carried out without any appreciation of why the problem has resisted attempts to solving it for so long.

It is sobering to realize that despite a hundred years' worth of effort, the actual confirmed knowledge about the Indus inscriptions has not moved all that much beyond what was stated in the observations made by Marshall in his 1924 article about the "seals... inscribed with legends in an unknown pictographic script" [Marshall 1924]. Among the "points ...worthy of remark" on the "strange pictographs which do duty for letters", Marshall had suggested quite intriguingly that "the marks... attached to many of the pictographs" were "apparently vowel signs". We are unsure on what basis Marshall stated this, but it could be that he had in mind the method of using diacritics to indicate the vowel used to modify consonants in Indic scripts, even though he went on to say that the signs "bear no resemblance whatever to any ancient Indian alphabet".

It is also possible that Marshall was hinting at the use of signs as phonetic markers or determinatives (to indicate the specific meaning) for a logogram that could not be uniquely interpreted otherwise. If this interpretation is correct, it could be quite prescient, as such use of signs to modify logograms have been seen in all the ancient writing systems of the Old World, including Sumerian cuneiform, Egyptian hieroglyphs and Chinese writing. Marshall also mentions a possibility that the civilization originated "further west", possibly meaning Elam, and as a supporting argument states that "there are linguistic reasons for believing that it was by way of Baluchistan that the Dravidian races... entered India". Note that the relation suggested here between ancient Elam and Dravidian predates by almost half a century the Elamo-Dravidian hypothesis proposed by David McAlpin in the 1970s [Renfrew 1989]. In fact, the resemblance of the Indus seals to Proto-Elamite tablets were pointed out within a week of publication of Marshall's article by the British linguist Archibald Sayce, the author of the first grammar of Assyrian in English [Sayce 1924], and the week after that by Cyril Gadd and Sidney Smith, both pioneers of Mesopotamian studies in the United Kingdom [Gadd 1924] in the pages of the same magazine.

Even more humbling is a study of the Indus inscriptions by G R Hunter that was submitted to the University of Oxford for the degree of Doctor of Philosophy in 1929 and that was published as a book a few years later [Hunter 1934]. It contains the first ever concordance compiled from all the known inscriptions at the time and remained the only corpus in existence till Mahadevan created a digital database in the 1970s.

Many of the conclusions drawn by Hunter based on painstaking manual analysis of the inscriptions, e.g., that most of the inscriptions were written from right to left, or that the patterns of sign co-occurrences of the seals have distinct characteristics, with those found in South Asia differing significantly from those found in West Asia, have later been supported by computerized analysis of sequence statistics performed on much larger corpora [Rao 2009, Ashraf 2018]. Given these insights obtained from an intense analysis of the structure of the inscriptions, some have gone to argue that the discovery of bi-lingual or multi-lingual inscriptions with Indus signs occurring in one of them, is all that is needed to decipher the script. Such an object(s), it is argued, will play the same role as the Rosetta stone in deciphering Egyptian hieroglyphs. Named after the town of Rashid or Rosetta in Egypt, where it was discovered in 1799 by Napoleon's army, the Rosetta stone is a stele inscribed with three versions of a decree, one in hieroglyphic, another in demotic and the third in Greek that was

inscribed in 196 BCE during the Ptolemaic dynasty of Egypt. As Greek was a known language, the stone proved instrumental in allowing Young and Champollion to make breakthroughs in deciphering ancient Egyptian writing.

Unfortunately, in the case of the Indus inscriptions, just the discovery of multi-lingual inscriptions is unlikely to be sufficient in enabling their reading. What is often missed in accounts of the Rosetta stone is that even after its discovery, it took more than two decades of work before hieroglyphics could be deciphered. This is because it is crucial to have a correct idea of the language of the script (which Champollion correctly surmised must be related to Coptic, the liturgical language used in Egyptian churches) that we are trying to decipher. To demonstrate this, we show in Figure 1, a bilingual inscription written in English and Chinese (Mandarin). In the absence of any knowledge of the Chinese language, the reader would be hard put to link the graphemes of the two inscriptions and thus, will be unable to “decipher” Chinese. As Michael Coe points out, while giving the specific example of Etruscan, whose writing system is known while the language is unknown, “No script has ever been broken, that is, actually translated, unless the language itself is known and understood” [Coe 2012]. The problem is compounded when the nature of the writing systems used for writing the different scripts in a multi-lingual inscription are different. For example, of the two scripts shown in Figure 1, English is a pure alphabet while Chinese is an ideographic system, making any mapping between segments of the two inscriptions based on their relative length quite problematic.

