# ARTICLE

Interrogating Indus inscriptions to unravel their mechanisms of meaning conveyance

Bahata Ansumali Mukhopadhyay (ID) [1]

ABSTRACT   This study conducts an epigraphic analysis of the yet undeciphered inscriptions of the ancient Indus Valley civilization and seeks to prove that just like proto-cuneiform administrative tablets of ancient Mesopotamia, or modern fiscal stamps and ration tokens, Indus seals and tablets too were formalized data-carriers that used both document-specific and linguistic syntaxes to convey messages. Analyzing various combinatorial patterns of Indus signs (e.g., typical co-occurrence restrictions present between some signs; tendency of some signs to form collocations; syntactic order(s) maintained between certain signs; strong positional preferences demonstrated by some signs; and the capability of certain signs to occur alone in inscriptions), this paper argues that Indus signs represented different content-morphemes and functional-morphemes—not phonograms used for spelling words—the majority of the inscriptions were logographic. Categorizing several Indus logograms into nine functional classes, it explores the way each sign-class has played unique functional roles for conveying complete messages through the brief inscriptions. By glossing the signs of the inscriptions using the names of their respective functional classes the study unravels the formulaic phrase-structures maintained by the majority of the inscriptions, where signs identified as phrase-final signs typically occur at the terminal positions of the semantic phrases, while signs used as connective-morphemes join semantically autonomous constituents in certain subordinating and coordinating ways, to form longer composite inscriptions. The study identifies certain Indus signs as numerical and metrological signs, which frequently collocate with specific lexemes, and clearly quantifying them in certain ways. This article also analyses the underlying relationships between certain metrological and phrase-final signs; examines the compositional nature of Indus collocations; and argues that some of the repeated sign-sequences were probable examples of linguistic reduplication. Analyzing the occasional occurrences of certain attributive quantifier signs as substantive lexemes, this paper suggests that in some inscriptional contexts those quantifier signs possibly represented certain commonly used metonyms of Indus civilization. Finally, this study explores the compositional semantics of Indus inscriptions without assigning any sound value to the signs and without speculating about whether the script encoded the speech of any specific ancient language.

---

[1] Infor, Bengaluru, India. Correspondence and requests for materials should be addressed to B.A.M. (email: alapchari@gmail.com)

Introduction

Discovered from nearly 4000 ancient inscribed objects, comprising seals, sealings, tablets, ivory rods, pottery shards etc., Indus inscriptions, are one of the most enigmatic legacies of the Indus Valley civilization (henceforth IVC), which flourished between 2600 BCE and 1900 BCE, spread over a vast region of 680,000–800,000 square km of the Indo-Pakistani subcontinent (Kenoyer, 2010). Right from the first publication of an inscribed seal in 1875, Indus Script (henceforth ISC) has suffered several avowed decipherments, most bereft of substantial scholarly consensus. The major factors incommoding decipherment are: absence of bilingual texts; extreme brevity of the inscriptions; ignorance about the language(s) encoded by ISC (if any); and poor chronological control in the existing ISC corpora (Wells, 2011). The meagre points of scholarly consensus can possibly be summarized as: right-to-left direction of the majority of the inscriptions, numerical nature of certain stroke-like signs; functional homogeneity of certain terminal signs, and some generally adopted techniques of segmenting the inscriptions into initial, medial and terminal clusters. Barring these few points, huge controversies surround almost every aspect of ISC. For example, certain scholars consider the script as logo-syllabic (Parpola, 1994, p. 85; Wells, 2011, p. 116; Hunter, 1934; Wells, 2015, p. 32, 85), some others take it as logographic (Koskenniemi and Parpola, 1982, p. 10; Mahadevan, 1978, 1986, 2014), whereas some even deny that it encoded "speech" at all (Farmer et al. 2004). Phonetic and semantic interpretations of individual Indus signs also vary widely. For example, Mahadevan took ⊍, the most frequent sign of ISC, as a Dravidian "pronominal masculine singular suffix", sometimes conveying an ideographic meaning of a "sacrificial vessel with food offerings" (Mahadevan, 2014, p. 10, 17). Wells (2015, p. 88) considered ⊍ as a terminal marker that might have been a verbal ending with the sound value "-ay", which also means "cow" and "mother" in proto-Dravidian. But to Bonta (2010, p. 76–82) ⊍ is a predicate or root, expressing the meanings "be" or "belong to", that was possibly a general predicate-marker. Scrutinizing various decipherment efforts of ISC, Possehl (2002, p. 139) has stated that, "a certain amount of impatience on the part of some researchers" of ISC, has driven them to move "quickly from an initial hypothesis to a series of conclusions and readings".

In my view, the semantic comprehension of Indus inscriptions needs to build on a thorough understanding of their internal structure and archeological contexts. With no established relationship with other ancient Indic scripts, and with our little knowledge of its underpinning language(s) if any, ISC at present falls in the most difficult category of un-deciphered inscriptions. However, remembering how the Linear B script of ancient Greece was rescued from a similar condition through the methodical structural and contextual analysis done by Kober (1948) and Ventris (Ventris and Chadwick, 1953), an exhaustive structural analysis has been attempted here too, to understand the mechanisms used by Indus inscriptions to convey meanings. A contextual analysis of the inscriptions has also been performed, based on the excavation reports and corresponding analysis done by several leading archeologists. This article attempts to establish that the inscribed seals, sealings and tablets of IVC were "formalized data carriers" (a term coined by Nissen et al. (1993) to describe the proto-cuneiform administrative tablets of ancient Mesopotamia), which used both linguistic and non-linguistic (document-specific) syntaxes to convey their meanings. After analyzing the combinatorial patterns and graphemic features of Indus signs, and the type of objects in which certain signs frequently occurred, this study classifies several signs into nine functional sign-classes, and examines these sign-classes' roles in the phrase-structure of the inscriptions. This study also

formulates certain criteria to identify the lexeme-signs, probes the nature of the collocations and the repeated sign-sequences present in the script, and discusses the compositional semantics of Indus inscriptions. Finally, through analysis of the co-occurrence preferences and co-occurrence restriction patterns demonstrated by different sign-classes, and comparison of those patterns with the patterns permitted by "phonological co-occurrence restrictions" and "semantic co-occurrence restrictions" in natural languages, it strongly suggests that most of the Indus inscriptions were logographic in nature.

Materials and methods

Corpora and conventions. Mahadevan's digitized corpus of Indus inscriptions (Mahadevan, 1977), whose Input-Data-File (henceforth IDF-80), was further enhanced with provenance-and-iconography-related details in 1980, is the primary corpus used for generating all the statistics used in this study. However, the excellent corpus compiled by Wells and maintained by Fuls (Wells and Fuls, 2006), and Parpola's photographic corpora of inscribed objects (henceforth CISI) (Joshi and Parpola, 1987; Shah and Parpola, 1991; Parpola et al., 2010) are also used for accommodating certain inscriptions not included in IDF-80, and for visualizing the inscribed artifacts. All the serial numbers used here, for referring to the signs, artifacts, inscription-lines, and inscribed-sides, are sourced from IDF-80, if not mentioned otherwise.

Since save a few inscriptions that have radial or boustrophedon arrangements, most of the Indus inscriptions were read from-right-to-left, all the inscription-lines are represented here in a normalized right-to-left direction. For example, the inscription of seal-#1325 (Fig. 1a), originally engraved in left-to-right direction in intaglio, is shown in right-to-left direction in Fig. 1f, the way its impression on a sealing should be read. The inscribed-line sides and inscription-line are numbered as S1, S2 etc., and L1, L2 etc., whereas the only inscription-line present on a side, and the only inscribed-side of an object are referred to as L0, and S0, respectively (see Fig. 1f). Any doubtfully read sign is marked with an asterisk ("*"). However, before considering any such sign-sequence, I have doubly verified them using the other corpora mentioned above.

