Free Paper

Determining Statistical Signatures for Undeciphered Scripts and Corpora:

the case of the Indus Valley Signs

Arnaud Fournet

Abstract: The paper first makes an assessment of a number of arguments about the linguistic or purely semiotic status of the Indus Valley Signs. According to the author it is hard to believe that a well organized society with precise weights could not have some kind of writing system at the same time. It is shown that the basic units of linguistic description: words, phonemes and syllables, have different statistical signatures. As a consequence, if a writing system is embedded in the Indus Signs corpus then it should be a syllabary with a high number of syllables (several hundreds as in Dravidian syllabaries or the Yi syllabary).

Keywords: Indus Signs, Decipherment, Syllabary, Phoneme.

1. *Introduction*

Among the major civilizations of the Antiquity, the Indus Valley is conspicuous for remaining the last one with a potential writing system which has so far resisted attempts by decipherers to determine how this set of symbols works and what it stands for, thereby confirming that it is not potentially but indeed actually a writing system.

One of the most radical approaches to the Indus Signs issue is advocated by Farmer-et-al (2004): according to the three co-authors they would not amount to a writing system at all in the first place. At the other end of the spectrum a more constructive approach is exemplified by recent attempts to apply statistical methods to the corpus of Indus Signs, especially internal entropy analyses. These papers tend to claim that the statistical features of the corpus bear similarities with proven writing systems and the Indus Signs should therefore be considered a writing system as well in spite of there existing no decipherment that has gained any acceptance beyond its enthusiastic creator(s).

The paper examines a number of arguments put forward in the ongoing debate about the linguistic or semiotic nature of the Indus Signs, the use of statistical methods and advances new insight, showing that the Indus Signs have the same statistical signature as syllables.

2. *Assessing the purely semiotic hypothesis*

This is the quite extreme and somehow provocative point of view proposed by Farmer, Sproat and Witzel in 2004: the Indus Signs are acknowledged to be symbols but they are claimed not to encode language as a writing system would do. This raises the issue of the recognizable features of a writing system when compared to other corpora that may look like writing and another issue is the validity of the arguments advanced in favor of a purely semiotic hypothesis.

2.1. *Arguments against the linguistic hypothesis*

On the whole most arguments put forward by Farmer-et-al (2004) hinge on the huge typological oddities of the Indus Signs. As will appear in the next paragraphs, they display nearly none of the features observed in proven writing systems:

2.1.1. Spread out of the area of creation

What the history of Egyptian hieroglyphs and the (Proto-Sinaitic and Phoenician) alphabet shows is that a writing system tends to spread to contiguous areas and to neighboring languages. It can also be noted that there are at least two traditional orders of the alphabet, which seem to be equally ancient: the north-western Semitic one and the south Semitic one. They obviously originate in two different schools of writing and two different directions of spread. The third order of the Arabic alphabet is more recent and represents a rearrangement based on the shape of the letters of Arabic dating.

Another example is Cuneiform, a syllabic system. The system gradually arose ca. -3000 BC, lost its pictographic features in the next centuries and spread to write a large variety of languages spoken in and around Mesopotamia: Sumerian, Akkadian, Hurrian, Hittite, etc. There are at least two writing schools with different ways of using the signs: the Anatolian school of Old Akkadian origin, with a simple ~ geminate opposition, and the Assyro-Babylonian school, with a voiced ~ voiceless opposition, plus numerous geohistorical variants.

Another typical example is the Chinese system of semantic writing. The phonetic input is only marginally taken into account by the system and the number of signs is extremely high: several thousands are in frequent use. The number requested for minimal literacy is about 2000. Many more historical variants are listed and up to 60 000 signs are known to have been used at least once, most often as the result of misspellings of regular characters. Invented around -1500 BC in central China, the system spread to whole China and then became used by neighboring countries, Japan and Korea among others. Around -200 BC, a radical standardization and schematization occurred, triggered by political unification and the increasing use of the ink-brush, which was not invented before.

Linear A was invented in Crete and spread to Cyprus. This system is also attested in Syria in Rās Šamra.

When compared with proven writing systems, we can see that the Indus symbols never spread out of the area where they are attested, even though the Indus culture lasted for centuries if not millennia. More than half (over 2100) of the strings of Indus Signs come from Mohenjo-Daro alone.