In fact, even if the two scripts had both been alphabetic, we would have to correctly infer the possible language which the unknown alphabetic system has been used to write – so as to be able to assign sound values. A toy example can be shown using a well-known work of fiction, viz., *The Adventure of the Dancing Men*, written in 1903 by Arthur Conan Doyle, in which a gentleman seeks the aid of Sherlock Holmes in figuring out why (what he believes to be) a series of childish stick-figure drawings (Figure 2) is terrifying his American wife, Elsie. As Elsie refuses to explain the reason for her reaction to these figures, Sherlock Holmes has to solve the mystery by working out the meaning of the various sequences of stick figures.

Holmes is able to deduce that the figures represent a substitution cipher and is able to read the messages by assuming it to be in English and then mapping the different stick figures to their sound values through a combination of frequency analysis (e.g., the most frequently occurring stick figure is likely to represent “e”, the commonest letter in the English alphabet) and completing partially reconstructed words to express known proper names such as “Elsie”. While it may seem that Holmes was rather lucky in being able to decrypt the messages from a relatively small number of samples, one should note that in recent times, Markov Chain Monte Carlo (MCMC) method has been used to break such ciphers rapidly from small samples by efficiently searching the space of all possible solutions for the most likely plain text message.

While deciphering inscriptions rendered in an extinct writing system is quite different from trying to decrypt an encoded message, it may be instructive to note that similar frequency-based techniques were sought to be employed in actual decipherment exercises – most notably, by James Prinsep (see Figure 3) in his eventually successful bid to decipher Asokan Brahmi

inscriptions. In principle, one can conceive of a situation in which, faced with what one believes to be an alphabetic system, the underlying language can be correctly guessed and for which there is a reasonably large number of written texts so that the frequency distribution of various sounds can be robustly inferred.

In such a scenario, it is plausible that the MCMC method can map the graphemes of the unknown alphabet to their appropriate sound values. Thus, had Brahmi still been undeciphered today, an inspired guess that the inscriptions could be in Pali Prakrit could have been sufficient to be able to decipher Brahmi, provided a sufficiently large Pali corpus was already available. However, a note of caution is in order here. Brahmi, like the Indic writing systems that followed it, is an alphasyllabary rather than a pure alphabet, and thus the number of distinct graphemes, that comprise both pure consonants as well as various consonant-consonant and consonant-vowel combinations (and occasionally, consonant-consonant-vowel, as well as, consonant-consonant-consonant conjuncts) can be as many as a couple of hundreds.

The size of data sets required to robustly infer the frequency values for such a large number of signs would correspondingly need to be quite large – unless one is able, like Prinsep, to work out the marks made on a consonantal sign to indicate each of the vowel inflections. The situation would naturally be compounded in the case of a logo-syllabic script with its many hundreds of distinct graphemes. Thus, it is crucial that any attempt at deciphering Indus inscriptions first correctly infer the underlying language, in order to be able to make any progress in reading it.

As should be evident by now, we strongly believe that a brute force approach to using computational tools to decipher scripts is bound to fail without an appreciation of the possible varieties of writing systems – that have both phonographic and logographic aspects (see Figure 4) – and their corresponding features.

This is particularly the case when occasionally claims are made about the Indus inscriptions not being writing at all. While non-linguistic sequences certainly do occur, some of them (such as musical scores) can only arise in cultures that are deeply steeped in the tradition of writing so that the concept of representing speech visually can be abstracted to the point where even sounds without an articulated meaning (such as in music) can be symbolically represented on paper. Furthermore, while it is probably impossible to prove that a set of inscriptions is writing without actually deciphering them, if there is evidence of structural organization in such symbolic sequences – such as the occurrence of clusters of signs that are seen to occur frequently as subsequences in longer inscriptions – then its interpretation as some form of writing would seem to be the simplest hypothesis in the sense of requiring the least number of additional assumptions (the Occam's razor principle). After all, “peoples around the world on similar levels of cultural complexity have come up with extraordinarily similar institutional responses when faced with similar problems – for example, the invention of hieroglyphic writing systems as an answer to the needs of nascent political states” [Coe 2012].