Methods. This article mainly focuses on understanding 'how' Indus inscriptions conveyed meanings, rather than on deciphering 'what' meanings were exactly conveyed. In the contextual analysis, archeological evidence, culled from the works of several leading archeologists, are used and studied in the light of the script-internal patterns, to establish that the inscribed seals, sealings and tablets were formalized data-carriers. For studying the syntactic structure of Indus inscriptions, a computer-assisted corpus analysis of IDF-80 is performed. For classifying the signs, I have used a manual feature extraction process, which focuses mainly on the signs' positional preferences, co-occurrence-preferences and co-occurrence restriction patterns. Occasionally, the feature-engineering process also considers the graphemic similarities between combinatorially similar signs, and the archeological contexts of the artifacts, where the signs mostly occur in. While analyzing the morphological characteristics and combinatorial patterns of repeated sign-sequences (⍣⍣, ⵌⵌⵌ etc.) and collocations (║ ⊍, ⍣║ etc.), certain universal linguistic rules that may have caused such patterns in linguistic texts, are considered.

Often multiple Indus objects contain identical inscriptions (see Fig. 1d). Thus, while analyzing a sign's combinatorial patterns, only the distinct inscription-lines (henceforth DILs) containing
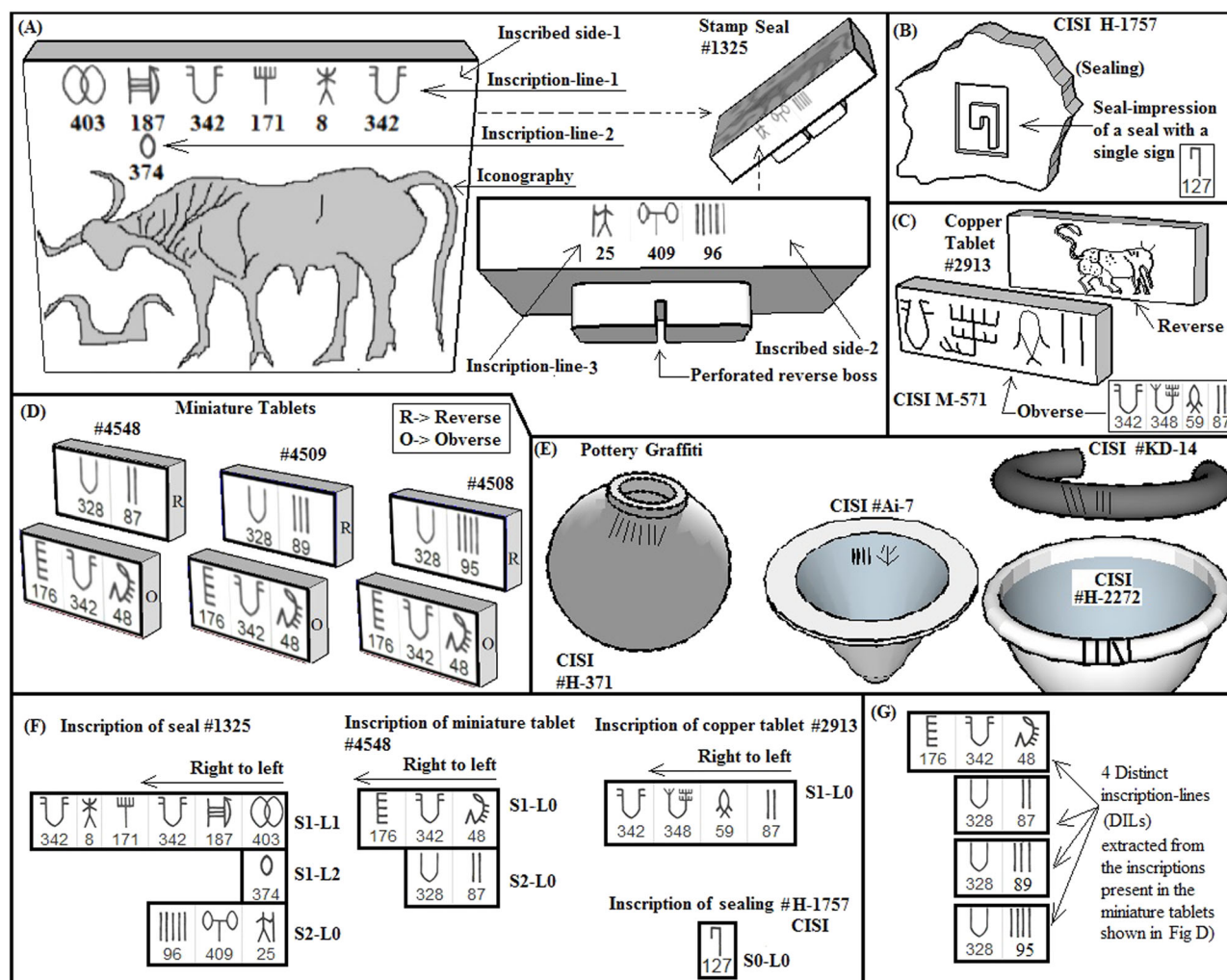
Fig. 1 Schematic representations of inscriptions found on different type of media. **a** An inscribed stamp seal of Mohenjo-daro; **b** An inscribed sealing of Harappa; **c** An inscribed copper-tablet of Mohenjo-daro; **d** Three inscribed miniature-tablets of Harappa; **e** Some inscribed pottery; **f** Representation of inscriptions using sign-numbers and side-line-numbers; **g** Distinct inscription-lines extracted from the miniature-tablets of **d**

the sign are considered (see Fig. 1g), so that repeated counting of a common inscription does not skew the statistics.

Since the corpus-analysis needs to exhaustively explore archeological and script-internal evidence, the essential points for each part of Section "Results and discussion" are included in the main text, while some part of the supportive detailed analysis is included in different sections of Supplementary-File Supp1.

Considering the typical brevity of Indus inscriptions (around 70% of the DILs contain only 1 to 5 signs), and the limited number of inscriptions found till date, this study prefers manual feature-engineering over unsupervised machine-learning algorithms, as the latter needs a much larger training set. Some of the Python programs used for searching the corpus are included in the supplementary-file Supp2.

Since ISC has been researched for nearly 140 years by several erudite scholars, an elaborate literature survey is provided in the supplementary file Supp3, which marks off the aspects of this study that are novel, from the ones that build on certain existing viewpoints, or differ from them.

Results and discussion

The results of this study are mostly obtained from two categories of analysis: (i) contextualization of the inscriptions using external evidence; and (ii) structural study of the inscriptions based on script-internal evidence. Both methods help us understand the mechanisms through which meanings were conveyed by these atypically brief inscriptions.

