STRUCTURAL READING AND EVOLUTION OF THE INDUS SCRIPT VIEWED AS A COMPLEX SYSTEM. I: METROLOGICAL READING, PART A[1]

by

P. Jeganathan

Indus script refers to a system of writing employed in an urban civilization called Indus Valley Civilization, dated from around the middle of the third millennium B.C. The texts are available only through very restricted unperishable mediums, typically in the form of stamp seals, each text on the average containing a text of only FIVE signs. The writing is still unread. Many of the keys that helped to understand other ancient writings are not available, but the texts have rich structural regularities.

A step by step, logically rigorous procedure (a sort of stochastic 'grammatical' inference) is employed to study the structure and the possible contents of the texts, based on the data provided by about three thousand texts. The paper presents overwhelming interlocking evidences that the writing on the seals and related objects represents an internally consistent system, possibly used to 'price' various goods and services in terms of the amount of a common currency, possibly a grain. (This does not mean that grain measures were always consciously perceived whenever the texts were used). Related forms of such a system of metrology are probably the ones that were in use until recent past in India.

A typical text starts with a sign signifying divinity, and then a certain amount is specified through a stage by stage accumulation (multiplication) process. The construction of the amount should not be viewed as abstract or conscious mathematical construction, but analogous to the construction of the words and phrases of the spoken language, possibly evolved in the barter system of trade and business. The construction procedure can be described through the following structural and characteristics classification of the signs.

The first class of signs may be called STRAIGHT NUMERALS. These are the signs in the form of closely grouped strokes. The numeral value of such a sign is the number of strokes involved in the sign. When two straight numerals are paired, with the signs clearly distinguished, the operation between them is multiplication. The second class consists of what may be called FIRST ORDER NUMERALS. These stand respectively for the numerals four to ten, some of these having more than one sign forms. When a first order numeral is paired with a straight one, the operation is multiplication. When the combination is between two first order

[1] August, 1993 (Technical Report 220, Department of Statistics, University of Michigan, Ann Arbor). Revised, March 1994. This version, December, 1995. The revisions have benefited from the helpful suggestions of Professors Peter J. Huber (Universität Bayreuth, Germany) and Petr Sgall (Charles University, Prague). I wish to express my thanks to them.

numerals, the operation is addition. The resulting numbers take the form of a first order.

The third class consists of what may be called SECOND ORDER NUMERALS. These have concrete backgrounds in the sense that they induce multiplicative operation when they are combined with the preceding classes of numerals, mostly with straight numerals. These include signs for ten, hundred, and thousand. We have been unable to identify numeral or approximate word values of the remaining signs of this class with confidence, but they appear to have been derived from a few basic signs.

The fourth class consists of what may be called basic METRICAL NUMERALS. One of the important ways in which new metrical numerals are formed is by internally modifying (ligaturing) a few basic metrical numerals, mostly using the first order numerals, and the first order ones have this distinguishing property. In fact, this and other types of compound signs form a substantial part of the sign forms. Each one of the metrical numerals stands for both a suitable unit of a (possibly capacity) measure and for the value of a numeral, taking either one of these characters depending on the context, but we have been unable to identify the exact numeral values of these also, though we derive and identify the approximate word values of many of these with those of a system of standard metrological units employed in the later historical periods. These are combined with the preceding classes of numerals to form what may be called CONSTRUCTION UNITS. The operation involved in such a combination is multiplication. These construction units (which include metrical numerals) are put together in a syntactic order, specifying the intended amount. The operation between the construction units is multiplication in the concrete sense: a construction unit measures (counts) its adjacent unit. The construction units have the tendency to align with certain other well-chosen construction units, indicating that the amounts are constructed through familiar amounts evolved in the spoken language in the form of words and phrases, and the texts were possibly understood in terms of such familiar words and phrases. Standard amounts constructed in the preceding manner are also 'multiplied', mostly by a few of the second order numerals, typically by placing the multiplier at the end of the line and by placing a conventional sign between the multiplicand and the multiplier.

The numeral values of the signs standing for four to ten, hundred, thousand, and the approximate word values and meanings of several other signs are found through characteristics classification analysis and limited etymological (in an evolutionary conceptual sense) studies. Following the opinion of the scholars, based on archaeological and linguistic data, a *form* of Dravidian is assumed in such etymological study, but the steps are designed so that any contradiction to this assumption will be easily revealed. It turns out that these numerals were evolved through concrete counting and metrical measurements.

Though the system we derive for the texts under study is a system of metrology, the analysis of this paper strongly indicates that some of the basic signs might also

have been used for a different system of purely phonetical writing in general. A detailed and systematic analysis of this and other evolutionary aspects is done in Part II of this paper.

1. Introduction and background to the data and methodology

In this section, we first give some preliminary description of the script (Section 1.1), and then give, in Section 1.2, some indications of the previous efforts, mostly associated with attempts at phonetical readings, together with some discussions, in Section 1.3, on the uncertainty as to the contents of the texts. Some indications of the approach taken and the methodology and the nature of the data employed are given in Section 1.4.

1.1. Introduction to the Script

The Indus Script poses certain problems which were not present in other writings of the Ancient World such as those of the Sumerians and the Egyptians. First, the texts are available only through objects of a very restricted unperishable medium, typically in the form of stamp seals made of steatite, each seal on the average containing a text of only five signs or graphemes. No bilingual text is available. Also, the language, or the language family, in which the texts are written is unknown.

On the other hand, as will become clear below, the texts have rich structural regularities which make them distinct in many ways from other ancient writings. Also, it is known that at least some of the seals were used in making seal impressions on clay bullas which were used as tags attached to bales of goods, indicating their restrictive character in the sense that they were intended for repeated, open-ended use. In addition, seals and sealings have been found in all parts of the excavated sites and levels (Vats 1940: 316)), which in particular include the dwelling houses of the common people, indicating that the seals were in widespread use, and their texts were understandable, or the aid or help needed to understand them was easily available among a considerable section of the wider public.

To proceed further, we now give a rough classification and description of the objects on which the inscriptions are found.

Type 1: Stamp seals made of steatite with texts deeply engraved on them in intaglio in the reversed direction and intended for making impressions. These constitute an overwhelming majority of cases, about 60 percent. Of these, about 90 percent are of square shape, typically a little more than one square inch size, containing a text in the upper register with below, a standard iconography. The most common iconography is a bull facing an unidentified object. The remaining are of rectangular shape containing only a text.

Type 2 (about 17 percent): Sealings or tablets of terracotta or faience which are positive impressions in bas-relief made by special mould. These are not intended for making impressions and are generally found inscribed on two or three sides, occasionally with pictorial motifs. These objects appear to have been mass produced from moulds since many of them bear identical impressions.

Type 3 (about 9 percent): Miniature tablets of various shapes made of steatite, terracotta or faience with incised inscriptions, containing texts on one, two, three or four sides. These are also not intended for making impressions. According to Vats (1940), these are tiny objects in the sense of measuring in length from 0.36in to 0.7, in width 0.25in to 0.6, and in thickness 0.05in to 0.13. The inscriptions on them contain many identical duplicates. Some of the shapes of these tablets are curious in the sense that they resemble some of the pictographic signs frequently employed in all types of objects, but there is no correspondence between the shape of an object and the pictographs it contains. Some of these shapes are also shared by some of type 2 objects. In addition, type 3 are found only at Harappa at lower levels since the comparable levels at Mohenjo-daro are not yet reached or reachable due to high water-level. Objects of types 1 and 2a (at Harappa) diminish both in size and numbers as the stratum levels go downwards and their place is taken by the type 3 objects (Vats 1940).

Other types of objects are copper tablets (about 4.5 percent) with inscribed or engraved inscriptions, generally similar to Type 2, seal impressions (of type 1 objects) made on burnt clay; most of them are tags attached to bales of goods (Rao 1979), and inscribed or impressed inscriptions on pottery, bronze implements, ivory or bone rods and other miscellaneous objects.

Thus, objects of types 1 and 2 appear to have been evolved from type 3 objects, when the purpose or function of type 3 objects gradually became intensive and widespread, necessitating mass production and/or simplifying the process by the device of stamping the texts. It can be

seen that type 2 tablets (see Parpola and Shah 1991: 316-333) that have shapes common or similar to those of type 3 objects have also the texts similar in character to those of type 3, suggesting that type 2 are of intermediary between type 1 and type 3, but there is no reason to exclude the possibility that all three overlapped or are of even simultaneous use especially of types 1 and 2.

We face the Indus script on the earliest type 3 objects in a fully developed and conventionalized form with an inherent syntactic structure, though the range of the texts is limited in view of the large number of duplicates, and the texts appear to have been separated often into blocks occupying different sides of the objects, possibly in view of the tiny size of the objects of type 3. In addition, the basic syntactic structure remains the same in later texts of types 1 and 2, though the complexity increases. These facts indicate that the beginning stages of evolution of the Indus script might have occurred mostly in perishable mediums different from the objects of the preceding types, but they have not survived though many of the 'potter's marks' of the societies of the stage prior to the mature phase resemble or even occasionally are identical to the signs of the Indus script; see for example, Parpola (1986: 403-407) for a summary and review.

The earlier levels of Harappa at which type 3 objects have been found are placed at 26th century B.C on the basis of MASCA corrected C14 dates in Allchin and Allchin (1982). Indus type seals (type 1 objects) have also been found at many sites of the Near East, which are placed generally from around 24th century B.C. Some standard references on the general archaeological aspects of Indus civilization are Marshall (1937), Mackay (1939, 1943), Vats (1940), Wheeler (1968), Fairservis (1975), Rao (1979) and Lal and Gupta (ed, 1984), though some of the inferences drawn in these sources are unfortunately of speculative nature. In particular, it may be noted that some of the earlier ideas regarding the formative stages and the 'decline' of the Indus have come under drastic revisions in view of increasingly fresh archaeological evidences, see for example Jarrige (1984a, 1984b) and Jarrige and Meadow (1980) which provide archaeological evidences for the chronological development of the 'Greater Indus' from hunter-gatherer and neolithic stages of the 7th millennium B.C to the later formative stages of mature phase.

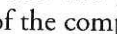
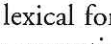
1.2. Some previous attempts at reading the script

There has been several attempts to read the texts ever since a substantial collection of them became available since 1920's; a partial list of such attempts can be found in Mahadevan and Rangarao (1986) and Parpola (1995). Among the notable achievements made so far, the publications of Koskenniemi and Parpola (1979, 1980, 1982) and Mahadevan (1977) of the corpus of available texts together with concordances deserve particular mention. That of Mahadevan (1977) in addition contains statistical tabulations of various aspects of the texts, upon which the present study heavily relies, as well as a fairly accurate reproduction of the sign forms. Earlier limited sign lists and concordances are due to Longdon (1931), Gadd and Smith (1931), Hunter (1934, Ph.D Thesis) and Vats (1940).