Thus, the archaeological evidence of a complex society with extensive trading networks, that would have required some form of long-distance communication robust to noise and the means to reliably keep accounts and store records, suggests that the weight of evidence favours the Indus inscriptions to be some form of writing.

At the Computational Epigraphy Lab (iCEL) in the Institute of Mathematical Sciences (IMSc), Chennai, we have embarked on a program of data analytics-based approach to understanding writing systems in general, with a special focus on the Indus script. Using initially the corpus compiled by Iravatham Mahadevan, and subsequently the corpus curated by Bryan Wells (the early version of which was in fact compiled at IMSc when Dr. Wells was a postdoctoral fellow there), we have compared the features of the Indus writing system with other systems around the world, both ancient and modern. One of the most promising techniques developed is the unsupervised segmentation of linguistic sequences based on sign co-occurrence statistics. In early work [Sinha 2011] we had used sign pair statistics, identifying signs that occur together much more often than would be expected had they occurred next to each other only by chance (e.g., given their individual occurrence probability in the corpus). This allowed us to merge sign clusters into larger and larger meta-symbols representing multiple co-occurring signs. Once it reduces the entire string to a single meta-symbol, - we invert the sequence to obtain the segmentation tree for the sequence, allowing us to see if the sequence (let's call it the "sentence") is built out of several sub-sequences (in this context, "words"). In subsequent work [Ashraf 2019], we have extended the analysis to include correlations that go beyond just a pair of signs.

By considering the most frequently occurring statistically significant sign clusters, and arranging them so as to align the sequences with the cluster functioning as the anchor (see Figure 5), we see that many of the Indus inscriptions are composed of relatively independent segments. By lining up these segments below one another, we can identify related graphemes that are linked to each other in terms of occurring in the same or similar contexts. An example is shown in Figure 5, where we see that two graphemes (lozenge in an oval and lozenge in a diamond) that are considered to be identical in the Mahadevan sign list but that are considered to be distinct by Bryan Wells, are most likely just allographs. By coupling such results with other features of the artefacts, such as their place of discovery, the material on which they are inscribed, etc. it may be possible to identify certain semantic correlates of the sign clusters.

We are currently in the process of constructing a database of 3-dimensional representations of the Indus artefacts in which these inscriptions appear, to provide material context that was so far absent from the data. We hope that the occasion of the centenary of the announcement of the discovery of the Indus Valley Civilization will enable stakeholders across India and around the world, who are working on different pieces of the Indus puzzle to come together in creating a digital data platform that will make the Indus inscriptions freely available online – along with their entire associated material spectrum – providing much needed context. Only by ensuring unimpeded access to the original source material to the public at large can we ensure that the next breakthrough will occur sooner rather than later.

Acknowledgments

We would like to thank the various collaborators and scholars who have enhanced our understanding of the "Indus puzzle" over the past one and half decades, especially Bryan Wells, Raj Kumar Pan and C Subramanian (Subbu) with whom we began our computational exploration of the Indus Civilization inscriptions, as well as, Mayank, who has played the vital

role of connecting the few serious scholars of the field in India. P. P. Divakaran (PPD) has been a constant source of encouragement and a fount of information on early numeral systems.

We are grateful to Andreas Fuls for maintaining the online database of Bryan's concordance that has been a vital resource for the community. We fondly acknowledge our interactions with the late Iravatham Mahadevan, whose infectious enthusiasm we miss greatly. The Indus Research Center at the Roja Muthiah Research Library, with G Sundar at its helm, has provided a wonderful meeting place for exchange of ideas. We would also like to thank Nandini Mitra, who has been instrumental in getting us interested in the decipherment of Brahmi by Prinsep, and Kurush Dalal for suggesting that we look at the non-linguistic system used by Mumbai Dabbawalas. We would also like to express our gratitude to the Department of Atomic Energy, Government of India, for their generous funding of the Computational Epigraphy Lab (iCEL) that has enabled us to continue work over the past several years.