Contextualizing the formalized data-carriers of IVC. By the term "formalized data-carriers" I refer to any object that carries information of a specific semantic scope, and uses certain pre-defined structures for their message conveyance. Ability to identify certain inscribed objects of IVC as "formalized data-carriers" assumes significance because such data-carriers use both non-linguistic document-specific syntaxes and linguistic syntaxes in their written content. For example, as shown in Fig. 2a, different types of information (e.g., the name of the issuing country, the purpose of the data-carrier, or its denominational monetary value) are placed in different predefined positions in the stamps used in modern India. Similarly, the obverse and reverse sides of currency-coins contain different categories of information (Fig. 2a). But, all these syntactical rules are purely document-specific, and have nothing to do with the informing languages used in the objects. Yet, even in brief messages of formalized data-carriers, linguistic phrases often obey certain grammatical, language-specific syntaxes. For example, since English mainly uses pre-nominal adjectives, the stamps of Fig. 2b contains constructs where the numerical adjectives preceded the monetary nouns

Fig. 2 Schematic diagrams of certain formalized data-carriers of modern times. **a** Some coins and stamps of India; **b** Stamps of some countries whose informing languages use pre-positive adjectives; **c** Stamps of some countries whose informing languages use postpositive adjectives

(e.g., "One Anna", "3 Cents"). Moreover, the word-orders used in the genitives follow the "adjective-noun" order so that phrases like "United States Postage" are never constructed as "Postage United States". Contrarily, stamps of Fig. 2c, written in languages like Italian, Romanian and French, generally use post-nominal adjectives, giving us constructs like "Lire 30", "Lei Zece Mii", "Poste Vaticane" and 'Colonie Italiane', where the substantive words precede the attributive words.

Now, as elaborated in Section-S1 of supplementary-file Supp1, this study identifies the seals, sealings, and tablets of IVC as "formalized data-carriers" based on the following evidence: (i) miniature size and portable nature of the seals and tablets; (ii) fixed positioning maintained between inscriptions and their iconography (see Fig. 3b); (iii) fixed and formulaic structures used in the sign-sequences of the inscriptions (Fig. 3b); (iv) enormously expensive, regulated and painstaking processes used for manufacturing the durable and intricately made seals and tablets; and finally (v) standardized usage of identical and near-

identical inscriptions across distant Indus locations (as distant as 600–900 km) (see Fig. 4). Each of these features can be compared to the characteristics of various "formalized data carriers" of the present time, such as the revenue stamps, currency notes, or trade permits issued in a modern country, every one of which must: (i) be portable and durable in nature; (ii) have a specific pre-defined format for conveying their information; (iii) be made in a regulated way with intricate security features to resist counterfeiting; and (iv) be used in a standardized manner in every distant part of a country.

This study further claims that, the domain of usage of these formalized data-carriers was mainly some commercial activities and related administrative activities of IVC in which metrology and standardization played crucial roles. The archeological evidence that leads to this claim are: (i) the inscribed seals and tablets were almost always found concentrated near craft areas, such as bead and shell workshops, or near fortified city gates where traded goods were supposedly measured and taxed; (ii) the
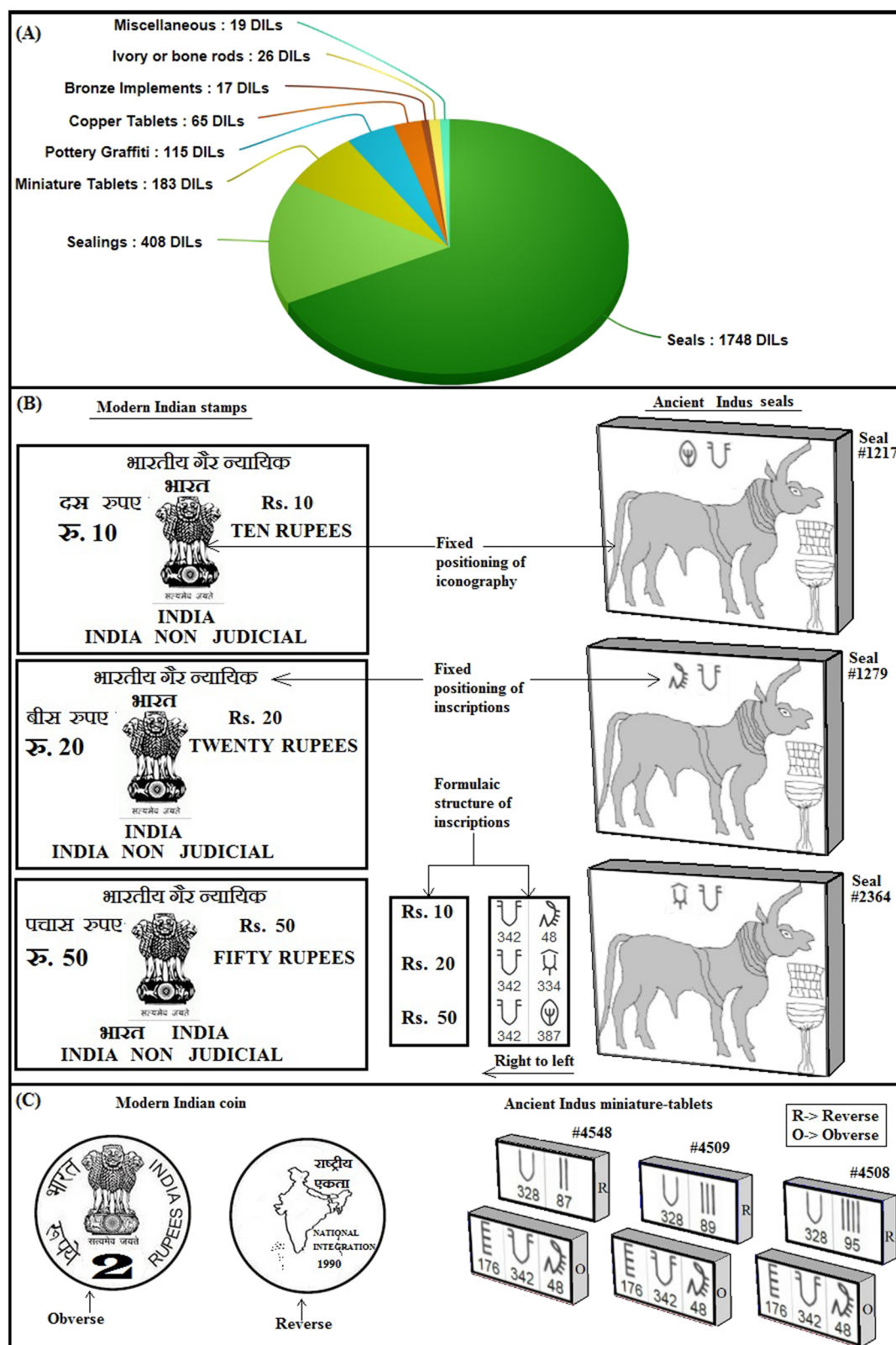
Fig. 3 Distribution of Indus inscriptions across artifact-types **a**, and their structural similarities with the structures found in modern data-carriers **b**, **c**

seals were quite often found along with standardized weights of IVC; (iii) seal impressions were found in clay-tags attached to packed merchandises; (iv) inscribed objects were rarely found in religious contexts such as in grave goods; and (v) they were often discarded as trash after use (Kenoyer, 2010; Parpola, 1994; Bhan, 2011; Wells, 2011; Possehl, 2002; Kenoyer, 2005; Mackay, 1931). Section-S1 of supplementary-file Supp1 provides further details.