In addition to the preceding efforts, the following two investigations have become widely known: one by the Russian scholars Yurij V. Knorozov and his teams (1965, 1968, 1970), and the other by Finnish scholars Asko Parpola and his teams, see Parpola (1995) for an integrated synthesis. These investigations also involve some kind of structural analysis (but not in the sense of the present paper) of the texts, but with the specific *apriori* aim of classifying the signs or sign combinations into linguistic units, such as standing for root morphemes, attributes and grammatical suffixes, and then to read the texts phonetically, adapting a form of Dravidian as the underlying language. Their phonetical reading is based on the external interpretations of the pictorial form of individual signs through the principle of homophony (rebus principle) in terms of later Hindu mythologies and tradition, so as to read into the texts things such as proper names composed of divine names and/or with determinatives such as ranks, titles, etc. The same type of analysis has also been made for instance in Mahadevan (1986). Detailed reviews of these and other similar readings can be found, for instance, in Burrow (1969), Clauson and Chadwick (1969), Zide (1970), Zide and Zvelebil (1970a, 1970b, 1976) and Zvelebil (1985). While these reviewers agree in general that some form of Dravidian is a plausible, but not a certain, language of the texts, none of them agrees with phonetical readings. A typical opinion is expressed by Zvelebil (1985): '...the careful positional-statistical analysis of the texts notwithstanding, their further procedures are entirely speculative and intuitive, the conclusions possible but entirely speculative and above all absolutely unverifiable' (see also below).

What may be viewed as the only rigorous step that has been achieved so far in the above works appears to be the possible division of some texts into probable words or phrases based on stable sign combinations (see Smith 1931 and Hunter 1934), which is now fairly apparent from the concordances mentioned earlier. Another step consists of the determination of the direction of reading (in the impressions of the seals), which is widely viewed by many authors to be of a right to left direction, see Mahadevan (1977) for a summary and review. Unfortunately, the 'internal' evidence used in this respect in, for instance, Hunter (1934), Mahadevan (1977) and Parpola (1995) is based on certain deviations accidentally noticeable in a few exceptional texts, which may not be considered logically sufficient to reach a conclusion (as will be discussed in more detail later in Section 4.2.1). Indeed, unless at least a partial understanding of the texts is achieved, as was done in the case of for instance Mycenaean Linear B, one may not claim the availability of any reliable internal evidence. In the same way, the direction of reading may not be concluded from the external exceptional features such as cramping or overflow of end signs, a procedure based on the assumption that the direction of reading is the same as the direction of writing, as is done for instance in Lal (1966). The reason is that the signs, which on the average are only five in each text, are not written, rather they are carved (on the original negatives) together with other iconographic motifs in a tiny space and many of the seals actually constitute a work of fine representational art, so that the criteria of the direction of carving would also have been based in part on convenience and other factors.

As to the readings themselves of the texts of the above authors, unfortunately the primary importance is given on obtaining the precise, phonetical or otherwise, readings based on unverifiable external *apriori* suggestions or interpretations of each individual sign or occasionally text in itself, or among a few conveniently chosen sign combinations or occasionally texts in themselves, rather than attempting to isolate an internally consistent system. For the purpose of illustration, we now indicate what we think is a faithful account of the essential nature of the analysis in Parpola (1995) in this respect; the analyses of other authors are also of the same nature though they differ in individual interpretations. First, a particular sign  is identified (1995: 179) as the depiction of a fish, which has a valid basis. Second, it is observed that a certain lexical form for fish is almost identical (in Dravidian) with that for star , thereby suggesting (ibid.:182) that the fish sign is

used to signify a star by the principle of homophony. This gives phonetical readings of the names of a few stars (ibid.:194) for a few combinations, such as  , formed by the fish sign with a few of the numeral signs. Then some of the modified forms of the fish sign is argued into having also the phonetical readings of the stars. For instance, in the case of the compound sign  formed by the fish sign and an another sign that is identified as resembling a roof, it is argued (ibid.:197) that the a lexical form for the roofing is similar to that for black or dark, thereby suggesting that the compound sign in question signifies a star named 'black-star' which in turn is identified to mean the planet Saturn. Similarly (ibid.: 196), Mercury is identified with  , interpreted as 'halved fish' which in turn is interpreted as 'green star' by homophony. Similar interpretations are argued for a few other sign combinations; see Parpola (1995: 275-277) for a list of them.

The difficulty with this approach is that (assuming for the moment that the arguments involved in the derivation of homophones are valid in some sense), it is rather unclear, unfortunately, how the meanings thus derived for a few conveniently chosen sign combinations in isolation fit into a possible meaning for each of the texts as a whole and how such meanings for the individual texts in turn fit into the form of a system as a whole that is represented by the writing, since the notion of the 'meaning' itself becomes meaningful only under the framework of a system. It appears that much of the force of the arguments is derived on the *apriori* assumptions such as that most of the texts constitute noun phrases such as the names of divinities or proper names composed of divinities and other attributes, thereby arguing that the fish signs represent the names of the divinities identified with stars. In other words, much of the effort appears to force the arguments to fit into these *apriori* assumptions, that is, the nature of the system itself is assumed *apriori* instead of deriving it from the analysis of the data of the texts. More specifically, it is possible to have two systems in which the individual interpretations of each of respective entities (signs) of the systems in itself may partly overlap, yet the two systems can be entirely different, that is, the respective structures, or relationships between various entities, of the systems can be quite different. For instance, the meaning of the ordinary numbers, when they are involved in the usual decimal place value system, is entirely different from that when they are involved in the system of license or identification plates of automobiles, or in the system of phone numbers, so that the later two systems are not of much use in inferring the meanings associated with the first

system. Some of the assumptions, such as that the language represented being a form of Dravidian, would be reasonable since it is partly supported by the linguistic and archaeological data and is common to a multitude of other systems as well, but again the further analysis needs be devised so that any contradiction to such assumptions would be easily detected. The difficulty also lies in the fact a system that serves some common purpose of the society as a whole involves established conventions governing the structure of the system, so that the above associations, if true, of 'black' by 'roofing' and 'green' by 'halved' by the society must involve a complex conventionalization *process* which needs to be identified in some form.

In the present paper the writing will be viewed as an evolving complex system and a detailed study of its structure at many different levels will be made. In particular, even though it will follow from the study that, contrary to the expectations of the above authors, a precise phonetical reading is unfortunately highly improbable to obtain, the general contents as well as possible functions and evolutionary and several other important factors of the texts and the writing are still possible to obtain from such a study; see Section 1.4 below for more details.

1.3. Uncertainty as to the contents of the texts

As indicated above, an interpretation that many scholars, including the critiques such as Zvelebil (1985), have inclined to accept is that most of the texts in part constitute noun phrases such as the names of divinities or proper names composed of divinities or titles, ranks, etc. The reason for this is the assumption that the seals and related objects were used either in trade and administration, which has some valid basis, or the *apriori* speculation that they were used in something related to religious worship, and hence the assumption that they must have served as objects of some form of identification. While the possibility of such interpretations may not be *apriori* excluded in any attempt at the understanding of the texts, it is rather unclear if they can form a conclusive basis for the phonetical reading. One can give many reasons. For instance, we are in effect dealing with an archaic language and a historical period where the urban life centered on agriculture was either in the process of making or had only a short history. In particular, the language itself would only be still in the process of evolving adequate

means to express many of the thoughts associated with abstract notions, at least in the beginning phase of urbanization, as was noted by Diakonoff (1983) in a different context. Since the notion of divinity itself is an abstract one, one can doubt if there could have been so many divinities with distinct names. To give a simple analogy, it would be far from reality to claim that whatever mathematical knowledge available in the Indian sub-continent at the turn of the Christian Era would also have been available around 2500 B. C, even though some of the ideas might be in the process of evolving in the specific concrete situations in which the ideas had the opportunity, if at all, to be employed. It may also be noted that we are dealing with a period, at least the initial period, in which writing in other adjacent contemporaneous cultures, such as Sumerian and Proto-Elamite (see Nissen (1986) and Vallat (1986)), were basically still in the form of one of many technologies necessary to ease the elements generally associated with the process of urbanization, in particular to ease the process of redistributive system associated with complex economy, a complexity which might have been present in some form even prior to the phase of urbanization; for instance our modern system of currency is just one such instrument we now employ in our centralized redistributive system. One may then wonder if such a precious technology on the seals was used by the **wider public** in such an environment just for the purpose of identification.

Several scholars also draw the analogy with the cylinder seals of the ancient Near East, which are believed to have been used for personnel identification. However, it may be pointed out that the seals documented for instance in Collon (1987) appear to be different in character from that of the Indus. Specifically, the Near Eastern seals invariably contain elaborate figurative representational arts including the form of humans and human-like deities, whereas such things are a rarity in the Indus. For instance, among 2906 objects considered in Mahadevan (1977), 1993 contain some field symbols, of which 1385 are type 1 objects (seals) among which only 19 contain humans and human-like figures. In fact, the pictorial motifs in the Indus seals are quite standardized, an overwhelming majority of them being in a stereotyped form of a bull. In addition, the proportion of near eastern seals having some form of writing is rather small (and without any apparent common structure). In fact, Collon (1987: 105) mentions that '... relatively few seals were inscribed and at some periods all seals were uninscribed'. On the other hand, Indus seals with no inscriptions are rare. Furthermore, according to an analysis done by Parpola (1986,

and 1995, Fig 7.17; see also Lal 1974), a substantial number of burnt clay tags recovered in a warehouse at Lothal (Rao 1979) contain multiple seal impressions (two, three or four) and, in addition, almost 50 percent of them are interconnected in the sense of having shared seal impressions. These facts do not appear to be entirely consistent with the assumption of personnel identification.

1.4. The data and methodology of the present paper

It follows from the foregoing discussions that further analysis of the functional characteristics of the signs and the texts is essential in order to obtain a better and more reliable understanding of the character of the texts. Such an analysis was done initially for instance in the case of Linear B (see Ventris and Chadwick 1956) before any attempt at phonetical reading was made, but the texts of Linear B had their own advantages. For instance, the texts are on the average much longer than that of Indus, and it was possible to identify the numerical and metrical units, and hence to isolate the system associated with them, since they occurred in certain positions and simple interlocking operations such as adding the numerals and metrical units, converting them into higher units, recording the total, etc., were done in the same text. This step helped to the analysis and possible identification of the signs that occur regularly and in isolation before the numerals since they are likely to have individual meanings. These identifications then helped to the overall identification of the general character of the texts, in particular to isolate and to identify the character of that part of the writing which involved a phonetical system.

Unfortunately, such a procedure consisting of relatively few steps does not seem to be sufficient in the case of Indus since the functional character of most of the signs appears to be tied or interconnected to that of too many of the signs, at least through the intermediary of other signs. For this reason, analysis will be done through several steps at many different levels. Such a step by step, logically rigorous procedure appears to be essential if one desires to avoid *apriori* speculations of any kind. In fact, we are aiming at logical reasonings as close as possible to those used in exact applied sciences. In addition a basic aim of this approach will be to isolate a suitable internally consistent system. *In this sense, precise external interpretations of any individual sign in itself will be less important than the nature of its functional relationships with respect to*

other signs and its role as part of the consistent system isolated. For instance, suppose that a certain sign is classified as having the character of what will be called a second order numeral, in addition to having the numeral value hundred. Then the conclusions of this paper regarding the system as a whole will remain unaffected even if one finds reasons to doubt the validity of the specific numeral value, as long as the characteristics classification of the sign as a second order numeral remains unaffected. Further, if the pictorial form of that sign resembles, for example, that of the lay-out of a city in general whose etymology is conceptually related to, among others, that of a (possibly unspecified) numeral in a suitable concrete sense, then the possibility that the sign is used for the meaning of a city will be excluded as it will be inconsistent with the system isolated and other interlocking evidences.