References

1. Ashraf, Md Izhar. *Network Representation of Symbolic Sequences: A Computational Analysis* (PhD Thesis, Crescent Institute of Science & Technology, Chennai, 2019) Chapter 3.
2. Ashraf, Md Izhar and Sitabhra Sinha. "The 'handedness' of language: Directional symmetry breaking of sign usage in words", *PLoS ONE*, Vol 13, 17 January 2018, e0190735.
3. Coe, Michael D. *Breaking the Maya code* (3rd edition, Thames & Hudson, New York, NY, 2012).
4. Gadd, C.J. and Sidney Smith. "The new links between Indian and Babylonian civilizations", *Illustrated London News*, 4 October 1924, pp 614-616.
5. Hunter, G. R. *The Script of Harappa and Mohenjodaro and its Connection with Other Scripts* (Kegan Paul, Trench & Trübner, London, 1934).
6. Marshall, John. "First light on a long-forgotten civilization", *Illustrated London News*, 20 September 1924, pp 528-532 and 548.
7. Parpola, Asko and Jagat Pati Joshi. *Corpus of Indus Seals and Inscriptions. 1. Collections in India* (Suomalainen Tiedekatemia, Helsinki, 1987).
8. Prinsep, James. "Note on inscription no. 1 of the Allahabad column", *Journal of the Asiatic Society of Bengal*, Vol III, March 1834, pp 114-118.
9. Rao, Rajesh P. N., Nisha Yadav, Mayank N. Vahia, Hrishikesh Joglekar, R. Adhikari and Iravatham Mahadevan, "Entropic evidence for linguistic structure in the Indus Script", *Science*, Vol 324, 29 May 2009, p 1165.
10. Renfrew, Colin. "The origins of Indo-European languages". *Scientific American*, Vol 261, October 1989, pp 106-114.
11. Robinson, Andrew. *Lost Languages: The Enigma of the World's Undeciphered Scripts* (McGraw-Hill, 2002)
12. Sayce, A. H. "Remarkable discoveries in India", *Illustrated London News*, 27 September 1924, p 526.

13. Shah, Sayid G. M. and Asko Parpola. *Corpus of Indus Seals and Inscriptions. 1. Collections in Pakistan* (Suomalainen Tiedeakatemia, Helsinki, 1991)
14. Sinha, Sitabhra, Md Izhar Ashraf, Raj K. Pan and Bryan K. Wells. "Network analysis of a corpus of undeciphered Indus civilization inscriptions indicates syntactic organization", *Computer Speech & Language*, Vol 25, July 2011, pp 639-654.
15. Wells, Bryan K. *The Archaeology and Epigraphy of Indus Writing* (Archaeopress, Oxford, 2015).

序言

鉴于对人类家庭所有成员的固有尊严及其平等的和不移的权利的承认,乃是世界自由、正义与和平的基础,

鉴于对人权的无视和侮蔑已发展为野蛮暴行,这些暴行玷污了人类的良心,而一个人人享有言论和信仰自由并免于恐惧和匮乏的世界的来临,已被宣布为普通人民的最高愿望,

鉴于为使人类不致迫不得已铤而走险对暴政和压迫进行反叛,有必要使人权受法治的保护,

鉴于有必要促进各国间友好关系的发展,

鉴于各联合国国家的人民已在联合国宪章中重申他们对基本人权、人格尊严和价值以及男女平等权利的信念,并决心促成较大自由中的社会进步和生活水平的改善,

鉴于各会员国业已誓愿同联合国合作以促进对人权和基本自由的普遍尊重和遵行,

鉴于对这些权利和自由的普遍了解对于这个誓愿的充分实现具有很大的重要性,

因此现在,大会,发布这一世界人权宣言,作为所有人民和所有国家努力实现共同标准,以期每一个人和社会机构经常铭念本宣言,努力通过教诲和教育促进对权利和自由的尊重,并通过国家的和国际的渐进措施,使这些权利和自由在各会员国本身人民及在其管辖下领土的人民中得到普遍和有效的承认和遵行;

Preamble

Whereas recognition of the inherent dignity and of the equal and inalienable rights of all members of the human family is the foundation of freedom, justice and peace in the world,