This contextualization enlightens the subsequent structural analysis, when we proceed to distinguish document-specific
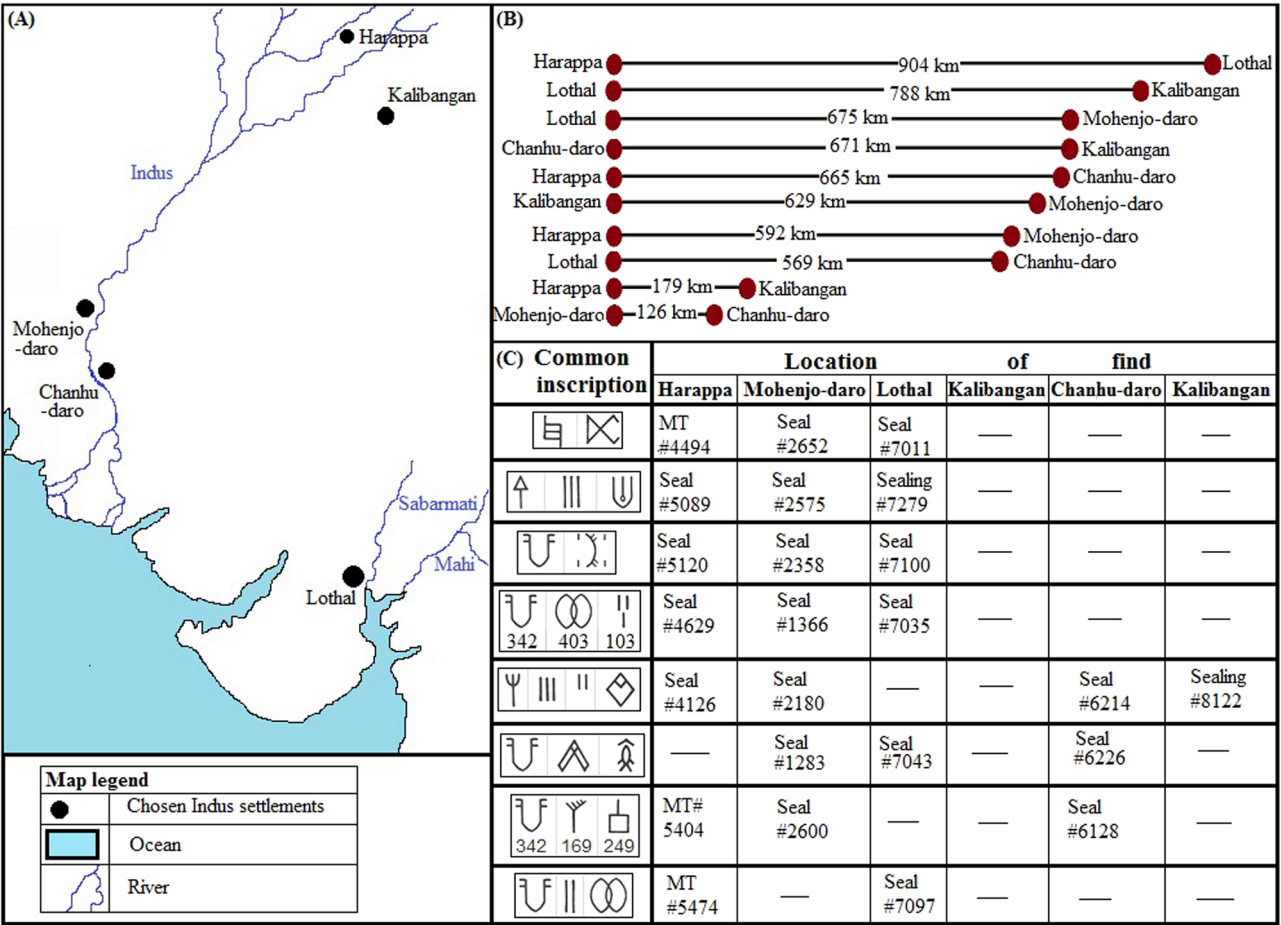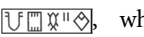
Fig. 4 Identical Indus inscriptions found in distant Indus locations. **a** A map showing selected Indus locations (adapted from a map included in Mahadevan's corpus (1977, p. 29)); **b** Visual comparison of pair-wise distances between Indus locations (all distance-calculations are taken from Possehl, 1999, cited in Yadav, 2013.); **c** A grid listing the common inscriptions found in different locations

syntaxes used in Indus inscriptions from syntaxes more likely to be language-driven.

Structural analysis of Indus inscriptions. This structural analysis hinges on the basic postulation that the complete inscription content of each unbroken seal, sealing and tablet (independent to the iconographies, if any) was a semantically complete message, complete with respect to the context and the purpose designated to it. As discussed above, the standardized use of these pains-takingly made artifacts strongly indicate that their formulaic inscriptions cannot be just random scribbles or decorative designs. They must have conveyed complete, meaningful messages of great importance to the people of IVC. However, the inscriptions found on potshards, and the inscriptions of the artifacts which were broken or damaged so irretrievably that some part of their inscription-contents are no more readable, are excluded from the list of semantically complete messages.

Even though this postulation sounds obvious and platitudinous, its methodical pursuit helps immensely in the structural analysis, especially in identifying various Indus signs as logograms (see Section "Identification of lexeme-signs based on semantic completeness of inscriptional units").

Since 79% of the 2906 inscribed objects of IDF-80 could convey their messages using only single inscription-lines, I treat all inscription-lines as semantically complete phrases, and use them as the basic units of this structural analysis. The only exceptions might be the 'split-sequences' (Mahadevan, 1977, p. 12), where

possibly due to space-constraints faced by scribes, a continuous sign-sequence was sometimes split between more than one inscription-lines carved on the same sides of the artifacts. For example, the inscription-line ⊔▦𝕏"◇, which carries the complete message of seal #2618, got split into the sequences "◇ and ⊔▦𝕏 that occur in separate lines on the same inscribed-side on seal #6112.

Sections "Identification of lexeme-signs based on semantic completeness of inscriptional units"–"Unclassified Indus signs and their probability of being lexemes" are mainly dedicated to classifying the functionally similar Indus signs into separate sign-classes, exploring the roles of the sign-classes in Indus phrase-structure, and formulating certain criteria for identifying the lexeme-signs.

Identification of lexeme-signs based on semantic completeness of inscriptional units. Using the basic postulation stated above, we can straightaway recognize more than 10% of the Indus signs (49 signs) as lexemes (LEX-signs), based on the four simple criteria stated below (see Fig. 5a–d).