It also appears that the methodology of this paper, which relies on simple data analysis, is akin for instance to what is called 'stochastic grammatical inference', but unfortunately we have been at present unable to isolate an explicitly formal description of all the important points of our data analysis into any of the existing theories. Note however that within the restricted context, many levels of descriptions of this paper of the functioning of various structures involved are as formal as possible but much of the results appear to be still concentrated in the detailed analysis and discussions. In this connection it is important to note that we do not make any *apriori* assumptions regarding the meanings or contents of the texts, or on the specific nature of the system represented in the texts, whereas the usual applications of the indicated subject heavily rely on such assumptions. In this sense it appears that the present methodology has not been employed even in a limited way, as far as we are aware of, in similar contexts with a view of the present broader aim and scope. In viewing the language as an evolving complex system with inherent interrelated structures at many different levels, we have relied mostly on our limited understanding of F. de Saussure (1959). In this sense, the formal framework described in the book by Sgall, Hajičová and Panevová (1986), as well as other publications related to this work, might be of importance in obtaining a formal description of the present results that will go beyond the specific context of the present work. (See further Jeganathan (1997)). For further details of the methodology employed, see Section 3.1.0 below, especially the Remark in that section.

Section 2 is devoted to the analysis of the construction pattern of the texts and the classification of the signs, that is, loosely speaking, to the analysis of the syntactic aspects of the structure of writing, without making any linguistic assumption.

The results of Section 2 naturally force us into a certain direction of characteristics (structural meanings) analysis of the signs, done in Section 3. We have employed a limited etymological study (in an evolutionary conceptual sense, see the remark in Section 3.1.0) in such an analysis, though not absolutely essential at this stage, in order to obtain stronger logical support for the conceptual meanings suggested by other means. This limited etymological study (basically for the numerals four to ten and hundred), is solely for the purpose of understanding the concrete meanings, in a broad conceptual sense, behind a certain restricted class of words associated with certain systems of numerology and metrology, as it will turn out from the analysis, that were employed during the historical period. This then leads into a detailed analysis and understanding of the structure. These steps are carried out in Section 3, where it is also noted that one can read the given arguments in such a way that the etymological arguments may be omitted at first, obtaining tentative results which are then supported by the given etymological arguments. The resulting structure (indicated in the Abstract above), is more completely described, with the help of a selected number of texts, in Section 4 of Part II where many other issues such as the direction of the writing, possible contents and purpose of the texts, are dealt with in some detail.

Though the system we derive for the texts under study is a system of metrology, the analysis of Section 3 strongly indicates that some of the basic signs might also have been used for a different system of purely phonetical writing in general. A detailed and systematic analysis of this and other evolutionary aspects is done in Part II of this paper.

In connection with the preceding methodology, the readers may note the several limitations, outlined in Section 3.1.0, of the etymological study of this paper in order to avoid any possible misunderstanding in this respect, especially in view of the unfortunate bad reputation the present subject has acquired due to the great many claims of 'decipherments'. It may also be mentioned at the outset that no definite conclusions are generally intended in isolation at any single stage of the analysis, and the evidences for any suggestion are in general spread out throughout the paper, including Part II. The present paper also does not have the usual aim of text by text

phonetical decipherment; in fact the present results indicate that such an aim may not be possible entirely to achieve, unfortunately.

As was indicated earlier, the structural analysis of the present paper is based in part on the concordance and the statistical data presented in Mahadevan (1977), which in addition contains the corpus of the texts on which the concordance is based, provided by 2906 inscribed objects of the types described earlier. Here, an inscribed object can have, by definition, only one text consisting of one or more lines, inscribed on one or more sides of the object. The total number of lines of texts is 3573, but the majority of the texts are of single lines only. The unit of structural analysis of the present work is in general a line of text, which is also the unit of concordance as well as the source for statistical data. Note that there are texts, especially in type 3 objects, which appear to have the character of a single line but divided into blocks which are represented on different sides, but we shall deal with such possibilities later after the character of the texts are properly understood. The sign list of the the texts presented in Mahadevan (1977) constitutes 417 signs, broadly interpreted, with a total frequency of occurrences 13372. Of these, 112 signs occur only once. Further, the number of signs having frequency at least 10 is 155, which together have total frequencies of about 95 percent. However, a great many of the signs are clearly combinations of one or more, or modifications, of the remaining signs. In addition more than 30 of the signs are what has been called '*numeral signs*', that is, signs consisting of closely grouped strokes, constituting about 11 percent of the total sign occurrences. (*We do not make the assumption that these stood for numerals*). A glance at the table given on page 17 of Mahadevan (1977) reveals that relatively few signs take most of the functional load of the texts. In Appendix B, we have presented a class of signs, the total frequency of them constituting about 91 percent, where we have grouped together all forms of the above indicated numeral signs under one category, for the reasons that will become clear later. While all the signs will be utilized to gain possible insights, the preceding class is quite sufficient for the purpose of structural and characteristics analysis and classification. In addition, they adequately represent the overall characteristics of all the signs and they together with their variants take almost all of the functional load.

In Appendix A; see also Part II, Section 4.3.0, a list of about eighty texts is given for the purpose of illustrations. There they are not accompanied by their respective identification numbers of either of the

preceding corpuses since they are easily identifiable from the concordances.

Note that the texts in Mahadevan (1977) include also identical duplicates, predominantly of objects of types 2 and 3. This raises the possibility of over-representation of certain signs in their respective frequencies, since the duplication process might be mostly only implicit in the case of type 1 objects. However, our checking shows that such signs have also high frequencies with respect to type 1 objects, so that relative down-weighing of frequencies of these signs is not essential for the present purpose.

2. Analysis of the construction pattern and classification of the signs

2.1.0

Our aim in this section is to classify the signs into several groups, based on the construction pattern of the texts, without assuming any linguistic nature, such that the signs within each group have certain common characteristics. This will lead to a description of the construction pattern and the interacting relationships between the signs. First we shall classify those signs that have certain clearly identifiable positional and functional characteristics. One of the important and crucial facts that emerges in the initial stages is that many of the signs that have linear graphic forms classify themselves together with what has been called numeral signs.

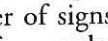
What is implicitly involved in this section is the analysis of the frequencies of the different possible sign combinations, that is, relative frequencies (transition probabilities) with which different signs follow, or alternatively precede, a given sign or sign combinations, together with the analysis of certain clearly identifiable positional and conventionalizational pattern of the construction. Explicitly, no mathematics other than computing the percentages will be involved.

In what follows the identifications (T.1), (T.2), etc. stand for the texts presented in Appendix A, below. There they are divided into appropriate units which should be ignored for the present.

2.1.1

One can choose the initial step in many ways. Perhaps a convenient way is to start with the signs and . These occur mostly at the left end of a line, with very high frequencies, see the texts (T.1) - (T.79). Even when they occur at the middle, appropriate sign sequence with one of these signs as the left end has the character of a complete text elsewhere, see (T.16), (T.50), (T.60), (T.61), (T.68). (There are also cases, with reasonable frequencies, where these signs have the meaning of occurrence at the left end but are joined together with another sign, see (T.56)). These two signs have the respective frequencies 10.75 percent and 1.7 percent. If we take into account the total number of signs, these percentages are abnormally high. In fact is the most frequently occurring sign. In addition, they occur on at least 55 percent of the objects. (In computing the above percentages we have excluded the cases indicated above where these signs are joined together with other signs). In other words, the inclusion of one of these signs as the left end of a line or a sign sequence appears to be a matter of established convention. For these reasons, these two signs are grouped together. This class will be noted by **C1**.

2.1.2

Next the conventional signs in the forms of superscript strokes  see for instance (T.11), (T.13) and (T.48), and in the forms of single and double strokes  and  of the middle register, see (T.60), are grouped together to form class **C2**. The reason is that each of these signs occurs in the middle portion of a line, with the sign combination on the left of it having the character of a complete text elsewhere and on the right side of it having either a sign sequence, often having the character of a text elsewhere, or the single occurrence of just one sign. The total frequencies of these signs is 7.6 percent, with the superscript sign alone having the frequency 4.85 percent, the second most frequent sign.

2.1.3

Among the texts having the conventional signs from the preceding class **C2**, the ones having the occurrence of a single sign on the right side of the signs from **C2** form a substantial part. Such single occurrences are dominated by a few signs. Of these few, some of them occur mostly in such positions. Even when they do not occur in such positions, they often occur in the final right most or quasi-rightmost positions of a line. These are given by

The total frequency of these is a little more than 6 percent. These are grouped together under the class noted by **C3a**. These signs also occur in the middle of a line with certain characteristics, as we shall see later, which will also be shared by a few other signs that will then be classified as **C3b**. Several texts illustrating the class **C3a** can easily be identified among (T.1) - (T.79). Here it may be noted that the sign  also occurs in the form .

2.1.4

Now let us look at the sign  . This sign occurs in the leftmost position of the lines in almost all of its occurrences, similar to those in **C1**, see for instance (T.2) and (T.3). But it cannot be classified under **C1**, since in most cases it occurs at the immediate left side of the signs in **C1**, that is, it cannot play the role of those in **C1**. (It also appears to have been frequently ligatured with other signs, taking the role of the sign as will be discussed later). It has a high frequency (355) but it occurs in objects of type 1 only 89 times, but still reasonably high, and almost all of the rest on types 2 and 3 objects, which is in part the reason for its high frequency in view of the duplicated nature of objects of types 2 and 3. It appears to intend to imply something regarding the sign sequences on the right side of it, but we shall leave this sign unclassified for the present. Another sign that shares the character of this sign is .

Another important sign in the construction of the texts is the man sign , see (T.44) - (T.55). This sign seems to have certain unique

characteristics not shared by other signs, as will become clear later. Further it occurs, either alone or in ligatured form with other signs, with high frequency, about 6.34 percent. For this reason it is isolated for individual study.

2.1.5

The preceding steps then lead to the consideration of the sign sequences, consisting of one or more signs, of the following types:

(a) Sign sequence forming a complete line when apparently no signs from C1 and C2 are involved, see (T.44), (T.54) and (T.56).

(b) When no signs from C2 occur but a sign from C1 occurs at the left end or quasi-left end, the sign sequence forming the right side of the sign from C1, see (T.1), (T.4), (T.6) and (T.9).

(c) Similarly, when a sign from C2 occurs at the middle of the text and a sign from C1 is apparently absent, the sign sequence forming the left side of the sign from C2, see (T.49) and (T.58) - (T.60).

(d) When both signs from C1 and C2 occur, the sequence occurring in between them, see for instance (T.7) and (T.8).