2.1.2. Attestations in foreign lands

Apart from spreading to other languages and areas, another typical feature of writing systems is that they are most often attested out of their original location on a large variety of supports:

- Etruscan is attested in Egypt on a mummy's shroud, in Croatia on a tile and in the Aegean sea in Lemnos on a rock. These attestations found out of Etruria are the longest documents in the language,

- Carian, with about one hundred inscriptions, is attested as graffiti in Egypt on walls, door-pillars, rocks, epitaphs, as far-away as the second cataract of the Nile in Sudan,

- Eteo-Chypriot, with fewer than fifteen inscriptions, is nevertheless attested at least once in Egypt as a wall graffito,

- Chypro-Minoan, with fewer than ten inscriptions, is attested in Cyprus and Syria.

So it is quite incredible that the Indus Signs, with thousands of attestations, are not attested more frequently on walls, door-pillars, rocks, stelas or other material supports in Mesopotamia or whichever foreign land where Indus people are known to have sojourned and be present.

2.1.3. Increasing and divergent schematization

Another typical feature of writing systems is the increased schematization with time. There is nothing left in the letters A and B that they used to be drawings of a cattle-head and a house. In addition to schematization, writing systems usually develop several "styles" when used for a specific language: for example, capital letters or hand-script cursive letters. Hieroglyphs which never ceased to have decorative purposes nevertheless have hieratic and demotic "styles", which reflect ordinary profane use of the system rather than chronological evolutions of the system. The same is true for Chinese characters, which have several ways of being written, drawn, carved or painted: Dazhuan Seal style, imitating the archaic shape, regular Kaishu styles and Caozhu style with artistic rather than writing purposes. It can be noted that the adaptation of the original Phoenician alphabet to other languages lead to divergent shapes as is now visible in the numerous alphabets around the world.

As for the Indus Signs corpus, Farmer-et-al (2004:33) underlines that it apparently did not undergo schematization and talks about a "suspended state of development for six centuries". But this argument is weaker than it seems. One of the reasons for the apparent homogeneity is that most seals and artefacts with Indus symbols were produced in very few places. In addition as we do not understand what the signs stand for, we are not in a position to tell which signs may be schematized variants of each other. This argument is in fact much more a consequence of the pending status of the corpus as an undeciphered writing system than an argument against the corpus being an actual writing system.

2.1.4. Bilinguals

It is worth noting that even the writing systems with a very limited geographical extension and a paucity of attestations have revealed bilinguals. After the discovery of the monolingual Mitanni letter, the next finds written in Hurrian were bilinguals and even quadri-lingual documents. The longest document in that language is a Hurrian-Hittite bilingual. Eteo-Chypriot in Amathous, with fewer than fifteen inscriptions, has a bilingual with Gree. Carian also has bilingual inscriptions with Egyptian in Egypt and Greek in Caria. Such a poorly known language as Hattic has bilingual texts with Hittite. Libyco-Punic has bilinguals with Latin. Irish Oghamic has bilinguals with Old Norse.

Here again, the Indus semiotic system, in spite of thousands of attestations, does not have any bilingual with any other deciphered language. This is strikingly improbable. Considering the testimony of other scantily attested writing systems, about 5 to 10% of the Indus corpus should be bilinguals. How is it possible that no Indus Valley king, high-priest or leader ever wrote to anybody else nor ever received any kind of answer, when Mesopotamia, Asia Minor, the Aegean sea and Egypt display an extensive correspondence between kings, pharaohs, princes, etc. for centuries in a large array of writing systems?

In such a large area where the Indus civilization existed, several languages are expected to have existed. With the same size, Mesopotamia had Sumerian, Akkadian, Assyro-Babylonian, Elamite, Kassite, Hurrian, Urartian, Aramean, Luwian, etc. representing several linguistic families. Does this mean that the Indus Valley was linguistically more homogeneous than the Near East and remained isolated? Parpola (2005:30) suggests that "Terracotta models of bullock carts attest to improved transport in the Indus Valley. This led to considerable cultural uniformity over a wide area." This may be an explanation.

2.1.5. Variety of material supports

As rightly criticized by Farmer-et-al (2004) a kind of remedial hypothesis to explain the absence of long strings of Indus symbols is that the long strings were on perishable supports, like cotton or palm leaves, and got lost because the climate is too harsh. But it can be noted that one of

most ordinary media of writing for thousands of years was clay tablets and there exist thousands of them, which got baked in accidental fires or arson, in Mesopotamia and in Crete. It is amazing that not a single potentially existing clay tablet has been found in the thousands of Indus villages, built with clay bricks, during the centuries of existence of this society. One of the most typical features of writing systems is that when people have a writing system they tend to write on the largest specter of material supports possible. How come there is no rock, no stella with long Indus symbols on it when Libyco-Punic displays such a large variety of supports in spite of there being so few Libyco-Punic inscriptions. How come no such artefacts can be found for the Indus semiotic corpus, when they can be found for Oghamic, Runic, Libyco-Punic, Egyptian, Chinese, etc., that is to say nearly all writing systems?