Criterion-1. Aloneness in an object: Since the total inscription-content of a complete and unbroken seal, sealing or tablet must be a semantically complete message, if a single sign occurs as the total inscription content of such an inscribed object (Fig. 5a-Column1), that sign must contain some lexical meaning on its own, as it single-handedly expresses the complete message of that object (Fig. 5a-Column2 identifies 28 LEX-signs using Criterion-1).
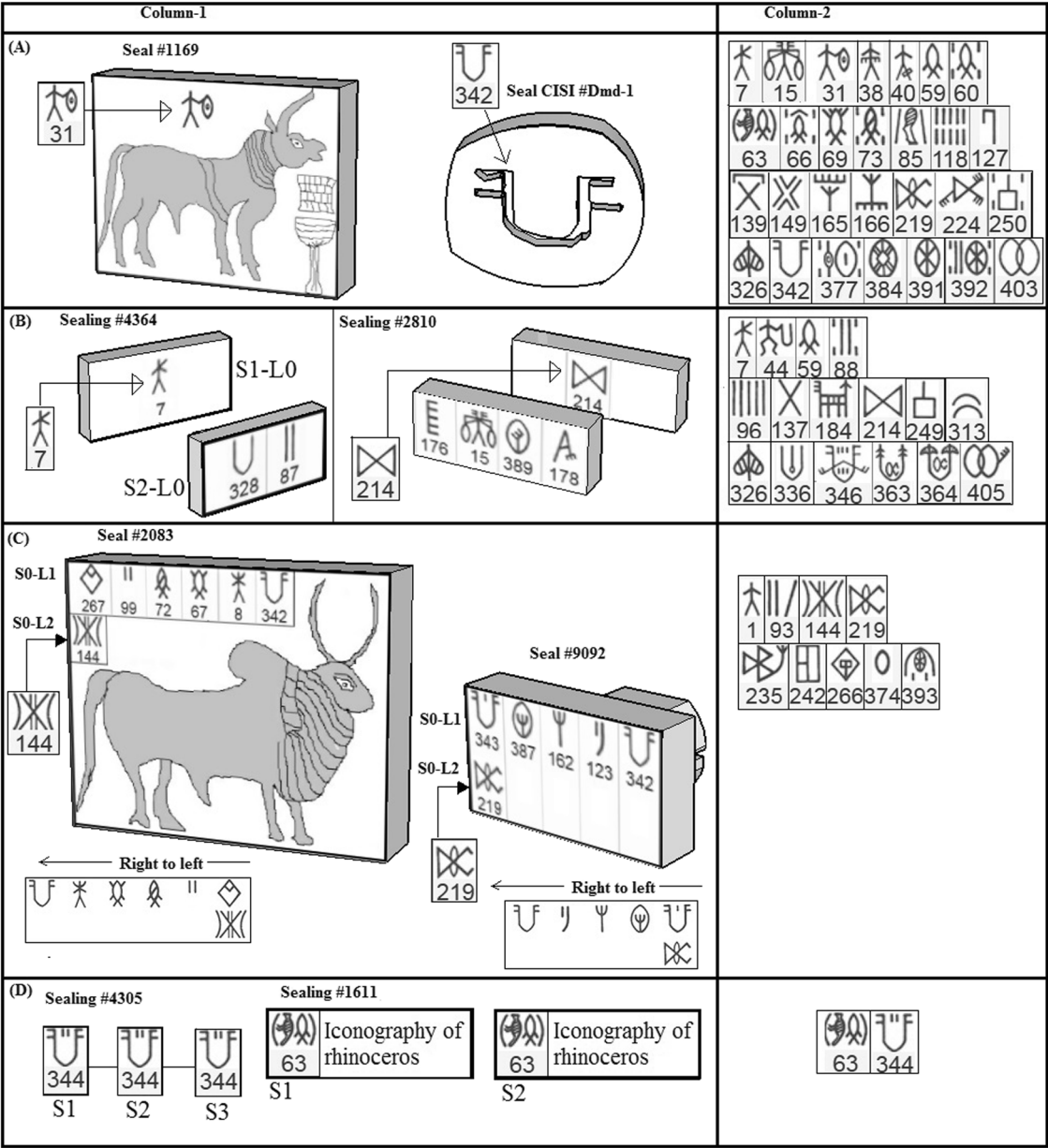
Fig. 5 Certain isolable occurrences of Indus signs. Single signs occurring alone: in an object **a**, in an inscribed side **b**, in the last inscription-line of an inscribed side **c**, and in all sides of the same object **d**. Column-1 provides representative examples of each scenario. Column-2 provides the list of signs identified using the corresponding scenario

To preclude inclusion of casually scribbled signs, or decorative symbols as lexeme-signs, Criteria 1–4 exclusively considers inscriptions found in seals, sealings, and tablets (formalized data-carriers) only. The inscriptions of pottery, bangles, bronze-implements etc. are not considered.

Criterion-2. Solitariness in an inscribed side: If a sign occurs alone in an inscribed side of an unbroken artifact and the inscriptions on the other sides are recognizable as semantically complete phrases, then this single sign cannot be a syntactic continuation of the inscription on other sides, and must hence be a semantically autonomous lexeme-sign (Fig. 5b demonstrates 16 LEX-signs, identified using Criterion-2).

As discussed in later sections, semantically complete phrases are recognizable if they end with well-known terminal-signs, such as phrase-final signs, crop-like signs, and encapsulated signs, or

occur independently in other inscribed objects. Since Indus scribes seldom split inscription-lines on different sides, each inscribed-side of a seal/tablet functions as a syntactic boundary (Parpola, 1994). Moreover, as discussed before, generally the obverse-side and reverse-side inscriptions demonstrate clearly distinct semantic scopes. All these facts reinforce Criterion-2.

Criterion-3. Aloneness in a syntactically isolated inscription-line: If in an artifact with multiple inscription-lines on a single side, the last inscription-line contains a single sign, while the penultimate inscription-line is recognizable as a semantically complete inscription, then that single sign of the last line should be a semantically autonomous lexeme-sign (Fig. 5c-Column1). Nine such signs are identified as lexemes (Fig. 5c-Column2).

Criterion-4. Single sign repeated on all sides of an object: If each inscribed-side of a sealing or tablet contains a single sign,
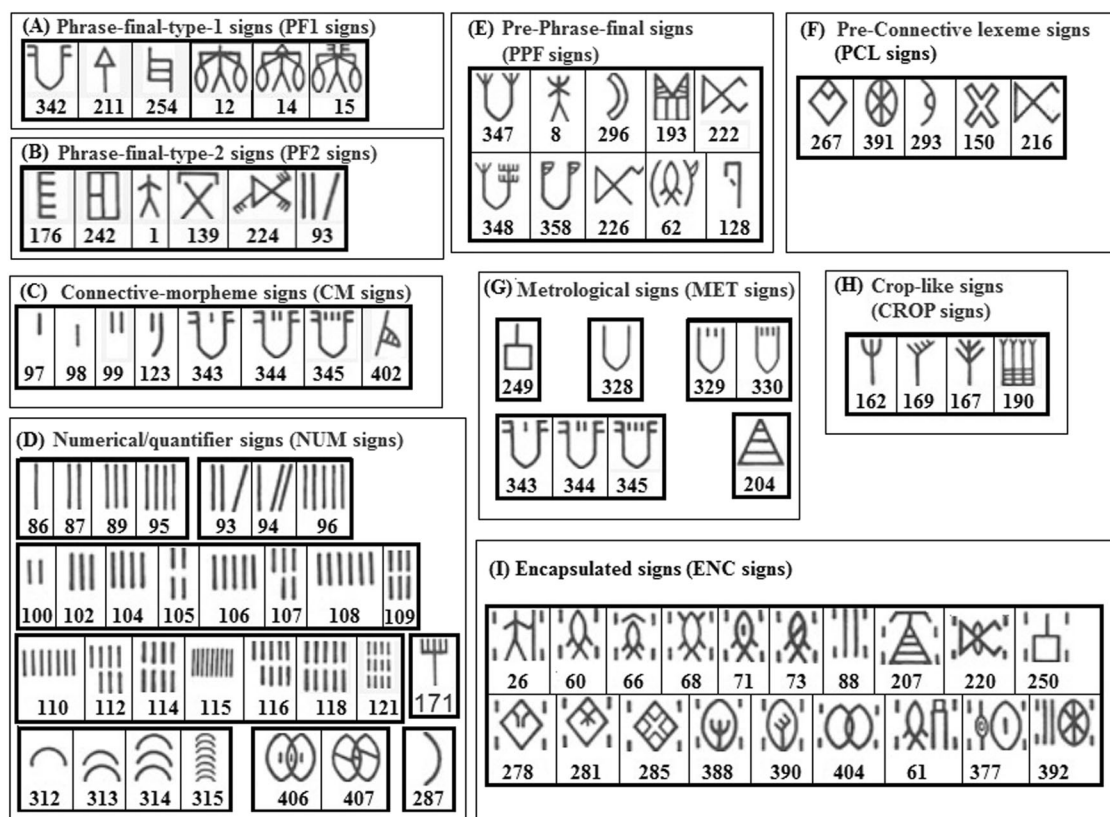
Fig. 6 Indus sign-classes along with the list of their member signs

then the meaning of that sign must be the complete message conveyed by that object, and that sign must be a semantically autonomous logogram. As shown in Fig. 5d, two such logograms are identified.