Whereas disregard and contempt for human rights have resulted in barbarous acts which have outraged the conscience of mankind, and the advent of a world in which human beings shall enjoy freedom of speech and belief and freedom from fear and want has been proclaimed as the highest aspiration of the common people,

Whereas it is essential, if man is not to be compelled to have recourse, as a last resort, to rebellion against tyranny and oppression, that human rights should be protected by the rule of law,

Whereas it is essential to promote the development of friendly relations between nations,

Whereas the peoples of the United Nations have in the Charter reaffirmed their faith in fundamental human rights, in the dignity and worth of the human person and in the equal rights of men and women and have determined to promote social progress and better standards of life in larger freedom,

Whereas Member States have pledged themselves to achieve, in co-operation with the United Nations, the promotion of universal respect for and observance of human rights and fundamental freedoms,

Whereas a common understanding of these rights and freedoms is of the greatest importance for the full realization of this pledge,

Now, therefore,

The General Assembly,

Proclaims this Universal Declaration of Human Rights as a common standard of achievement for all peoples and all nations, to the end that every individual and every organ of society, keeping this Declaration constantly in mind, shall strive by teaching and education to promote respect for these rights and freedoms and by progressive measures, national and international, to secure their universal and effective recognition and observance, both among the peoples of Member States themselves and among the peoples of territories under their jurisdiction.

Fig. 1: An example of a bilingual set of inscriptions from the modern era, viz., the preamble of the Universal Declaration of Human Rights proclaimed by the United Nations General Assembly in Paris on 10 December 1948. The English text on top has been retrieved from the website of the United Nations (www.un.org) on August 8, 2025, while its Chinese (Mandarin) translation at the bottom has been retrieved on the same day from the website of the United Nations Human Rights Office of the High Commissioner (www.ohchr.org).

Criminal's (Abe Slaney staying at Elrige's Farm) first message



Criminal's second message



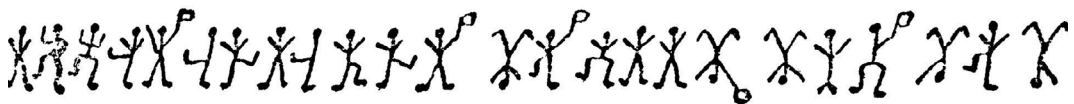
Criminal's third message



Elsie's response



Criminal's final message



Sherlock Holmes message to the criminal



Fig. 2: The messages written using a substitution cipher in the Sherlock Holmes short story "The Adventure of the Dancing Men" (1903) by Arthur Conan Doyle. Holmes decrypts the message by guessing that the underlying language is English and then using knowledge of relative frequencies of letters in English texts, along with the intuition that some of the "words" spell out proper nouns corresponding to certain names of people and places associated with the events.

Classification of the ancient character, No. I.