2.1.6. Monumental writing

Several civilizations have carved inscriptions on mountains in gigantic size and here again the Indus Valley civilization is conspicuous for providing no instance of that kind. This can be compared with Abu Simbel (Egypt), the Inscription of Behistoun (Old Persian) or numerous Chinese inscriptions on mountain cliffs as for example in Taishan (Shandong).

2.1.7. The representation of writing

Another feature of civilizations and people who know writing is that there are explicit signals that they can write in the first place: writing devices are depicted, for example the Hieroglyph Y3 representing a writing set, some artefacts are unmistakenly surfaces and supports designed for writing, learning how to write and they bear long strings of signs, as the Marsiliana Tablet (Etruscan Alphabet ca. -700 BC), people are represented writing. This can be compared with the absence of attested writing devices, the absence of any representation of writing in the several hundred Indus Signs and the absence of any statue of a scribe. Hieroglyphs testify to the existence of writing, of writing tools and of scribes. In spite of being fairly figurative and numbering more than four hundreds at the least, the Indus Signs do not bear testimony to any activity of that kind. It should nevertheless be noted that some finds have been interpreted as writing devices (Cf. Parpola 2005:39).

2.2 *Arguments against the purely semiotic hypothesis*

Graph1: The internal structure of Indus Valley weights

As described in the preceding paragraph the Indus Signs present huge typological oddities when compared to proven writing systems but at the same time it seems exceedingly strange that an urban

civilization that seems to be very well organized such as the Indus Valley would have nothing similar to a writing system or an embryonic substitute of a full-fledged writing system. In addition contrary to a claim oft-repeated by Farmer, it takes only about half an hour to get the following graph[1] which shows that the Indus Valley civilization did have a standardized weight system:

It is strikingly obvious that an overwhelming number of the weights represent multiples or fractions of a particular unit: ¼, ½, 1, 2, 10, 100, etc. It is impossible to believe that this internal structure is a chance coincidence. This is all the more impossible to believe as ¼ of the theoretical unit embedded in the weight chart of the Indus Valley coincides with 1 karsha[2] of India's present-day traditional weight system with a 2% error margin. Cf. Parpola (2005:30) for similar observations.

In other words if we dismiss the corpus of Indus Signs as being a writing system we are then faced with the paradox of a well organized society with planned urbanization and precise weights but no writing system at the same time. Is that really possible?

3. *Determining statistical signatures with Entropy analysis or cumulative contribution*

As the Indus Signs have so far resisted decipherment a new approach based on statistics has emerged in recent years. It is aimed at showing that the internal patterning of the Indus Signs is similar to that of proven writing systems and therefore at substantiating the hypothesis that this corpus is indeed a writing system and not a random set of drawings for example. Rao's recent papers (2009, 2010) typically exemplify this approach. These attempts are not really aimed at deciphering the writing system for good (if it is indeed one) but rather at refuting the radical theory that the Indus Signs are not at all a linguistic script. A similar approach can be found in Lee-et-al (2010) about Pictish Glyphs. The idea in these papers is to show that the corpora have an entropic profile which is reminiscent of writing systems and can be distinguished from those of a number of corpora of diverse origins (like DNA, music, kudurrus or Fortran). Incidentally Sproat suggests to add Boy Scout merit badges and highway signs to the set of corpora.

I have no opinion whether these analyses based on (conditional) entropy may or may deliver conclusive insight on the Indus Signs (or Pictish Glyphs). But generally speaking it must be noted that the characterization of writing systems retained in these papers is often flawed from the start. What is more the rest of the present paper will show that more direct approaches than entropy are possible on undeciphered scripts or any kind of semiotic corpora of that kind.