It would be convenient to introduce a terminology for such a sign sequence, to be called a PRINCIPAL BLOCK. Note that there are sign sequences having the character of principal blocks which do not come under the preceding classifications (a) - (d). For instance, there are texts of a single line that can be decomposed into sections, each of them having the character of a line elsewhere, see (T.50), (T.61), (T.62) and (T.68). In fact (T.50) is composed of three separate sections corresponding to the texts (T.32), (T.37) and (T.48). In such cases, the preceding classifications need to be applied to each such sections. Also, the sign sequences on the right side of signs of C2 have often the character of principal blocks, see for instance (T.46). In general, once the formation of principal blocks is understood in a vast majority of cases where the preceding procedures (a) - (d) apply, separating the principal blocks in other cases will be clear. The idea of considering such a principal block is that it has the character of forming by itself a

complete text or line, so that once the general characteristics of such principal blocks as well as those individual signs that constitute them are understood, at least structurally, then a better idea of the functional nature of the signs in C1 - C3 can also be formed.

2.2.0

If one looks at the corpus or concordance of the texts, a visible and striking nature to be noticed of the principal blocks is that what has been called 'numeral signs', that is, the signs in the form of repeated strokes, such as  ,  ,  etc, are regularly interspersed among other signs. In fact numeral signs form 11 percent of sign occurrences. Therefore a convenient way to start analyzing the signs forming the principal blocks is to start with these signs. To begin with one cannot assume that these signs actually stand for numerals, but one can attempt to see if they have any common functional characteristics to group them into a single group.

One easily visible characteristic of the numeral signs is that they are regularly aligned with many other signs to form STABLE COMBINATIONS. Here, a stable combination roughly means a combination of two, occasionally more, signs having the characteristic of a single unit in the sense that it has a frequency high enough or as high as many of the frequently occurring individual signs. For example, the combinations

are such combinations. It is difficult to give a more precise definition of a stable combination, but it will be clearly identifiable from the concordance.

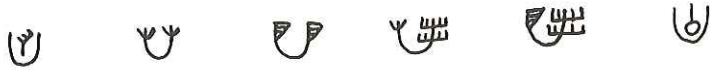
2.2.1

To get the possible characteristics of the numeral signs and to see if any other sign shares the same characteristics, we shall then have a closer look at such stable combinations. For convenience and concreteness, we consider the sign that looks like a vessel or container. (*It is not*

assumed that this actually stand for a vessel-like object). This sign occurs aligned with numerals almost exclusively in type 3 and type 2 objects in the forms

completely occupying one of the sides of the object. These combinations have total frequency 267. This high figure is in part due to the duplicated nature of type 3 objects, but the frequency is high enough even if we ignore duplications. This means, the numeral signs and can be grouped together, the group having the property that its members regularly form stable combinations with .

Now, one of the important characteristics of Indus signs is that new signs are formed by combining two or more separate signs. There are indications that such COMPOUND SIGNS, also called LIGATURES, actually stand in general for stable combinations; that is, in certain special cases stable combinations, in the course of their gradual evolution, are simplified, when the simplification is possible, into a single ligatured compound sign, in order to economize the space and/or possibly to incorporate additional meanings. For instance, the compound sign , which occurs at the left end of a line (with frequency 126), stands for the combination . This is clear since this compound sign is never preceded by that has the character of occupying the left end or quasi-left end of a line like an established convention, being a member of C1. As an another example, the combinations and are occasionally simplified into and . Thus, the sign continues to occur in later type 1 and type 2 objects but in the form of ligatures with other signs. Frequently occurring such ligatures are



We now isolate the signs that modify in the preceding ligatures. In this connection the following important remark may be kept in mind throughout.

REMARK: In isolating the signs that are ligatured with signs such as it is important to note that forming a compound sign is analogous to forming a compound-word. When a compound-word gradually acquires its own life and takes the form of a single word, the individual morphemes that formed the compound-word might undergo some changes within the compound-word, due to the process of mentalization and the eventual breakup of the intimate relation that once existed between a compound-word and the component morphemes. In fact, it appears that we face the Indus script at a stage where many of the compound signs have already attained a similar such character. Thus one should expect that the form of the signs within the compound sign might not coincide exactly with their individual sign forms. Even within the compound signs there might occasionally be variations. For instance, the compound sign occurs also in the form occasionally. However, it should in general be possible to identify the signs at least through external statistical structural similarities, provided other factors such as for instance the need for introducing additional semantic suggestions were not brought into the construction. For instance in the case of the sign , one can isolate the signs and which also occur individually, but it is not clear in what sense these were combined since it appears that other factors might also be involved in forming the compound sign involved, that looks like a man bearing a load on his shoulder. In such ambiguous cases, the signs that appear compound need to be treated in general as an indecomposable single sign, though it might be possible to use them to support a possible general meaning of, for example, the man sign in the preceding case.

Now, the sign involved in the ligature is easily identified to be , since the ligature stands for the combination . Similarly, another sign involved in is . Further, the sign has almost identical functional form and frequency with the sign . In fact, Koskenniemi and Parpola (1979, page 13) argue that these two are just variants of one and the same sign, extending the same conclusion to the pair and . We need not assume such an identity now, but appears to be the only sign whose external structure matches with the sign that can be isolated from that has also variants of the forms such as . Such an identification is also in part confirmed by where it is clear that the sign is involved. Unlike the sign , the sign is not easy to

attach and duplicate at two places within another sign, and this explains the morphological change within the compound sign. Similarly, the only sign that matches the external structure of the sign involved in is , but we shall provide more internal evidences later regarding this. Thus, we classify the signs

together with the numeral signs identified earlier. These have respective frequencies 212, 240, 132 and 170 (for their individual occurrences, that is, excluding their occurrences within a compound sign). The sign occurs, either alone or in ligatured form with other signs, with frequency of about 6.5 percent.

2.2.2

Now let us call the sign a ROOT, and the preceding signs used to modify , MODIFIERS. Also let us call the modifiers in the form of numeral signs NUMERAL MODIFIERS and those in the form of linear signs, such as , LINEAR MODIFIERS. We now use the modifiers we have derived to identify further roots and modifiers. The sign occurs in the modified forms

Keeping in mind the remark of the previous section 2.2.1, we have already identified the modifiers in the first three of these. (It appears that the one involved in the third one can be taken to be either or , but we shall return to this point later in 3.5.3.) Let us identify the modifier in the last of these. There are signs of the form which is clearly an attempt to ligature with . It is possible that in view of the difficulty involved, this ligature is simplified into . (Similar situation occurs with respect to as will be discussed later in section 3.7.0). The sign has total frequency 102. There is no sign other than this one having the structural match with the possible sign that can be isolated from . Thus, we classify with the preceding linear modifiers and with . The

sign , either alone or in ligatured form with other signs, has total frequency of about 4.5 percent. (Here and in what follows, the given percentages refer to the situations where the occurrence of the sign in question can be identified without ambiguity).

REMARK 1: Note that we have not included in this discussion since it has already been classified into C3, and hence has the character quite distinct from the compounds such as . In addition, the sign that can possibly be isolated from this is of the form which does not coincide with any of the sign forms. The ones that look similar to this, such as , have very low frequencies, individually or in combination with others, and later we shall show that they are related to, but different from, the sign . Thus the sign is possibly not intended to be a ligature of and , since a general character of a linear modifier is that it occurs with a reasonable frequency. Also one cannot take the sign as the one involved in since ligaturing usually simplifies the structure of the modifier within the ligature. Similar remarks can be made with respect to or its variant .

Next, the sign forms stable combinations with numeral signs as well as with the modifier . It also occurs in the ligatured form , with high frequency (216). Hence is classified as a root with , and is classified as a linear modifier. The sign occurs individually with a relatively low frequency 14, but it is extensively aligned with a few other signs (in addition to possibly forming a basis for signs such as). The sign which has the appearance of a fish occurs either alone or in modified forms with frequency of about 9.2 percent.

Similarly, the sign in is classified as a modifier. The sign occurs separately with a reasonable frequency (26). In addition, the sign , with frequency 49, has the possibility of having the same value . The same is possibly the case with which occurs with a relatively low frequency. (In the pre-urban pottery of Rehman-Dheri mentioned earlier, it appears that the sign forms , , , , are used interchangeably, see Parpola and Shah 1991: 352-378, or section 2.3.4 below).

REMARK 2: Note that the sign does not occur exactly in this individual form, but it will become clear that it itself is a modified form of the sign , analogous to , , , etc. It appears that

there is some ambiguity in the identification of the sign in the concordances, since it normally occurs in the form in its individual occurrences. The reason for lengthening one of its legs to have the form is just to adjust its height to the height of signs that occur adjacent to it. A closer study of the pictographic forms of this sign using the photographic corpus of Parpola et al. (1987, 1991) clarifies this point. In particular, for example, the sign represented in the form in the concordances is actually a compound sign formed by the signs and . Thus the root involved will be taken to be , and the form will henceforth be considered to be its variant. These two, either individually or in ligatured form with other signs, have the combined frequency of about 1.4 percent.

The only other sign that can possibly qualify as a linear modifier appears to be since it is involved in the compound signs such as , ; but note that these compounds have relatively low frequencies in comparison with the preceding compounds. The sign has the frequency 23.

It may be noted that the exact nature of the sign , which is a modified form of and has a high frequency (279), is not clear at this point, but several suggestions point to the possibility that it might stand for the ligature of and , but much more evidence is needed to verify this. We shall return to this later in Section 3.7.0 of Part IB.

2.2.3

Now, the preceding procedure of identifying the roots and modifiers involves compound signs. One can attempt to extend this procedure to stable combinations, but unfortunately, such an extension runs into several ambiguities. For instance, since we have already identified as a root, one can use it to identify further modifiers, now with the help of stable combinations. It turns out that many other numeral signs can also form stable combinations with . For this reason all numeral signs and other modifiers derived earlier can possibly be grouped together. On the other hand, it also turns out that many compound signs such as , formed by other roots and can also pair with , with reasonable frequency, a fact which is not much helpful at this point. Alternatively one can attempt to use the modifiers to

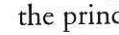
isolate further roots with the help of stable combinations. It turns out that numeral modifiers themselves are paired with other linear modifiers, as well as that linear modifiers can pair among themselves. Thus, stable combinations (that are not in the form of compound signs) are not much helpful in isolating further roots and possible modifiers without ambiguities. Also, if we confine ourselves to compound signs, there are no further linear modifiers with reasonable frequency. Thus, it would be convenient to consolidate what we have derived. That is, we group all the numeral signs to form class C4a, and group the linear modifiers derived above together to form class C4b. The distinguishing analytical property of C4 is that the members of it are used to INTERNALLY modify the roots or to form certain clearly identifiable stable combinations. We have also included in C4 the numeral signs with higher order strokes, even though they are not used to internally modify the roots, since they were probably not used in ligaturing due to difficulties in forming compound signs with them. For instance, the combination occurs with high frequency and here it is not easy to form a compound sign combining and . For convenience we collect together the linear modifiers:

, , , , , , , , ,

2.3.0

It would be now convenient to study the characteristics of the remaining signs in relation to their properties with respect to class C4, defined above. To this end, the following rough pattern that emerges from the concordance will be helpful to devise further steps. First, for the moment, let us assign the combinations formed by the signs within the class C4 the same characteristics as those of individual signs in C4. The reason for this is that in forming the compound signs such as , the linear modifiers , , are grouped together having possibly the meaning of a single unit. There are also compound signs of the forms , , etc., formed among the linear modifiers. Second, by classifying a sign as a root let us understand for the present that it roughly means a sign whose dominant property is that it is regularly used as a basis to form new units in the form of ligatures or stable combinations.