Interestingly, signs like [sign], [sign], and [sign] are identified as logograms by more than one criteria stated above.

All the sign-classes identified in the sections below, are listed together in Fig. 6, for convenience of readers.

Phrase-final (PF) signs and their subcategories. Excluding the 89 DILs which contain only single signs, and 323 DILs whose terminal signs are irretrievably lost, 1998 DILs of IDF-80 remain. Among these, in 1293 DILs (i.e., 65%), the terminal signs belong only to the very selective set of the 12 signs listed in Fig. 6a, b. Since Indus inscription-lines are semantically complete phrases, and these 12 signs predominantly occur in the final positions in the majority of such phrases irrespective of their inscription-lengths (see Fig. 7a, b), they are classified as "phrase-final" (PF) signs. The PF-signs are further categorized as phrase-final-type-1 (PF1) and Phrase-final-type2 (PF2) signs, since the signs classified as PF2s ([signs]) predominantly follow the signs classified as PF1s ([signs]), thus constructing <PF1 PF2> sequences (PF-clusters) such as [signs], etc.

It is important to note that terminal occurrences in a few inscriptions alone cannot entitle a sign to be classified as a PF, since one of the main classification criteria is the high percentage of terminal occurrences. For example, considering only completely legible DILs, such percentages for the PF1-signs are: [sign] (90.8%), [sign] (88.75%), [sign] (87.1%), [sign] (89.4%), [sign] (100%), and [sign] (90.8%).

Another important classifying feature for a PF-sign is that it should be syntactically and semantically detachable from its preceding sign-sequence, playing a phrase-level syntactic role, rather than being a semantic continuation of its preceding sequence. For example, the bigram [signs], which repeatedly got followed by PF1 [sign], could independently occur in sealing #4823 (Fig. 8a), proving the detachability of [sign] from its preceding sequence. Similarly, comparing the complete inscription-lines [signs], [signs] and [sign], [sign], we realize the detachability of PF2-signs (Fig. 8b). These patterns prove that the phrase-level PF-signs are not integral parts of the meanings conveyed by their preceding sequences. Thus, the phytomorphic signs ([sign], [sign] etc.) which occur quite frequently in terminal positions, are not classified as PF-signs (see Section "Crop-like signs (CROP-signs)"), as they demonstrate a strong affinity towards their preceding stroke-signs (e.g., [signs] etc.), rather than playing a generic and detachable phrase-level role.

Though in some inscription-lines the PF-signs occur in apparently medial positions, as demonstrated in Fig. 8c, those inscription-lines actually comprise multiple shorter juxtaposed semantic messages, and the medial positions of the PFs are actually phrase-final positions of the semantically complete constituents.

Since certain PF1-signs ([sign], [sign]) and PF2-signs ([sign], [sign], [sign]) are already identified as lexemes (see Fig. 5), the other PF1s and PF2s must also be lexemes, considering their functional homogeneity (Fig. 9a, b, respectively, show how the PF1-signs, and PF2-signs often occur in mutually similar inscriptional-contexts).

Figure 9c imparts an important insight regarding the functional nature of PF-signs, showing how signs with similar graphemes show a strong affinity towards specific PF-signs. For example, in constructs like [signs] etc., the pincer-like signs are typically followed by PF1 [sign], whereas the fish-like-signs usually show a special affinity for PF1-sign [sign]. Now, in a script where lexemes are represented by individual signs (Section "Identification of lexeme-signs based on semantic completeness
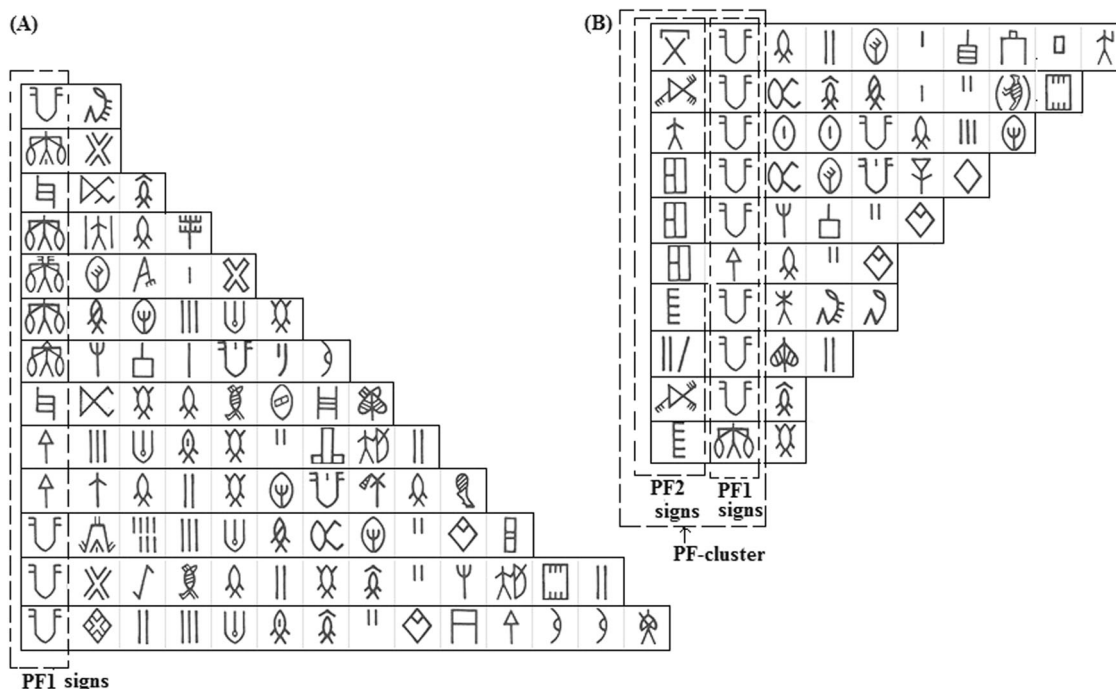
Fig. 7 Phrase-final occurrences of signs in inscription-lines. **a** Terminal occurrences of PF1-signs in inscription-lines of different lengths. **b** Terminal occurrences of PF1-signs followed by PF2-signs
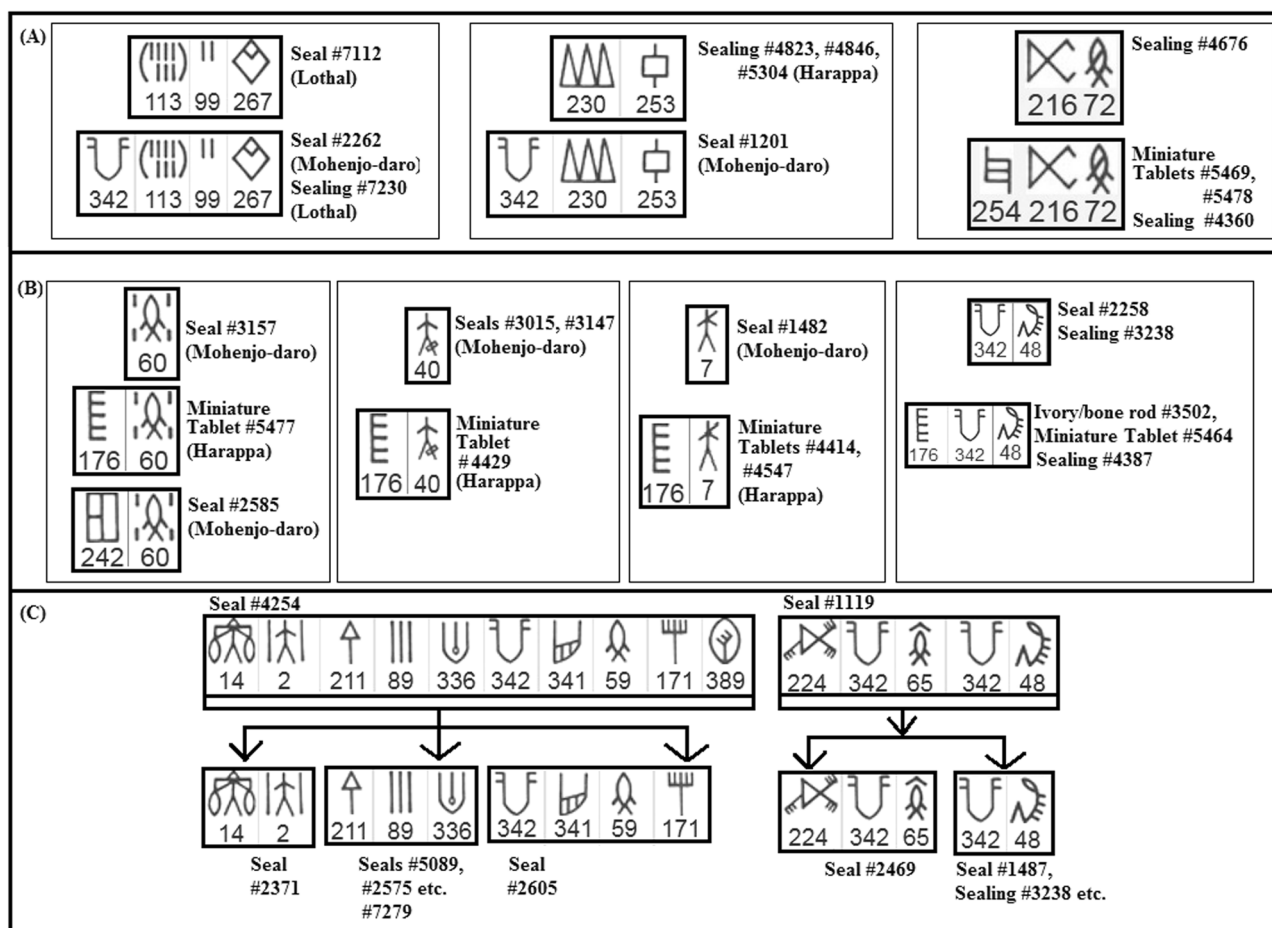