One interesting and apparently conclusive argument advanced by Farmer-et-al (2004) is Figure2 which represents the relationship between the number of signs ordered by decreasing frequency (horizontal axis) and the cumulative contribution of the signs (vertical axis). Some comments can be made on Figure2:

  - The seven represented corpora include: the Indus Signs (according to Mahadevan and Wells), Egyptian hieroglyphs, Sumerian, Scottish heraldic blazons, Chinese (according to headlines or texts). The choice of Scottish heraldic blazons is made to "show" that a purely semiotic corpus can have the same apparent profile as proven writing systems, but in my opinion this corpus is possibly neither representative nor relevant. In all cases semiotic corpora with a naturally linear disposition of the elements would be better.

  - It can also be noted that Sumerian is an obscure language, with a shaky decipherment and no known genetic relationship against which the decipherment could be checked, and that its testimony is therefore at best controversial. In addition Sumerian is a mixture of phonetic and ideographic signs, and so is Egyptian with its biliteral, triliteral signs, its redundancies and

---

[1] Starting from the data in: http://www.safarmer.com/Indo-Eurasian/weightchart.jpg, divide all values by 13,67g then select a logarithmic vertical scale.

[2] Cf. Sanskrit कर्ष.

numerous semantic determinatives. Both systems have a considerable number of signs when compared to alphabets. They are certainly not the most representative writing systems of phonetic nature, even though they may have been chosen because they are roughly contemporary with the Indus Signs. In all cases Egyptian and Sumerian are not "pure" enough to be used as clear references.

- The case of Chinese is also debatable: the footnote indicates that the number of signs for the headlines and texts is respectively: 1553 and 1312. This is incredibly few when compared to the bracket of 2000-4000 signs that can be expected from an average competent adult reader of Chinese. This cannot be considered representative of Chinese.

- As it is Figure2 would tend to "prove" that Chinese does not have the same profile as that of the other corpora, which the Indus Signs happen to share, hereby "proving" that writing systems have different profiles and suggesting that semiotic corpora can happen to have the same profile as some of the proven writing systems. But the major problem about the "proof" is the nature of the axes retained: absolute values for horizontal axis and percentage for vertical axis. In fact the position of Chinese below the other corpora is a consequence of the higher number of signs for Chinese and has nothing to do with some intrinsic features of morphemic writing systems when compared to phonetic writing systems.

In other words the corpora are far from being satisfactory and representative. What is more, Figure2 tells nothing about intrinsic features of the corpora and is an artefact directly based on the deceptive choice of absolute value for X-axis but percentage for Y-axis. Two other arguments put forward is that signs are hardly ever repeated in the same string of Indus Signs, a falsified claim on account of the Dholavira 17-sign string, and that strings are supposedly too short to encode language, a claim that is ultimately irrelevant, as we do not know what these strings actually meant or were designed for by their creators.

4. *The statistical signature of the linguistic items: morphemes, phonemes, syllables*

Assuming that the Indus Signs may contain a writing system, the issue is then to determine what kind of writing system it might be. This requires to define objective criteria against which the corpus of Indus Signs could be compared. As will appear below the different linguistic items: morphemes, phonemes, syllables do not have the same statistical signature. Although our method seems fairly simple it would appear that it has never been proposed or used before.

4.1. *A linguistic typology of writing systems*

On the whole, writing systems can be divided into two major classes:

a) phonetic systems relying on what languages sound: a1) phoneme-based systems and a2) syllable-based systems
b) semantic or morphemic systems relying on what languages mean.

In addition writing systems can be defective, redundant or approximative, which may render a decipherment yet more difficult. Typical examples of sound-based systems are the Egyptian hieroglyphs and the alphabets. The alphabetic system is characterized by a low number of signs: most often fewer than 30. Apart from the Korean alphabet, which is an independent invention imitating the shape of Chinese characters but not the principle, all alphabets ultimately derive from the adaptation and rationalization by North-Western Semitic people of the complex Egyptian hieroglyphs. The original Egyption system can write individual sounds and also two or three sounds with one sign. It could have evolved into a true alphabet, which it actually contains, but Egyptian people apparently never tried nor desired to discard phonetic redundancy and semantic determinatives. There were about 750 signs, 160 of which are truly phonetic: 30 uniliters, 80 biliters and 50 triliters. It can be noted that

the high number of signs of the Indus Signs corpus is not intrinsically incoherent with being a writing system as Egyptian hieroglyphs are also quite (uselessly) numerous.