Then a rough description of the CONSTRUCTION PATTERN is as follows. The root signs such as , , , and their ligatures as well as several other signs that have the character of roots, together with possible combinations of signs of these categories with signs of character C4, are put together in a syntactic manner to form what we have called principal blocks. Let us call these constituent units, with roots as their bases, that form the principal blocks, CONSTRUCTION UNITS.

What we mean by construction units can be described more precisely through some examples. For instance, in the text  the principal block is  . This is a combination of three construction units , and , where is a root, is a ligature of the root with  , and is a stable combination of and , with being the ligature of the root with .

As an another example, consider the text  . Here, the principal block is . The combination is formed by the members of C4 and has been given the characteristics of those of C4. Then the block may be considered as the combination of two construction units and , where the first unit is the combination of the ligature with and the second one is the ligature of and .

It is important to note that there are also texts consisting of basically a single construction unit with the nature of the root involved left unspecified or to be understood only implicitly, that is, texts formed by the suitable combinations of signs from within the class C4.

2.3.1

The signs from C3a are also involved in the construction of principal blocks, though with frequency noticeably less than that associated with their role with respect to the character of class C3a. But one can notice, from the concordance, certain features in the pattern of their occurrences, in the sense that their relation to roots is similar to that of linear modifiers of C4, except that they are not used to internally modify the roots, whereas their relation, when there exists, to C4 is similar to that of roots. Out of these two characteristics, the first one is more dominant than the second one. A few other signs that can be isolated to share this characteristic in forming the principal blocks are

These will be noted by C3b. Note that roots and their ligatures themselves have to some extent this characteristic in forming the principal blocks, but their defining dominant property is that they are regularly used as bases in the formation of ligatures or stable combinations. It is also important to note that the relationship of some of these signs, such as , with C4 is not very clear since they form stable combinations mostly or only with C4a, numeral modifiers, so that some of them might be actually closer in character to C4b, linear modifiers. Such ambiguity will not be important since in such cases; only their relationship to those having the characteristics of the roots needs to be taken into account.

Another crucial sign in the construction of texts is the man sign, , but, as noted earlier, this has the unique character of being ligatured both with the signs in C4 as well as with the roots and the signs that have the character of the roots. In fact its occurrences often give the impression that an entire sequence is a sort of being ligatured by it. Further, it occurs, either alone or in ligatured forms with other signs, with very high frequency, about 6.34 percent. These and a few other facts that will become clear later, suggest that this sign does not classify itself into any of the preceding classes.

2.3.2

It is clear from the preceding discussion that so long as the nature of the construction units is formed by the roots such as , , and as their bases, they share certain distinguishing characteristics analogous to linguistic units such as words and compound words. Then the next step would be to group all those signs that form the bases of such construction units. In this connection note that, in an extended sense, the ligatures such as can themselves be taken to be such bases since they occur with high frequencies. For instance and its variant have respective frequencies 102 and 134. Such an extended approach appears to be essential with respect to some of the signs since, unlike the case where the root involved is clear, there are signs which appear to be compound signs but are so

conventionalized or modified that attempts to isolate possible roots might run into risks of several kind. We have already seen one such possibilities in the form . Another example is . By looking into the analogy , which is a variant of , the combination is possibly involved in the compound , but the nature of the root involved, if any, is not at all clear. Still another example is (which has the vague appearance of a bull), and since a deliberate effort appears to have been made to preserve this complete outer form with individual details throughout the entire period, one can argue that this sign is a sort of combination of three signs , and , but there is no guarantee at this point of analysis. Even if this is true, there is no guarantee as to the exact meaning attached to such a combination. A similar argument can be given to the sign (which has the vague appearance of a hoe).

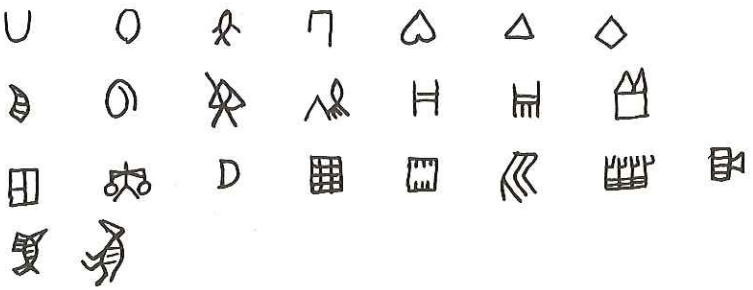
When the preceding extended approach is taken, it indeed turns out, through a careful study of the concordance, that a vast majority of the remaining signs have the common characteristic of forming, by themselves, or as suitable bases for, the construction units, as will be specified below. We shall call such signs EXTENDED ROOTS, which in general also include what was earlier called roots. However, there are a few signs whose nature of construction needs to be analyzed with more care. Specifically, there are a few signs occurring both in an 'empty' form and in an 'infilled' form. For instance the linear modifier , a member of C4, also occurs in the form with reasonable frequency, and there are reasons to believe that these two forms have different meanings .

To see this, consider the block , which occurs three times in the corpus. Here is a root so that is attached to . Also stands for the combination , which has the character of C4. This means, if and are given the same value, the entire block needs to be taken as a single construction unit which is very unusual in form. Also, the question arises as to why only is infilled whereas and are left unaltered. The only consistent plausible interpretation appears to be that we have two construction units and , where and have the character of the roots. Thus, infilling forms appear to take the *character* of extended roots. A few other such forms are and obtained respectively from and . This does not mean that empty forms of an infilled form cannot have the character of an extended root. For example, both forms and

appear to have the character of extended roots, though these two do not occur in the same text, indicating that the infilled form has the possibility of having a value different from that of the corresponding empty form. Later we shall give internal evidences to the possibility that the infilling forms such as stands for the combination or . In this connection, it may also be recalled that the signs in C3, when they occur within the principal blocks, share to some extent the characteristics of extended roots. *In this sense, it is a possibility that some of the extended roots, depending on the context, were used within the principal blocks with emphasis analogous to those in* C3, *or some of the signs classified as extended roots might even be closer to the character of* C3 *in their relation to other extended roots. For this same reason it would occasionally be convenient to treat the members of* C3 *as extended roots for the purpose of decomposing the principal blocks into construction units.* Such minor ambiguities cannot be avoided and will not affect the understanding or structural reading of the texts.

2.3.3

With the preceding discussions in mind, let us list the basic forms of the (extended) roots associated with signs listed in the Appendix B:



Note that by listing these as basic (extended) roots, we also give the meaning of (extended) roots to the signs that are derived from these, either in the form of ligatures or modified forms. For instance, is included means, its modified forms, suggested by the results of later sections, such as

that is, the signs in row (E6) of the Appendix B, also take the forms of extended roots. This does not mean they have the same meanings. Similarly, the extended roots corresponding to the roots

are given in the respective rows, (E1) - (E5). The signs in rows (E1) - (E7) form the class of roots and extended roots. It is a possibility that a few other extended roots may also be classified into these roots, for instance  into , but we have avoided them in order to isolate those cases where we can give unambiguous interpretations. These classifications will gain more and more strength as the steps progress.

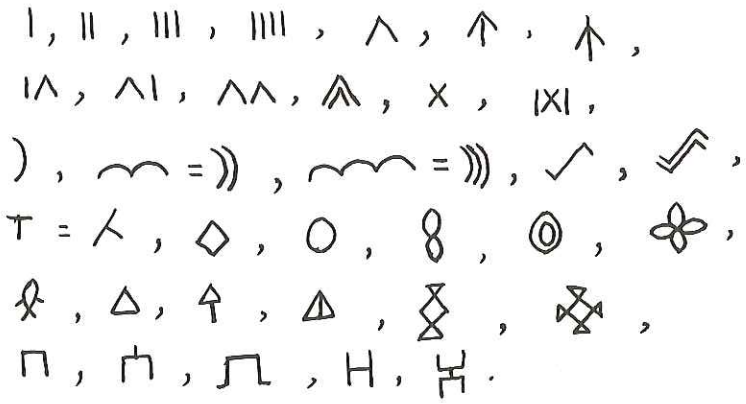
It is also important to note that some of the basic root forms, such as , occur only rarely in comparison with some of their modified forms, even though in the pre-urban phase at Rehman-Dheri, only basic forms are involved, as will be indicated below. This is not unusual since for instance the root  may be considered rare (about 0.06 percent) in its basic form in type 1 objects, in comparison with its very high frequency of occurrence in the earlier, less frequent, types 2 and 3 objects. Similarly, the root  occurs in the pre-urban phase in a more basic form  which appears to be absent in the later stages under consideration.

One unfortunate aspect of the Indus Script is that there appears to be no way of dealing with the evolutionary trends of the sign formations that will indicate the precise differences in meanings between different signs derived from possibly the same basic form, except when they are in the form of clearly identifiable combinations such as ligatures. We can not even determine if some such similar differences in forms are due to their use in different time periods, the total span of which may be at least 600 years, since there appears to be no reliable method of classifying the texts according to different time periods, except the broader classification in terms of the types of objects given earlier. In addition, the main medium of writing appears to be different from that of seals and has not survived. The texts themselves may not be of much help in this respect since even though the meaning of different forms may remain the same, they might not have been employed in the contexts of same texts or construction units in different time periods. Thus the suspicions such as that  and  might have the same value, which is also strengthened by the fact

that these are not used in the same texts, cannot be verified. We shall give more such examples later. Thus it is possible that there are some redundancies in the sign forms. However, these difficulties will not in any way affect our attempt at structural reading and understanding of overall characteristics of the texts.

2.3.4

REMARK : For later purposes we now list the signs occurring on the pottery at Rehman-Dheri. The sign forms of this list appear to differ from those of Durrani (1981), the reason being the difference in the angle of viewing the signs, as they are in the form of isolated marks with no definite rule with respect to the angle of their markings so that different markings of the same sign can be misunderstood as different signs. The sign forms below are obtained by viewing them in order to have internal consistency as a suitable system as well as to be consistent with the sign forms of the period under consideration, as will become clear as this paper progresses. We have used the photographic corpus of Parpola and Shah (1991: 352-378), in obtaining the sign forms. Durrani's list contains a few more signs such as which we are unable to identify clearly, possibly since the material he had, being the excavator, might have been more extensive. Many more sites similar to Rehman-Dheri are also said to have produced 'potter's marks' resembling Indus signs, see Parpola (1986) for a brief review, but unfortunately we have not researched them. It may also be noted that this data is used here only as an additional support of our results, that is, the validity of our results does not depend on this data.



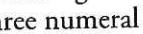
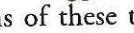
3. Characteristics and structural analysis and possible metrology

3.1. Introduction to the chosen approach and possible alternatives

3.1.0

Having described the classification of the signs and the associated construction pattern of the texts, we shall now see if these results suggest any direction as to the next stage of the analysis. The technical terminologies introduced in the previous sections will be used in the rest of the paper without further explanations. For convenience we recall them: PRINCIPAL BLOCK, STABLE COMBINATION, COMPOUND SIGN or LIGATURE, ROOT, EXTENDED ROOT, MODIFIER, NUMERAL MODIFIER, LINEAR MODIFIER, CONSTRUCTION UNIT and CONSTRUCTION PATTERN.