Reference	Simple letter	united with the vowel marks.					with η	other forms occurring
		\acute{a} ?	\acute{e} ?	\acute{i} ?	\acute{o} ?	\acute{u} ?		
1	४ ³²	४ ³	४ ¹¹	४ ³	४ ²	४ ¹	४ ⁰	४ ४ ४ ४
2	८ ³⁰	८ ¹²	८ ¹²	८ ¹⁵	८ ¹	८ ²	८	} ८ ८ ८ ८ ८ written by Lt. Burt.
3	८ ²	८ ²	८ ¹²	८ ⁵	८ ¹	८ ²	८	
4	८ ³⁰	८ ¹¹	८ ⁴	८ ⁶	८ ⁴	८ ¹	८	८ ८ ८ ८ (n ^o 2 with r?)
5	४ ²⁶	४ ¹²	४ ¹²	४ ¹⁰	४ ²	४ ¹	४	४ ४ ४
6	४ ¹	४ ¹	४ ²	४ ²	४ [?]	४ ¹		
7	८ ²⁷	८ ⁶	८ ¹⁹	८ ³	८ [?]	८ ¹	८	८ ८ ८
8	८ ²⁹	८ ⁸	८ ⁸	८ ¹⁷	८ [?]	८ ⁴	८	८ ८
9	८ ¹⁷	८ ⁸	८ ⁷	८ ²²	८ ³	८ ²	८	८ ८ ८ ८
10	८ ²⁰	८ ⁶	८ ¹⁰	८ ⁴	८ [?]	८ ³	८	(८ in Burt)
11	८ ²¹	८ ⁷	८ ⁴	८ ⁴		८ ¹	८	८ ८ ८ ८ ८ ८ ८
12	८ ²⁴	८ ¹	८ ¹	८ ¹	८ [?]	८		
13	८ ¹⁷	८ ²	८ ¹	८ ²		८ ²	८	८ ८
14	८ ²⁰	८				८ ⁹	८	८? ८ (the rest effaced?)
15	८ ¹¹	८ ⁶	८ ²	८ ⁷	८ ²	८ ²	८	८ ८ ८
16	८ ⁴	८ ²	८ ¹	८ ²	८ ¹	८ ¹		८ ८ ८
17	८ ⁹	८ ⁴		८ ⁷	८ ²	८ ¹	८	८ ८
18	८ ⁵	८ ²	८ ⁴				८	
19	८ ⁶	८ ¹	८ [?]	८	८			८ ८
20	८ ⁵	८ ⁶	८					८ ८ ८
21	८ ³			८	८			Same as 2?
22	८ ²	८		८	८	८		doubtful ८
23	८ [?]	८		८	८	८		८ ८ ८ ८ ८ ८
24	८	८	८	८	८			८
25	८ ⁴	८ ³		८ ¹	८ ¹			half six
26	८	८		८ ¹	८	८		ditto
27	८ ²			८		८ [?]		ditto
28	८ ¹	८ ³	८	८ ²			८	८ do
29	•	:	:	:				
30	Compound of ४ + ८ [?]	८	८	८				

The Initial word of 2 inscriptions on this & 5 on the Delhi Column. As. Res. VII. 180.

८ ८ ८ ८ ८

The figures against each sign shew its frequency of occurrence on the Stone del.

J.B. Tassin lith.

Fig. 3: Prinsep's frequency chart of graphemes occurring in the Allahabad Pillar inscription.