It can be noted that the so-called "rebus" principle is an inadequate way of stating that a system is phonetic not semantic. And the "logo-syllabic" principle supposedly underlying the Indus Signs corpus is a nearly oxymoronic construction which does not describe the corpus in an acceptable way. Writing systems are created on principles that are either semantic or phonetic. The Egyptian hieroglyphs are not "logo-syllabic": this system is a redundantly complex phonetic system, which got later on rationalized as a consonantic alphabet. It writes neither syllables nor words, but sounds, especially consonants in the first place. Neither is Cuneiform "logo-syllabic": it writes syllables of the Cv, vC or CvC types, in addition to using a certain number of signs which have been kept as ideograms. But Hurrian for example discarded most of the ideograms and is written in a near perfect unambiguous syllabic way.

4.2. *The statistical signature of morphemes and "words"*

The first test for the Indus Signs is to compare their statistical signature with that of semantic units: words or morphemes. During the 1960ies a thorough statistical survey of French texts, speeches and conversations was carried out for pedagogical purposes. The results are described in 'L'élaboration du français fondamental' [The elaboration of Basic French]. In this book EFF (1964:69-113), words are listed according to their decreasing frequency range. The first four words are: 'être' [to be], 'avoir' [to have], 'de' [of], 'je' [I]. These four words account for 14% of all words. This can be compared with four Indus signs accounting for 21% of all signs. The first twenty words account for 38% of all words. This can be compared with the twenty most frequent Indus signs accounting for 50% of all sign occurences. If the Indus Signs were a semantic writing system then they would be nearly meaningless strings of grammatical morphemes: articles, pronouns, deictics, auxiliaries, conjunctions. The conclusion is that the Indus Signs cannot be a semantic corpus. This is confirmed by their cumulative frequency profile when compared to French words (as exemplified in EFF) or English words (as exemplified in Geoffrey Leech's works):
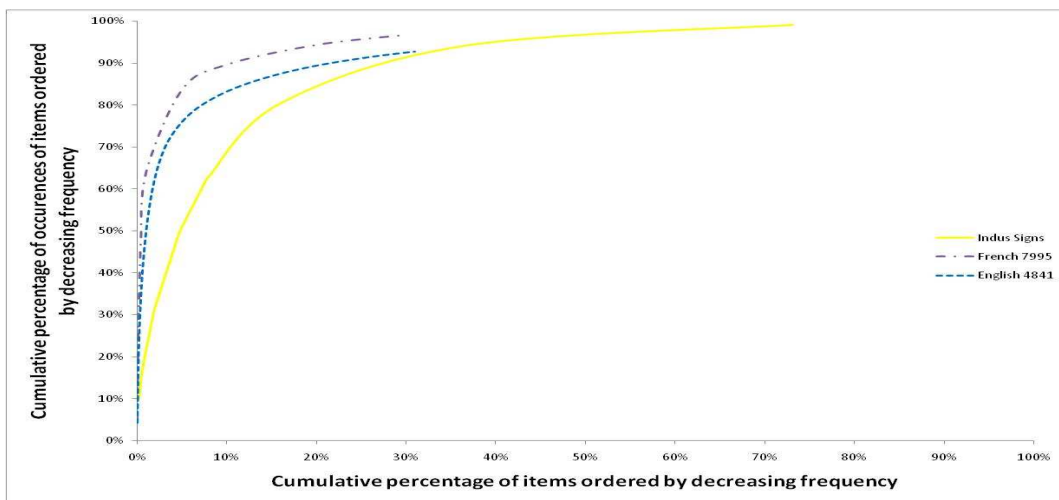
Graph2: The cumulative frequency of French and English words

As shown by Graph2 a typical feature of words is that a very limited number of words account for a considerable percentage of all occurences, leading to a near vertical start followed by a smooth increase. The profile of the Indus Signs does not fit that pattern at all. Only the first 30% are represented in Graph2 as the rest of the data is irrelevant.