First, note that it follows from the classifications derived in the preceding steps that the signs in C4 play a pivotal role in the construction of construction units and the principal blocks. Therefore any detailed understanding of the signs in C4 will lead to a better understanding of other classes of signs, such as roots and extended roots, and hence of principal blocks in particular. In fact, the signs in C4 account, either alone or in ligatured forms with other signs, for the frequency of nearly 30 percent.

Now, recall that the signs in C4 are classified into C4a and C4b, consisting respectively of numeral modifiers and linear modifiers, both having certain common characteristics, in particular with respect to internally modifying the roots. Then, in view of the preceding observations, there are two possibilities. The first one is that the numeral modifiers actually stand for numerals, either as a number concept or as a number word. Within this first possibility, there are several suggestions that the possibility of linear modifiers standing for the same cannot be excluded. For instance, when ligatures are formed with roots such as  , apart from the first three numeral modifiers  , only linear modifiers are employed, suggesting the possibility that linear modifiers are also continuations of these three numeral modifiers in their possible meanings. In fact, it is also rather curious to notice that the combination  , with numeral modifier of fourth order, that occurs with high frequency in objects of types 3 and 2, becomes absent in later type 1 objects, indicating the possibility that this combination might have been replaced later by an alternative, possibly simpler,

representation. One possible way to get some further clue regarding such possibilities is to attempt to retrace the etymologies for the number words for numerals and to see if such etymologies have any conceptual connection with the linear modifiers, as the forms of some of them are not entirely abstract and they do suggest, through internal and external evidences, some concrete objects such as a grain, a hand, a pestle, etc.; this will become clear below.

The second possibility is that the numerals were used for the sound value of the corresponding number words or their components, as linguistic building units in forming words and phrases. This also leads to the etymological study of the numerals.

As was mentioned earlier, it is the opinion of several scholars, based on archaeological and linguistic data, that a *form* of Dravidian is a plausible language of the texts. However, this is not certain, but one can handle this difficulty by devising the further steps and methods in such a way that any contradiction to the possibility of the language being a form of Dravidian will be easily revealed. Another type of difficulty has to do with the remoteness of the time period under consideration with respect to the so-called historical period, which starts only around the third century B.C. and for which reliable written documents are available. Still another difficulty is the geographical distance of the present location, which is the southern part of India, of major languages that have the apparent classification of being Dravidian, though it is the belief of many that Dravidian substratum is present in many ways in those that have been classified as Indo-Aryan languages of India. (*It may be noted that these classifications go back to the period of 'Comparative Philology' and is not based on the notion that a language is an evolving complex system with inherent inter-related structures at many different levels*). However, in the present situation, we are interested only in concrete objects or concepts such as numerals, so that one can hope to overcome these difficulties. **In fact our aim will NOT be to reconstruct the exact phonological shapes of the words spoken at the remote time period under consideration, but rather only to construct an approximation to their bases in order to see if they are in any way related to concrete objects, concepts or actions.** (See the remark below for more details of the limitations). For instance, one cannot ignore the possibility of numerals having been evolved from concrete counting or metrical measurements.

Unfortunately the etymological studies for numerals that we are aware of, for example, Andronov (1973) and Emeneau (1957) use

certain comparative methods based mostly on phonetic correspondences among the various present Dravidian languages, and hence it is not clear if the phonological shapes thus derived can be entirely relied upon for the present purpose. However, if one further takes into the possibility of the involvements of concrete counting and measurements, the etymologies presently available lead to a recognizable picture rather quickly. The details of this move constitute the next step.

At this point, a somewhat alternative approach may be indicated if one is willing to assume at this stage that the numeral signs actually stand for numerals, and leaving the verification of this assumption to a later stage. In view of the fact that the linear modifiers (class **C4b**) classify themselves together with numeral modifiers, one can then tentatively argue that linear modifiers also stand for numerals. Then the same arguments involved in the remaining subsections of the present section, but excluding the ones that involve etymological arguments, will lead tentatively to the results of this section which can then be supported by the etymological study of this section. The readers who may feel uncomfortable with etymological arguments can actually follow this course of reading for the remaining part of this section. However, we prefer to present the relevant etymological arguments (in the sense of the following remark and limited to some basic numerals) whenever the need for them arises, since such an approach appears to be a natural one for the present situation, as is suggested by the reasonings given above, in addition to the fact that the suggestions of each stage of the analysis will have much stronger logical support than otherwise. Further we would like to keep our methodology closer to one involved in the subject of 'stochastic grammatical inference' in which the meanings (semantics) aspect of the structure is an essential part. Another reason is the limited nature (basically involving, in this section, the numerals four to ten and hundred) of such etymological study, as indicated earlier and in the following remark.

REMARK : The limitations indicated and emphasized earlier of the etymological study naturally impose other limitations of this work which should be kept in mind throughout what follows in order to avoid certain possible misunderstandings.

First, the paper will be concerned with the approximate etymologies of only a certain restricted category of words related, as it will turn out, to some numerals and metrological units. Hence, any suggestions derived from such results may not apply to the possible forms of the

language as a whole for the period under consideration, since the nature of the linguistic changes with respect to a restricted system of words may not give clear indications of the corresponding changes for the entire language. It may also be noted that we are not aiming at complete phonetical reading.

Second, since our primary aim will be mostly limited to identifying the concrete meanings of a restricted system of words, and in view of the remoteness of the period under consideration, we will not always attempt to strictly follow the standard procedures, for instance in deriving the approximate phonetical change from the form of a compound-word to the form of a single word, employed in linguistic literatures. However, it is important to note, for instance, that when one identifies approximate changes with the above limited aim within a restricted system whose meaning is clearly understood at least in a broader sense, strict adherence to the 'principles' of the chosen linguistic method, if not often be only a secondary, need not be the sole reason for the validity of the meanings derived.

Sound values are invoked primarily to collect words that are conceptually related by a network of 'relative motivations' (see Saussure 1959: 131), and then the associated meaning, subject to the restriction indicated above, is isolated from such a network. Within such a limitation we aim the approach employed to be as rigorous as possible. Note however that we employ the Saussure's notion of 'relative motivation' in a somewhat broader sense than Saussure's discussions directly indicate (but see however his notion of 'mutability', ibid.: 74). For instance, in Dravidian, the lexical forms corresponding to 'four', 'several', 'everything', 'prosperity', and 'grain' are arbitrary, that is, unmotivated, in the sense of Saussure as he restricts the discussions to the synchronic (or static) aspects of the system. But when diachronic or evolutionary aspects of the system are taken into account it can be argued that all the above lexical forms are relatively motivated in the sense of sharing a relationship to a common conceptual background that is responsible for the closeness or similarities of the sound values involved. (See the discussions in Section 3.2.1 below). To indicate another axis in which the evolution takes place, consider the following lexical forms for mother: *ammā = aṉ-mā, annai = aṉ-uy, avve = ā-vēl, āthā = ā-ti, thāi = ti-uy.* Although these differ in their basic sound forms, all of them share a common conceptual background of divinity as the forms *aṉ, vēl* (*spear*), *ti* (*fire*) conceptually stand for divinity.

In view of these limited purposes and for the reason of personal convenience, or occasionally when no Proto-Dravidian forms are suggested in the literature (possibly by our limitations with respect to the subject involved), we often employ approximate forms closer to Tamil, one of the major Dravidian languages believed in general to have the characteristics, especially the phonological structure, closely related to the Proto-form (not necessarily the form coinciding with that of the Indus texts) than others of the family, whenever it is found that the chosen forms will serve our limited purpose at least as well as any other possible approximate forms. However, it will turn out that as the steps progress, one increasingly gains confidence in the approximations used, possibly in view of the fact that only simple word forms of a restricted system are involved. In this connection, it should also be emphasized that two languages, being evolving complex systems having different phonological structures, may very well have gradually evolved from a same language; the reason for the divergence can be many including the coexistence with another language having some difference at least in the phonological structure, so that the relationships of any one of the present languages of India, especially the ones that have been classified as Indo-Aryan, to those of the Indus cannot be concluded from a part or whole of the phonological aspects alone. (In this connection it appears that some form of the approach suggested in the previous paragraph might be of some importance in studying the relationships among the different languages of India).

3.2. Analysis of fitting some basic numerological units

3.2.0

Note that, as was indicated above, we first need to look into the possibility that the linear modifiers (class **C4b**) are continuations of the first three numeral modifiers, that is, from four onwards.

Let us now recall the numeral bases from four to eight.

Numbers	Numeral base
Four	*nāl* or *nālu*. Has the meaning several or everything
Five	*cai-*. Has the strong possibility of being connected with the word *kai*, meaning hand.
Six	*cāṟu*.

Seven	*eḻu* or *ēḻu*.
Eight	*eṭṭu*.

These forms for four to eight are commonly accepted as proto-Dravidian (see, for example, Andronov (1973) and Panikkar (1969)). The meanings for four and five stated above are as given in Andronov (1973). The precise form for nine is not essential for the present purpose, but according to Emeneau (1957), nine is defined subtractively from ten. (For eight, the adjectival form, such as when it occurs as the first constituent of a numeral construction, has the form *eṇ*.)

Since our aim is to fit the linear modifiers to the preceding numerals, let us recall them.

3.2.1

First let us identify easier cases. As stated above the numeral base for the five has the possibility of having the meaning hand. This possibility is also strengthened by the fact that, in Dravidian, *eṇṇikkai* (DED,678, see the following remark) means numbering, which might have initially stood for the meaning 'one to five', since the archaic numeral base for one is *oṇ* or *aṇ*, so that, if *kai* stood for five, one has *aṇ-i-kai* → *eṇṇikkai*. (In viewing the suggestions similar to this, limitations indicated in the preceding remark may be kept in mind). Then the linear modifier may be identified with five, since it is possible that this sign stands for a hand, though in a conventionalized form, as is also suggested by the compound signs such as .

REMARK: Here and in what follows DED stands for the work *A Dravidian Etymological Dictionary*, by Burrow and Emeneau (1961). (Unfortunately, the second revised edition of this was unavailable to us). Most of the lexical forms used in the present paper can also be found in *Tamil Lexicon, Vols. 1-6*, (1936), as it generally incorporates the lexical forms of other Dravidian languages also. Whenever, though occasionally, the source DED is not indicated it is to be understood that the form involved is taken from this lexicon.

Next, from the several variants of the signs  and  given in Mahadevan (1977: 785), it is clear that both stand for grain, which has the word value *nel* or *nellu* (DED,3112), the sound value of which approximates that of four given above. Hence, let us identify both of these linear modifiers with four.

It may also be noted that a grain is used to signify four; this is not entirely unexpected, since the lexical forms corresponding to four, several, everything, prosperity (by possessing 'everything') and grain were possibly not differentiated at an archaic stage, as their meanings are closely related (if one assumes that grains were identified with prosperity, as is natural as they for instance might have formed the basis of currency) in addition to having approximately the same sound values (in Dravidian).