Fig. 8 Certain combinatorial characteristics of PF-signs. **a**, **b** Examples where the preceding sign-sequences of PF1-signs and PF2-signs occur independently in other objects. **c** Pseudo-medial positions of PF-signs
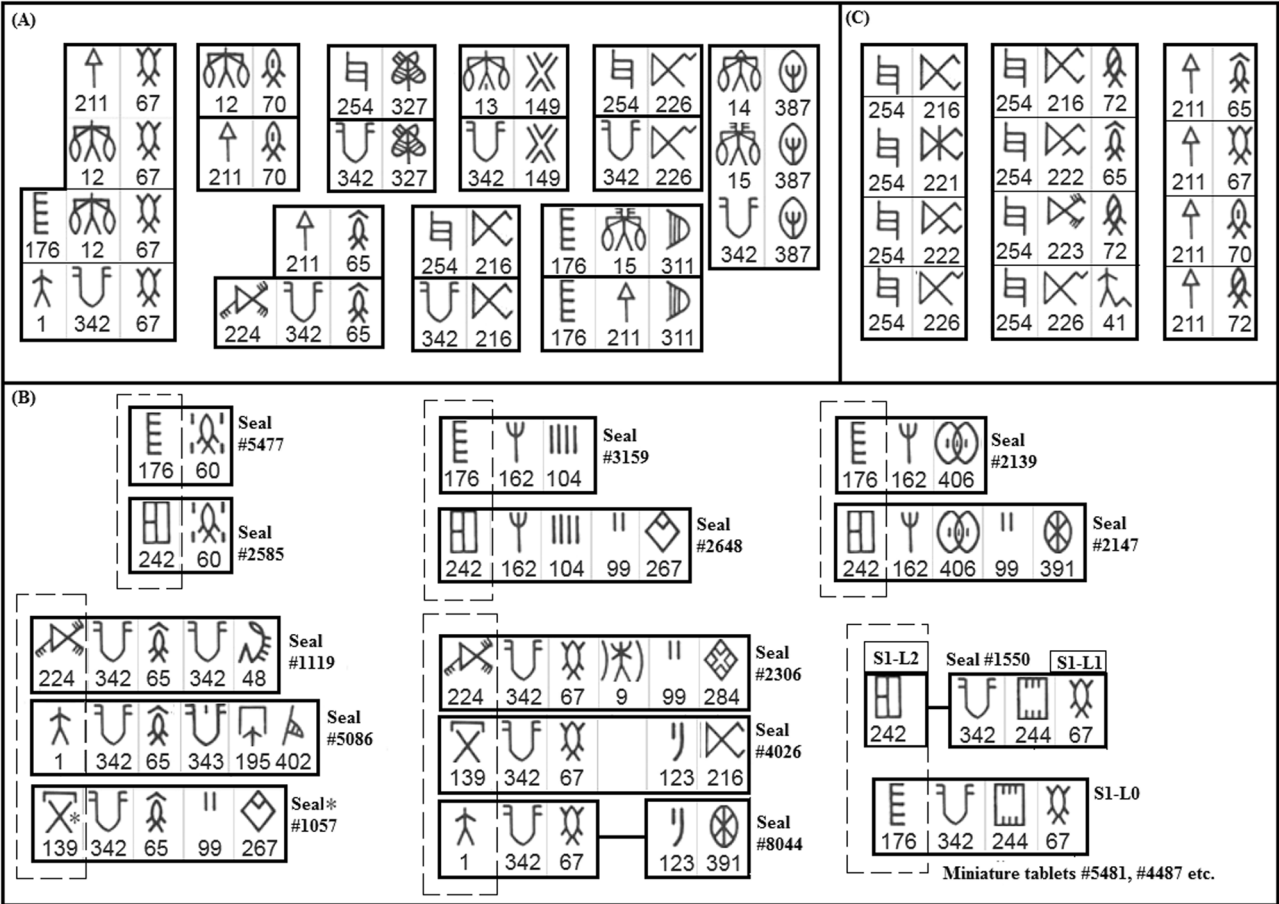
Fig. 9 Certain other combinatorial features of PF-signs. **a, b** Different PF1 and PF2 signs used in similar inscriptional-contexts. **c** Graphemic similarities between the signs that have special affinity towards specific PF1-signs

of inscriptional units"), choice of similar graphemes for a group of signs should logically be governed by the similar semantic meanings of those signs, proving that certain PF-signs were functionally more suitable to certain semantic groups.

Some other interesting observations, such as the affinity between specific PF1 and PF2 signs and the patterned occurrences where PF2-signs either appear without PF1-signs, or occur preceding them, are discussed in Section-S2 of supplementary-file Supp1.

Pre-phrase-final signs (PPFs). Although a large number of signs have regularly preceded PF1s in different inscriptions, a specific few show remarkable affinity towards the PF1s, so that they get positioned immediately before PF1s in almost all their occurrences (see Table 1). This study classifies 10 such signs (Fig. 6e) as pre-phrase-final (PPF) signs.