Now, we may also hypothesize that the Indus Signs may be a semantic writing system with omission of grammatical morphemes, as is fairly frequent in embryonic stages and exemplified by early cuneiform writing. As shown by Graph2 the Indus Signs do not fit that profile either:
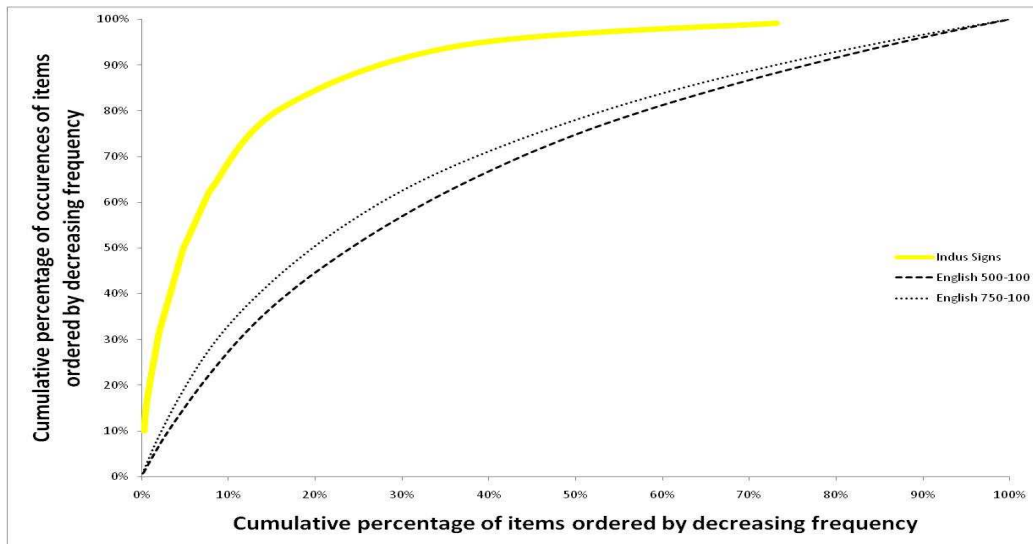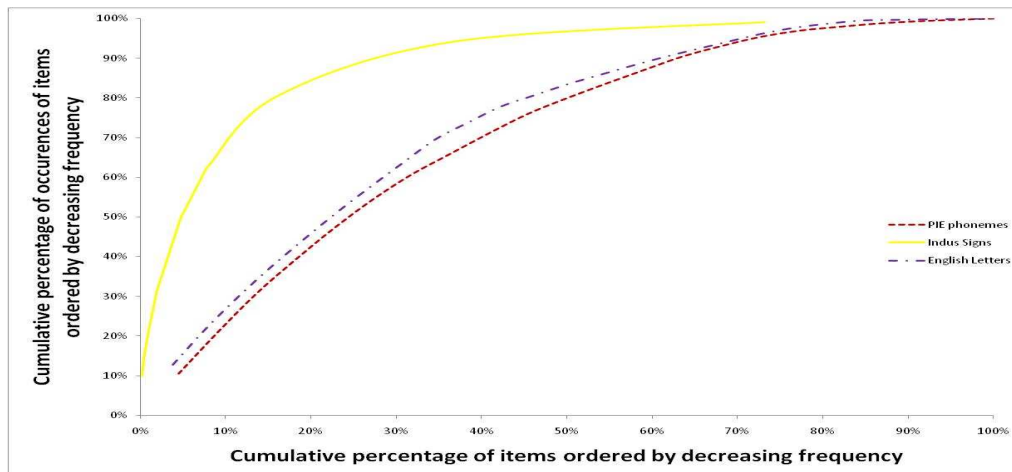
Graph3: The cumulative frequency of the most frequent truly lexical morphemes of English

Graph3 compares the cumulative frequency of the 400 and 650 most frequent English words, the 100 most frequent words of mostly grammatical or syntactical nature being removed from the calculation. As a general conclusion we can assert that the Indus Signs are almost certainly not a semantic writing system, be it exhaustive or defective. They do not encode morphemes or words. This analysis also tends to falsify the idea that they can be a "symbolic" system of any kind as we would expect meaningful "symbols" to present a profile similar to that of the truly lexical and non grammatical units of a natural language like English.

4.3. *The statistical signature of phonemes and letters*

The next test for the Indus Signs is to compare their statistical signature with that of individual sounds: phonemes or letters:
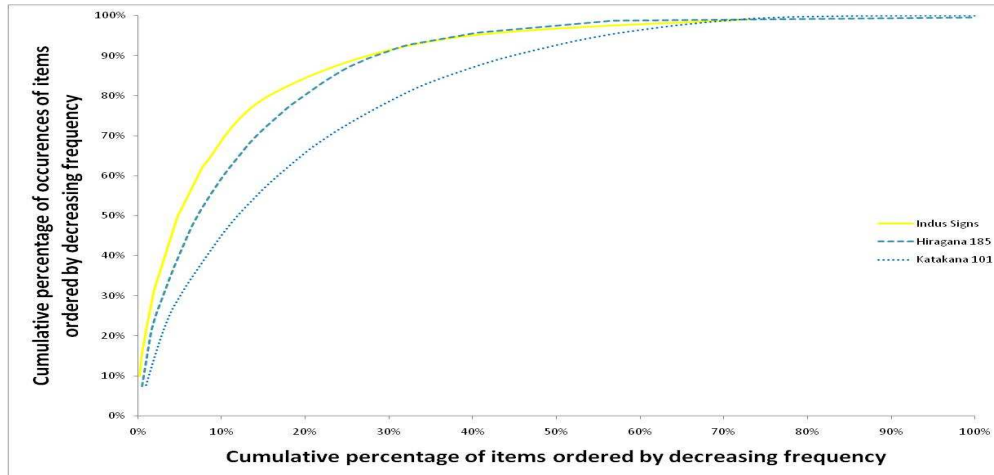