Note that the possibility that four having two sign forms cannot be excluded, since we are dealing with an evolutionary process. In addition, as was mentioned earlier, Koskenniemi and Parpola (1979) actually argue that these two forms are just variants of each other, in view of their almost identical frequency and positional and functional nature of their occurrences in the texts.

3.2.2

Now, let us look at the linear modifier  . This sign occurs as a modifier in the ligature  ,that has the appearance of a mortar, with high frequency (236). This suggests that the pictorial form of  might have initially corresponded to a pestle. The etymology for pestle is 'hand of mortar'. So, let us analyze the etymology for mortar. In an archaic stage it is likely that mortar, which is a kind of vessel and if it was viewed as such, was distinguished lexically from other types of vessels, if at all necessary, only through additional significations involving things such as concepts or actions. (Here and in what follows, by an ARCHAIC STAGE, we mean a stage in the remote past in which the word or the compound word under consideration was at the initial stage of the process of formation or evolution.) With this in view, a possible approximate form of evolution of the signification used for mortar is: *cōra - uḻu → cōrulu → orulu* or *uralu*, where the rightmost side gives the current forms of words for mortar (DED, 560), while in the leftmost side, *cōra* means a vessel or earthen pot (DED, 2355) and *uḻu* in general means 'work' and specifically means 'to plough', 'to dig

up' etc. (DED, 592). Thus, a mortar was signified at an early stage by 'vessel-work' or 'vessel-digging'. Note that the Proto-Dravidian form for six given above has undergone very similar phonetic change in the form *cāṟu → aṟu* or *āṟu*, where the right side gives the current form for six. In addition, the materials which are carried in the *cōra* have the word forms such as *cōṟu* (boiled rice, DED 2360) and *caru* (juice, sap, pepperwater, broth, etc., DED 2050). This means, both a specific vessel and some of the materials which were carried in it had the same lexical forms, confirming the archaic form of the word *cōra*.

Thus, these arguments suggest that the numeral base *cāru* for six might have originally stood for a specific kind of vessel, such as the one used for mortar, whose capacity served as standard volume measure that is six times the capacity of another standard vessel. This means, it is possible that the vessel that was used as mortar came to be identified with the numeral six, at least in a concrete sense, and then the abstract numeral six gradually evolved and signified by the notation of pestle , and the full notation was retained possibly for a certain volume measure, when the necessity to notationally differentiate between the concrete and the abstract arose. Note that both signs and occur in type 3 objects, in conventionalized forms that remain unchanged throughout the entire period. Thus the preceding possible evolutionary stages must have occurred long before the period of type 3 objects.

It is interesting to note that a sign similar to was used on clay tablets in the impressed form by Sumerian around 3000 B.C. to signify both the numbers one and sixty, differentiating only by the size of the sign, see for example Friberg (1978). In the present case also, if stands for an amount of one unit, then might possibly mean six times the amount . Later, we shall note a similar analogy with respect to the sign form for ten. Note that this similarity also holds with respect to the Proto-Elamites whose notations for some of the basic numerals, but not all, were similar to those of Sumerian.

3.2.3

Next, let us see that the linear modifier  has the possibility of signifying the numeral base eight. This is one of the signs that occur in the pre-urban pottery of Rehman-Dheri mentioned earlier, see Parpola and Shah (1991: 355-360), and also section 2.3.4. above. The same

word *eṯṯu* (eight) also stands for words meaning step or jump, to reach up, etc (DED, 669). Also, *ettu* which approximates the sound of *eṯṯu*, means kick. Now, the sign occurs in the form of ligature with the man sign (with frequency 20), which is suggestive of a man stepping or jumping. (Below in 3.2.6, it will be shown that the man sign has the possibility of being used for the meaning such as 'measure' or 'amount' through 'phonetical transfer'). Later (in Part II, section 5.1.2), we shall provide a few instances where the sign is used for the approximate sound value, either a suitable skeleton or the whole, of the word *eṯṯu*. (Another suggestion that will become clear later is that the signs and have the possibility of being derived from , having closely related, but not the same, meaning of which is suggestive of a length measuring device). All this suggests that a 'step' or 'jump' was used as a length measure at an archaic stage, and gradually acquired the numeral value eight in the concrete sense that eight times a lower unit formed a step or jump, which later evolved into a base for abstract numeral eight.

The preceding derivation then immediately suggests that both the modifiers and have the possible meaning of the sum of eight and one, that is nine. The same is possibly the case with .

3.2.4

Now let us look at the linear modifier . Unfortunately, we have been unable to identify any concrete object with respect to the outer form of this sign. However, we shall now show that there are several internal evidences to the possibility that this sign stands for the numeral base seven. First, note that the numeral base *eḻu*, that is seven, also in general means work involving physical effort, such as to rise (as from a seat), erect (a building), etc. (DED, 723). There are also many similar sounding words having similar meanings, such as *uḻu* (to plough, to dig up, etc., DED, 592), *eṛi* (throw, discharge, etc. DED, 731) and *ēṛu* (to rise, ascend, etc. DED, 776).

Now, the sign occurs in the form of the combination with high frequency, with occupying the left end position of the line. In fact itself has total frequency 73, of which it occurs 52 times in the form of this combination. Thus, gives the impression of having the characteristic closer to the class C1, and hence the possibility that it does not contribute a concrete

value with regard to this combination. Actually, in 61 cases it occupies the left end of the lines, in 7 cases it takes the left end position of the principal blocks, and in the remaining 5 cases, it takes the right end position 3 times and the remaining 2 in the form of the above combination. This possibly means, the combination has the same value as . Now, has the possibility of representing a plough, as is suggested by its sign variant, , and the word value 'to plough' (*uḻu*) in general also means 'work'.

Thus the preceding discussions indicate that and signify the same value and the precise value is associated with 'work' and indicated by , that is, part of the function of is to be a determinative in the usual terminology. Since the word value of the numeral base seven also means 'work', it is possible that the numeral base seven is evolved from a unit of work, possibly in the form of a unit of a time, in exactly the same way that six and eight were evolved from the metrical units respectively of a volume and length; the notation used to signify this is or . In addition, the archaic numeral base for seven could have been either *uḻu* or *eḻu* (or *ēḻu*), since these lexical differentiation might have evolved at a later stage.

In this connection, it is important to note that the unit of work or time corresponds only to a concrete concept and not to a concrete object or to any specific action. Then, one possible explanation for the involvement of is that seven, that is, work, was signified graphically by the representation of its sound value. This means, if the combination is taken to represent the sound value of work, then the sound value of work is decomposed into two basic units of which the first is represented by (*u*) and the second is represented by (*lu*), indicating a rather evolved form of phonetization. The possibility that 'work' was signified by phonetical transfer is also strengthened by the fact that when occurs not in the form , only rarely it appears to have been used for its full sound value (*uḻu*) signifying seven itself. Eventually only the sign was retained to signify seven, as the sign might have had other functions.

3.2.5

The preceding discussions also indicate that the linear modifiers have word values (in the context of writing on seals); in the case of four, five, six, and eight these were represented by concrete objects whose words have the corresponding approximate sound values, in addition to having corresponding related meanings in the cases of five, six and eight. In the case of seven, it is possible that the sound value of  might have approximated the word value for seven only partly, such as for instance having no or a different vowel value at the beginning, and hence further assistance was necessary when the word for seven occurred at the beginning of a line.

Regarding the remaining numeral modifier,  , we shall postpone the discussion of its possible meaning to a later step. For the present, we shall only note that it has the possibility of standing for the numeral base ten, but it is not the only form of ten employed.

The preceding steps taken together suggest the possibility that the linear modifiers have associations with numerals four to ten, either conceptually or for the value of their sounds. It is then even more plausible to have the same meaning for the numeral modifiers themselves, since both are in the same class C4. However, there might be additional meanings that distinguish numeral modifiers from linear modifiers since, as was indicated earlier, numeral modifiers form stable combinations with linear ones taking the forms such as  , which need not have the same meaning as the combination  , even though in both combinations  and  might stand for four.

3.2.6

Now, let us take up the study of the man sign  , though we shall see that it will have characters different from numerals. As was mentioned earlier, this sign constitutes an important building unit of the texts, occurring with high frequency (about 6.34 percent) in the forms ligatured with members of C4 as well as with roots and extended roots, so that any understanding of this sign will provide valuable clues regarding the characters of other signs, in particular linear modifiers, with respect to which it is ligatured.

The etymology of the word for man is *ā̲l* or *ā̲lu* (DED, 342). This has the approximate sound value with those of *a̲la* (DED, 252, to measure, measurement, size, number, capacity, magnitude, dimension, etc.), *ā̲l* (DED, 338, deep, plunge, depth, to sink etc), *a̲li* (DED, 235, to perish, decay, etc) and several other similar sounding words having similar meanings. Such similarities possibly go back to an archaic stage when the concepts related to measure, amount, grown-up, deep, perished or ruined (by being buried deep), etc. were not distinguished lexically and grammatically (Diakonoff (1983)), and the man was signified simply by 'grown-up'. Thus, the possibility that the man sign is used, through apparent phonetical transfer, to signify measure or amount cannot be excluded. This is also consistent with the numeral signification derived for C4. For instance, if two linear modifiers standing for four and five are paired, with one of them, say four, being ligatured with the man sign, it could mean that 'four is measured five times', resulting in twenty, whereas if four and five are paired without any indications, it could just mean four plus five, that is, nine, since the linear modifiers have the character of numeral modifiers, that is, they are abstract and disassociated with any concrete background. Later we shall also provide a few instances where the man sign has been used for its sound value in general.

The rest of Part I as well as Part II will appear in a following issue. The numerals that we have fit in Section 3.2 above are the First Order numerals indicated in the Abstract of this paper. In the same way, the remaining of Part I will in particular contain a detailed analysis of fitting the remaining class of numerals (see Abstract) to the appropriate classes of signs isolated in Section 2 above, as well as the operations associated with them together with the description of the resulting metrological system.

Department of Statistics
University of Michigan
Ann Arbor, Michigan, 48109

Appendix A

The horizontal bracket ⌊‾‾⌋ below indicates the principal block, and further subdivisions ⌊‾⌋ indicate the construction units of the text involved. Indications of these are in the main body of the paper, and they will be further explained in detail in the context of texts below in Part II.