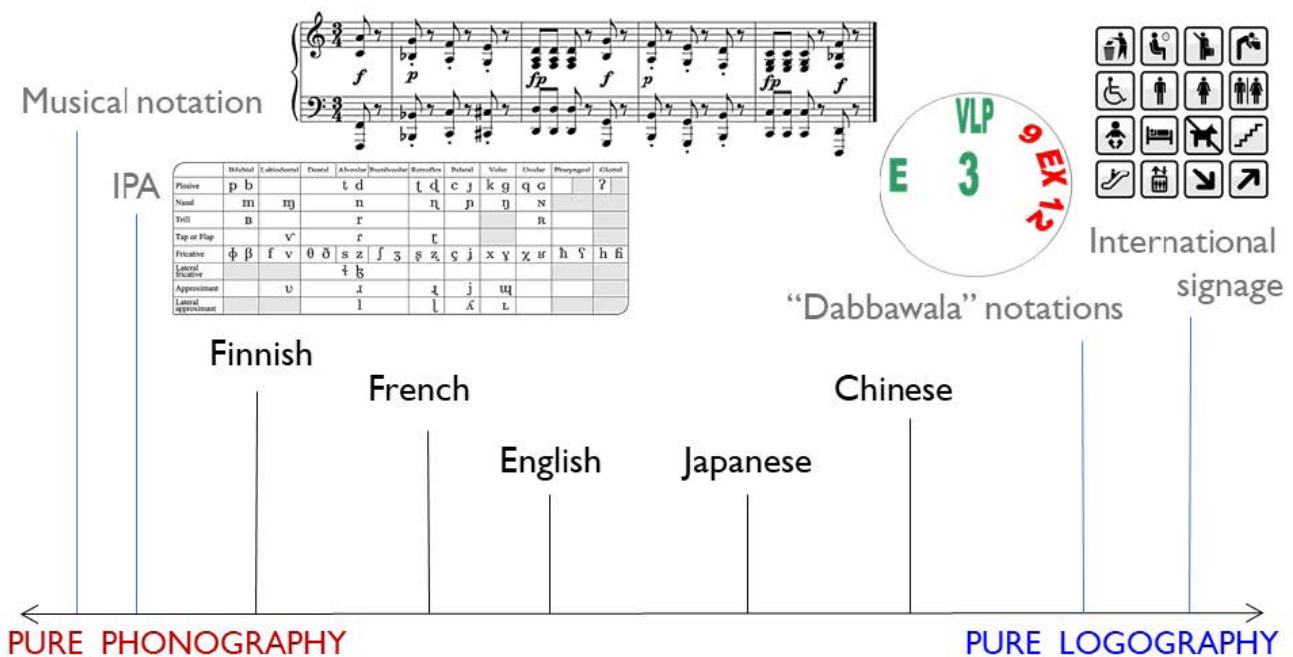


Fig. 4: All writing systems lie between the extremes of pure phonography (left) and pure logography (right) [figure modified from Robinson 2002]. Although Finnish, French and English all use pure alphabets, their differing rules of orthography makes it possible to reproduce the sounds of spoken Finnish relatively unambiguously from its written form compared to French, and even more so than English. Japanese is written using a mixed system that uses syllabic (kana) and alphabetic (romanji) elements in addition to logograms (Kanji characters that are borrowed from Chinese), and are thus more phonographic in nature compared to Chinese. Examples of specialized writing such as the International Phonetic Alphabet (IPA) and non-linguistic systems such as musical notation or international signage, are also shown. The case of the signage used by Dabbawalas (transporters of food from home to office – and back – in the city of Mumbai) can be considered to be writing used as a mnemonic device in that it uses letters and numerals (and also different colors) to indicate the end locations and way stations.

M 1480		𐎠 𐎢)	𐎠 𐎡 𐎢 𐎣	
M 525		𐎠 𐎢 ')	𐎠 𐎡 𐎢 𐎣	
M 857			𐎠 𐎡 𐎢 𐎣	
H 61			𐎠 𐎡 𐎢 𐎣 ' 𐎤 𐎥	
K 10			𐎠 𐎡 𐎢 𐎣 " 𐎦 𐎧 𐎨 𐎩 𐎪 𐎫	
M 816			𐎠 𐎡 𐎢 𐎣 " 𐎬	
M 1206			𐎠 𐎡 𐎢 𐎣 " 𐎭	
			𐎠 𐎡 𐎢 𐎣 𐎮 𐎯 " 𐎰 𐎱	
M 38			𐎠 𐎡 𐎢 𐎣 𐎮 𐎯 𐎰 𐎱 𐎲 " 𐎳 𐎴	
M 1548			𐎠 𐎡 𐎢 𐎣 𐎮	
M 1156			𐎠 𐎡 𐎢 𐎣 𐎮	
M 369			𐎠 𐎡 𐎢 𐎣 𐎮 𐎯 𐎰 𐎱 𐎲 𐎳 𐎴 𐎵 𐎶 𐎷	
M 377			𐎠 𐎡 𐎢 𐎣 𐎮 𐎯	
M 671			𐎠 𐎡 𐎢 𐎣 𐎮 𐎯 𐎰 𐎱 𐎲 𐎳 𐎴	
			𐎠 𐎡 𐎢 𐎣 𐎮 𐎯 𐎰	
H 12			𐎠 𐎡 𐎢 𐎣 𐎮 𐎯 𐎰 𐎱 𐎲 " 𐎳	

Fig. 5: A tabular arrangement for aligning sequences having in common a four-sign cluster of graphemes (“anchor sequence”, shown enclosed in a blue cartouche) that co-occur more frequently than is possible by chance, and hence must be an unit conveying meaning. This is supported by the fact that the cluster can occur independently as a sequence (3rd row). The other graphemes in the sequences can also occur together much more often than randomly. Such statistically significant sign clusters are enclosed in the red cartouches. Note that two signs (viz., lozenge inside an oval in the 6th row, and lozenge inside a diamond in the 7th row) occur in identical positions relative to the anchor sequence. Moreover, their close resemblance to each other, suggests they are allographs, i.e., stylistic variants of the same grapheme. Sequence photographs (cropped from the corresponding seal images), identified by their CISI id numbers, are from Parpola 1987 and Shah 1991. The three inscriptions not having CISI id are from Wells 2015.