Graph4: The cumulative frequency of letters and phonemes

Graph4 compares the Indus Signs with the frequency of letters in English and of the Proto-Indo-European phonemes as represented in PIE roots (my own calculations based on Pokorny's IEW). As can be expected from the huge number of Indus Signs (above 400 different items at least), this corpus does not have the profile of individual sounds.

This conclusion is obviously unsurprising.

4.4. *The statistical signature of syllabaries*

There remains a possibility that the Indus Signs would be a phonetic writing system relying on a syllabary. Japanese provides a set of syllabaries: Hiragana and Katakana with respectively 185 and 101 different signs. As shown by Graph4 (Data from Tamaoka-Makioka 2004) the Indus Signs definitely match the profile of the Japanese syllabary with the highest number of signs:

Graph5: The cumulative frequency of Japanese syllabaries

Even though Graph5 does not explicitly "prove" that a syllabary is embedded in the Indus Signs it leaves open the possibility that they may be a writing system of that particular kind. The frequency profile of the Indus Signs is neither characterized by the near vertical start of semantic units nor the somehow linear distribution of individual phonetic units. The profile has the smoothly increasing shape of a syllabary, steeper than that of phonemes or letters but not as vertical as that of words or morphemes.

Unfortunately it seems that there exists no available survey of India's syllabaries. They would provide feedback on syllabaries with high number of signs used for languages with areal if not genetic relationships with the language(s) once spoken in the Indus Valley.

It can be noted that the number of different syllables would be well above 185. This number can be compared with the syllabary of the Yi language (彝语 spoken in Southern China), comprising several hundred independent signs which cannot be reanalyzed alphabetically as is the case for the alpha-syllabic systems of present-day India. A syllabary with a high number of signs is therefore possible and the high number of Indus Signs is not an intrinsic obstacle to the hypothesis that they represent a writing system of that kind. Farmer-et-al (2004:29) states that "the number of Indus Signs is much larger than those expected in 'pure' syllabaries". This claim is egregiously false and is possibly based on the historical instance of the Greek syllabaries, which were a very approximative rendition of Greek. The present-day Yi syllabary probably has a higher number of signs than the Indus corpus used to have and falsifies Farmer-et-al's claim. The reason why the number of signs in the Yi syllabary is so high is because it provides signs for syllables of the type *CvC*, in addition to *v* and *Cv*. One may hypothesize that the Indus Signs only wrote *v* and *Cv* syllables as their number seems to be lower.

Because the number of Indus Signs is very high several scholars have concluded that the underlying script was probably a semantic-based writing as in Parpola (1994) or Possehl (1996). But we have seen above that this hypothesis does not match the profile of the Indus Signs. For that matter the hypothesis advanced by Farmer(-et-al) that the Indus Signs have religious or mythological "meanings" is also most probably false.

5. *Conclusions and methodological perspectives*

The paper shows that the different linguistic units: words, phonemes or syllables, have different and recognizable statistical signatures. Words or morphemes are characterized by a near vertical signature because a very small number of them accounts for most occurences. At the opposite point of the specter phonemes or letters, apart from being in a finite and limited number, are characterized by a near linear signature. In between one can identify syllabaries: their statistical signature is significantly steeper than that of phonemes and letters but does not reach the near vertical feature of words and morphemes. In theory it is therefore possible to determine the statistical signature of any kind of semiotic corpora, especially of those suspected to be of linguistic nature.