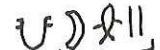
(T.1) 

(T.2) 

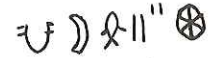
(T.3) 

(T.4) 

(T.5) 

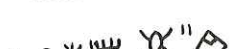
(T.6) 

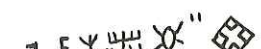
(T.7) 

(T.8) 

(T.9) 

(T.10) 

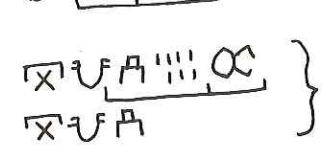
(T.11) 

(T.12) 

(T.13) 

(T.14) 

(T.15)

(T.16) 

(T.17) 

(T.18)

(T.19) 

(T.20) 

(T.21) 

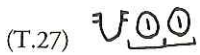
(T.22) 

(T.23) 

(T.24) 

(T.25) 

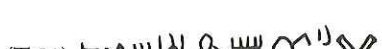
(T.26) 

(T.27) 

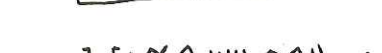
(T.28) 

(T.29) 

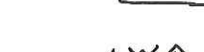
(T.30) 

(T.31) 

(T.32) 

(T.33) 

(T.34)

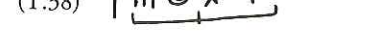
(T.35) 

(T.36)

(T.37)

(T.38)

(T.39) 

(T.40) 

(T.41)

(T.42)

(T.43)

(T.44)

(T.45)

(T.46)

(T.47)

(T.48)

(T.49)

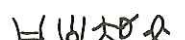
(T.50)

(T.51)

(T.52)

(T.53) 

(T.54) 

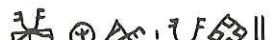
(T.55) 

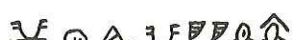
(T.56) 

(T.57) 

(T.58) 

(T.59) 

(T.60) 

(T.61) 

(T.62) 

(T.63) 

(T.64) 

(T.65) 

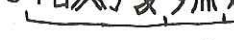
(T.66)

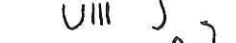
(T.67)

(T.68) 

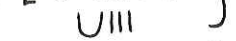
(T.69) 

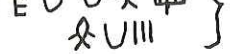
(T.70) 

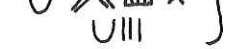
(T.71) 

(T.72) 

(T.73) 

(T.74) 

(T.75) 

(T.76) 

(T.77) 

(T.78) 

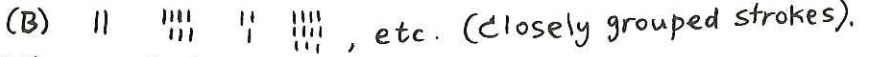
(T.79) 

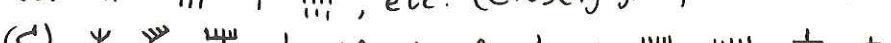
Appendix B

List of selected signs

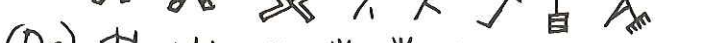
The arrangement given below is influenced by the results of this paper.

(A) 

(B) 

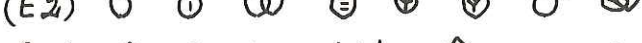
(C) 

(D) 

(Da) 

(E1) 

(E2) 

(E3) 

(E4) 

(E5) 

(E6) 

(E7) 

(F)

References

Allchin, B. and R. Allchin. 1982. The Rise of Civilization in India and Pakistan. Cambridge University Press.

Andronov, M. 1973. Dravidian numerals: An etymological study. International Journal of Dravidian Linguistics 5, p. 5.

Burrow, T. 1969. Dravidian and the decipherment of the Indus script, (Review of Parpola et al, 1969). Antiquity 43, p.149.

Burrow,T. and M.B. Emeneau. 1961. A Dravidian Etymological Dictionary. Oxford.

Clauson, G. and J. Chadwick. 1969. The Indus script deciphered?. Antiquity 43, p.200.

Collon, D. 1987. First Impressions: Cylinder Seals in the Ancient Near East. London.

Damerow, P. and R.K. Englund. 1987. Die zahlzeichensysteme der archaischen texte aus Uruk. In M.W. Green and H.J. Nissen: Archaische Texte aus Uruk, Band 2. Berlin: Gebr.Mann Verlag.

Diakonoff, I.M. 1983. Some reflections on numerals in Sumerian: Towards a history of mathematical speculations. Journal of the American Oriental Society 103, p.88.

Durrani, E.A.. 1981. Rahman Dheri and the birth of civilization in Pakistan. Bulletin of the Institute of Archaeology of the University of London 18, p.191.

Emeneau, M.B. 1957. Numerals in comparative linguistics: With special reference to Dravidian. Bulletin of the Institute of History and Philology, Taipei.

Fairservis, W.A. 1975. The Roots of Ancient India. 2nd Edition. Chicago: University of Chicago Press.

Friberg, J. 1978. The Third Millennium Roots of Babylonian Mathematics. 1. A Method of Decipherment, through Mathematical and Metrological Analysis, of Proto-Sumerian and Proto-Elamite Inscriptions. Chalmers University of Technology and the University of Göteborg.

Friberg, J. 1984. Numbers and measures in the earliest written records. Scientific American 250, p.116.

Gadd, C.J. and S. Smith. 1931. Sign-list of the early Indus script. In: Marshall 1931, p.406.

Gelb, I.J. 1963. A study of writing. 2nd Edition. Chicago: University of Chicago Press.

Heras, H. 1953. Studies in Proto-Indo - Mediterranean Culture. Indian Historical Institute, Bombay.

Hunter, G.R.. 1934. The Script of Harappa and Mohenjodaro and its Connection With Other Scripts. London: Kegan Paul, Trench, Trübner and Co.

Ifrah, G. 1985. From One to Zero: A Universal History of Numbers. New York: Viking.

Jarrige, J.-F. 1984a. Towns and villages of hill and plain. In: Lal and Gupta. 1984.

Jarrige, J.-F. 1984b. Chronology of the earlier periods of the Greater Indus as seen from Mehrgarh, Pakistan. In: B. Allchin (ed.), South Asian Archaeology 1981, Cambridge, p.21.

Jarrige, J.-F. and R.H. Meadow. 1980. The antecedents of civilization in the Indus valley. Scientific American 243, p.102.

Jeganathan, P. 1997. On the conceptual basis in modeling and inference. Technical Report 283, Department of Statistics, University of Michigan, Ann Arbor.

Joshi, J.P. and A. Parpola. 1987. Corpus of Indus Seals and Inscriptions: 1. Collections in India. Memoirs of the Archaeological Survey of India 86, and Suomalainen Tiedeakatemia, Helsinki.

Knorozov, Y.V., G.V. Alekseev, I.K. Fedorova, A.M. Kondratov, M.A. Probst, and B.Y. Volchok. 1965. Proto-Indica: 1965. Institut Etnografii AN SSSR, Moscow. (Translation and critique in Zide and Zvelebil (1976)).

Knorozov, Y.V., N.V. Gurov, and B.Y. Volchok. 1968. Proto-Indica: 1968, Brief report on the investigation of the Proto-Indian texts. Institute of Ethnography, Moskow. (Reprinted with commentary in Zide and Zvelebil (1976)).

Knorozov, Y.V., N.V. Gurov, and B.Y. Volchok. 1970. Proto-Indica: 1970. (Translated in Pande (1973)).

Koskenniemi, S. and A. Parpola. 1979. Corpus of Texts in the Indus Script. Research Reports 1, Department of Asian and African Studies, University of Helsinki, Helsinki.

Koskenniemi, S. and A. Parpola. 1980. Documentation and Duplicates of the Texts in the Indus Script. Ibid. 2.

Koskenniemi, S. and A. Parpola. 1982. A Concordance to the Texts in the Indus Script. Ibid. 3.

Lal, B.B. 1966. The direction of writing in the Harappan script. Antiquity 40, p.52.

Lal, B.B. 1974. Some aspects of archaeological evidence relating to the Indus script. Puratattva 7, p.20.

Lal, B.B. and S.P. Gupta (eds.). 1984. Frontiers of the Indus Civilization. New-Delhi: Books and Books.

Langdon, S. 1931. The Indus script. In: Marshall (1931), Vol. 2, p.423.

Lieberman, S.J. 1980. Of clay pebbles, hollow clay balls and writing: A Sumerian view. American Journal of Archaeology 184, p.339.

Mackay, E. 1938. Further Excavations at Mohenjo-daro, Vols 1-2. New-Delhi: Government of India Press.

Mackay, E. 1943. Chanhudaro Excavations (1935-36). New Haven.

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

Mahadevan, I. 1986. Towards a grammer of the Indus texts: 'Intelligible to the eyes, if not to the ears'. Tamil Civilization, Vol. 4.

Mahadevan, I. and Mythili Rangarao. 1986. The Indus script and related subjects: A bibliography of recent studies (1960-1986). Tamil Civilization, Vol. 4.

Mahalingam, T.V. 1974. Early South Indian Paleography. Madras.
Marshall, J. 1931. Mohenjo-Daro and the Indus Civilization, Vols. 1-3. Arthur Probsthan, London.
Nissen, H.J. 1986. The archaic texts from Uruk. World Archaeology 17, p.317.
Pande, H.C. 1973. Proto-Indica. Translation of Knorozov et al (1970). Field Research Projects, Florida.
Panikkar, G.K. 1969. Dravidian numeral constructions. In Dravidian Linguistics, Annamalainagar, India.
Parpola, A. 1986. The Indus script: A challenging puzzle. World Archaeology 17, p.399.
Parpola, A. 1995. Deciphering the Indus Script. Cambridge University Press. New York.
Parpola, A. and S.G.M. Shah. 1991. Corpus of Indus Seals and Inscriptions: 2. Collections in Pakistan. Memoirs of the Department of Archaeology and Museums, Government of Pakistan, Vol.4, and Suomalainen Tiedeakatemia, Helsinki.
Rao, S.R. 1979. Lothal, a Harappan Port Town (1952-62), Vol. 1. Memoirs of Archaeological Survey of India 78, New Delhi.
Saussure, F de. 1959. Course in General Linguistics. (English Translation by W. Baskin). McGraw-Hill, New York.
Sgall, P., E. Hajičová and J. Panevová. 1986. The Meaning of the Sentence in Its Semantic and Pragmatic Aspects. Edited by Jacob L.Mey, D.Reidel Publishing Company, Dordrecht.
Smith. 1931. In Marshall (1931), Vol.2, p.415.
Vats, M.S. 1940. Excavations at Harappa, Vols. 1-2. Government of India Press, New-Delhi.
Vallet, F. 1986. The most ancient scripts of Iran: The current situation. World Archaeology 17, p.335.
Venkatachalam, K. 1986. A study of weights and measures of the Indus valley civilization. Tamil Civilization, Vol. 4.
Ventris, M. and J. Chadwick. 1956. Documents in Mycenaean Greek. Cambridge.
Wheeler, R.E.M. 1968. The Indus Civilization. 3rd Edition. Supplement to the Cambridge History of India, Cambridge.
Zide, A.R.K. 1970. A brief survey of work to date on the Indus valley script. Journal of Tamil Studies 2, p.1.
Zide, A.R.K. and K. Zvelebil. 1970a. Review of A. Parpola et al 1969. Indo-Iranian Journal 12, p.126.
Zide, A.R.K. and K. Zvelebil. 1970b. Review of Knorozov et al, 1965, 1968; A. Parpola et al 1969. Language, 46, p.952.
Zide, A.R.K. and K. Zvelebil. 1976. The Soviet Decipherment of the Indus Script: Translation and Critique. Paris: Mouton.

Zvelebil, K. 1985. Recent attempts at the decipherment of the Indus valley script and language (1965-1980): A critique. In: Indus Valley to Mekong Delta: Explorations in Epigraphy. H. Karashima, (ed.,1985), Proc. Seminar on South and Southeast Asian Epigraphy, Tokyo, 1983. Madras.