As preliminarily discussed in the paper there is little doubt that the Indus Signs display a very strange collection of typological oddities when compared to proven writing systems. But at the same time the cumulative frequency profile of this corpus nicely matches that of syllabaries and on the contrary does not match that of phonemic writing nor that of morphemes, words or "symbols". I hope that the present paper will help researchers and scholars to focus on the probable hypothesis that the Indus Signs are a syllabary with a high number of signs and that statisticians will be able to determine adequate methods to help crack this "mysterious" writing system. The huge number of signs (much above 200 and possibly up to 400) is not a major obstacle as the syllabary of the Yi language is an attested and living instance of such a huge corpus of graphically independent symbols used to write a language. And it would appear that the Indus Signs are another instance of such a huge corpus of graphemes used to write a language, which is for the time being unidentified, or maybe several languages.

References

Farmer, Steve, Richard Sproat, and Michael Witzel,

  2004   The collapse of the Indusscript thesis: The myth of a literate Harappan Civilization.
              Electronic Journal of Vedic Studies 11 (2): 19–57

Kenoyer, J.M.,

  1991   *Urban Process in the Indus Tradition:A Preliminary Model from Harappa.*
              http://www.harappa.com/harappa_1986_1990/Harappa1986-90_04_Kenoyer-
              UrbanProcess.pdf .

Knight, Kevin, and Kenji Yamada,

  1999   A computational approach to deciphering unknown scripts. *Proc. of ACL
              Workshop on Unsup. Learning in Natural Lang. Processing.*

Lawler, Andrew,

  2004   The Indus script: Write or wrong? *Science*, 306:2026–2029, December 17.

Lee, R., Jonathan, P. & Ziman, P.

  2010   Pictish symbols revealed as a written language through application of Shannon
              entropy. *Proc. R. Soc. A* 466. (doi:10.1098/rspa.2010.0041)

Mahadevan, Iravatham,

    1977   *The Indus script: Texts, Concordance and Tables*. (Memoirs of the Archaeological Survey of India, 77.) New Delhi: Archaeological Survey of India.

Parpola, Asko,

    1994   *Deciphering the Indus Script*. Cambridge University Press.

Parpola, Asko,

    2005   *A Study of the Indus Script*.
                http://www.harappa.com/script/indusscript.pdf

Parpola, Asko,

    2007   *Is the Indus Script indeed not a Writing system?*
                http://compling.ai.uiuc.edu/2007Workshop/Slides/parpola.ppt

Possehl, Gregory,

    1996   *The Indus Age: The Writing System*. U. of Pennsylvania Press.

Rao, Rajesh

    2010   Probabilistic analysis of an ancient undeciphered script. *IEEE Computer*, pp. 76–80, April.

Rao, Rajesh, Nisha Yadav, Mayank Vahia, Hrishikesh Joglekar, R. Adhikari, and Iravatham Mahadevan.

    2009   Entropic evidence for linguistic structure in the Indus script. *Science*, 324 (5931):1165.

Sproat, Richard,

    2010   Ancient symbols, computational linguistics, and the reviewing practices of the general science journals. *Computational Linguistics*, 36(3).

Tamaoka, Katsuo, and Shogo Makioka

    2004   Frequency of occurrence for units of phonemes, morae, and syllables appearing in lexical corpus of a Japanese newspaper. *Behavior Research Methods, Instruments, & Computers*, 36 (3), pp. 531-547.

Vidale, Massimo,

    2009   *The Collapse Melts Down*. East and West, Volume 57, pp. 333-366.

Wells, B.,

    1999   *An Introduction to Indus Writing*. 2nd edition. Independence, Missouri: Early Sites Research Society.

Witzel,Michael,

> 1999   *Early Sources for South Asian Substrate Languages*. Mother Tongue.
> http://www.people.fas.harvard.edu/%7Ewitzel/MT-Substrates.pdf

Witzel, Michael,

> 2005   *Archaeological Fantasies: How pseudo-archaeology misrepresents the past.*

Witzel, Michael, and Steve Farmer,

> 2000   Horseplay in Harappa: The Indus Valley Decipherment Hoax. *Frontline* 17(19):4-
> 11. http://www.safarmer.com/frontline/

Witzel, Michael,

> Autochthonous Aryans? The Evidence from Old Indian and Iranian Texts. *EJVS*.
> http://www.people.fas.harvard.edu/%7Ewitzel/EJVS-7-3.pdf

Yadav, Nisha, MN Vahia, Iravatham Mahadevan, and H Joglekar,

> 2007   *A Statistical Approach for Pattern Search in Indus Writing*.
> http://www.harappa.com/script/tata-writing/indus-script-paper.pdf