An intriguing feature of PPFs (also used as their identification criterion) is that, if an inscription contains a PPF-sign, it is extremely unlikely that the inscription's PF1-sign would follow any sign other than the PPF. For example, in inscription-lines and , sign directly preceded . But whenever the PPFs and occur in similar inscriptions (e.g., or), sign gets distanced from , as the PPFs get the preference to occur immediately before . Figure 10a, b contains more such examples.

Figure 10c demonstrates another interesting fact about the PPF-sign . As observed by Mahadevan (1986), sign , which

Table 1 The percentage of occurrences of pre-phrase-final signs in positions that immediately precede the PF1 signs in inscriptions

Sign	Count of Pre-phrase-final occurrences in DILs	Count of Total occurrences in DILs	Percentage of Pre-phrase-final occurrences
	88 (8*)	95 (4*)	92.63%
	76 (1*)	86 (3*)	88.37%
	23	26	88.46%
	19	21	90.48%
	19	19	100.00%
	18 (1*)	20 (2*)	90.00%
	13	14	92.86%
	10	11	90.90%
	8	9 (1*)	88.89%
	6	8 (1*)	75.00%

The bracketed counts marked with "*" denote the count of instances containing "doubtfully read" PPF signs

occurred in only one DIL preceding the PF2-sign , is graphically constituted by combining the graphemes of PPF-sign and PF1-sign . This study argues that the construction of this ligature confirms the semantic affinity shared by the PPF and PF1 signs.

Identification of more logograms in relation to the PF-signs. The phrase-level syntactic roles played by the PF-signs, their
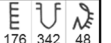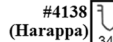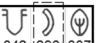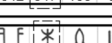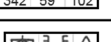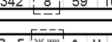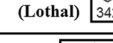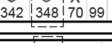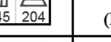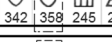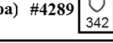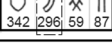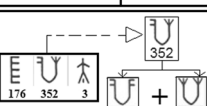
Fig. 10 Combinatorial and graphemic characteristics of PPFs. **a** Inscription-lines ending with PF1-signs and not containing any PPF-sign. **b** Inscription-lines containing the same sign-sequences present in **a**, with an extra PPF-sign positioned immediately before the PF1. **c** A ligature sign constituted of conjoined PPF and PF1 signs

occurrences as the boundaries of the semantically complete phrases, and their syntactic and semantic detachability from the preceding sign-sequences, help me propose the following criteria for identifying lexeme signs.

Criterion-5: Solitary occurrences after PF1-signs: Since PF1-signs and PF-clusters denote the boundaries of semantically complete messages, a sign occurring alone after such a sequence in a complete inscription-line (Fig. 11a) cannot be an indispensable part of its preceding sign-sequence, and must be a lexeme-sign that signifies some meaning on its own. Figure 11a lists 22 such lexeme-signs.
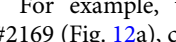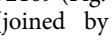
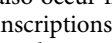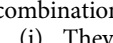**Criterion-6: Signs occurring alone with PF1-signs or PF-clusters:** Here I make a crucial proposition that if any sign occurs alone with a PF1-sign or a PF-cluster in a complete inscription-line (Fig. 11b) at least once, that sign must be a lexeme-sign. The lexemes identified by Criterion-6 are henceforth referred to as Alone-With-Phrase-Final (AWPF) lexemes.

The basis of this proposition is primarily the type of functional roles played by the PF1-signs. The rigid preference for terminal positions makes PF1s the most predictable part of Indus inscription-lines. From Shannon's theory of self-information (Shannon, 1948) and its applications in semiotics (Floridi, 2015), we know that in any semantic domain, the information content of a message is inversely proportional to its expected value or probability. Thus in the semantic domain of Indus inscriptions, the information content of the highly predictable PF1-signs must have been restricted to a specific and limited semantic range. Moreover, since as many as 1293 distinct inscription-lines needed

PF1-signs as part of their messages, the semantic role of PF1s must have been associated with the general usage of those inscriptions. So, given the generic type of meanings conveyed by PF1-signs, obviously the remaining parts of the inscription-lines had to convey the main information that semantically differentiated the message of one seal from the other. For example, in the modern stamps of Fig. 2a, the country-name, stamp-type, and monetary-units are generic common information, whereas the denominational monetary values expressed by the numeral nouns (Two, Four, Eight, Ten etc.) carry the main information that semantically distinguishes one stamp from the other. Similarly, in the inscription-sets in Fig. 11b (e.g. 目�, 目�, 目�, and 目�), different single signs occur alone with the same PF-sign. Therefore, these inscriptions' messages semantically differed only by these single signs preceding the PFs, indicating that they must be semantically autonomous lexemes.

Seventy-nine distinct signs are identified as AWPF lexemes by using Criterion-6 (Fig. 11b-Column2). Considering that all the PPF-signs have occurred at least once as an AWPF sign, I claim that all the PPF-signs were lexeme-signs.

Connective-morphemes (CMs) and composite inscriptions. Connective-morphemes (CMs) are possibly the most important entities in the structural compositionality of Indus inscriptions. In case of natural languages, the term "connective" is generally defined as "conjunctions, prepositions, adverbs and other particles which share the function of encoding semantic relations between sentences, or rather, between semantic objects, some of which can be meanings of sentences" (Blühdorn, 2010). Interestingly, many inscription-lines of ISC can be represented in the form <X CM Y>, where X and Y are either single-sign lexemes or semantically complete sign-sequences that have occurred independently as meaningful inscription-lines in one or more inscribed objects (see Fig. 12).

For example, the inscription-line 囲 ↑ ↑ 〈 ↗ Ɗ ∪ of seal #2169 (Fig. 12a), comprises smaller constituents Ɗ ∪ and ↑ ↑ 〈 (joined by CM-sign ↗), which are semantically complete messages of seals #6109 and #1225, respectively. Generally the signs that occur in between such semantically complete messages, belong to a very selective set (∥, ﾘ, ↗, ∪, ∪, ∪, ﹐, and ┆). Interestingly, often their combinations (∪ﾘ, ∪ﾘ, ∪ﾘ, ﾘ↗, ∥↗, ┆↗, ﹐∪, ∪∥, ﹐ﾘ) also occur in between the autonomous constituents of composite inscriptions. The reasons for which these signs and their combinations are classified as connective-morphemes are that:

(i) They form clearly visible juncture points in Indus inscriptions (elaborated below).

(ii) The signs ∥, ┆, ﹐, and ﾘ, almost never occur alone in inscriptions, and rarely occur in any isolable position that characterize the lexeme-signs.

(iii) Some of these signs, i.e., ∥, ┆, and ﹐, are much smaller in size (almost one-third) than all the other full-length Indus signs, indicating a graphemic feature consciously chosen by Indus scribes, to make their juncture-property more visual.

To demonstrate the obvious juncture-property of these signs, let us analyze the 88 composite inscriptions listed in Fig. 13. Almost all of the Y-constituents of these DILs are semantically complete messages that occurred independently in other objects, and terminated with either PF-signs or CROP-signs (see Section "Crop-like signs (CROP-signs)" for CROP-signs). On the other hand, the X-parts consisted of single signs belonging to a very selective set (◇, ⊗, 〉, ⊠ and ⊠). Thus, the semantic-scopes of the X and Y parts clearly differed, and quite arguably the semantic role of the signs positioned at the juncture points was that of connective-morphemes.
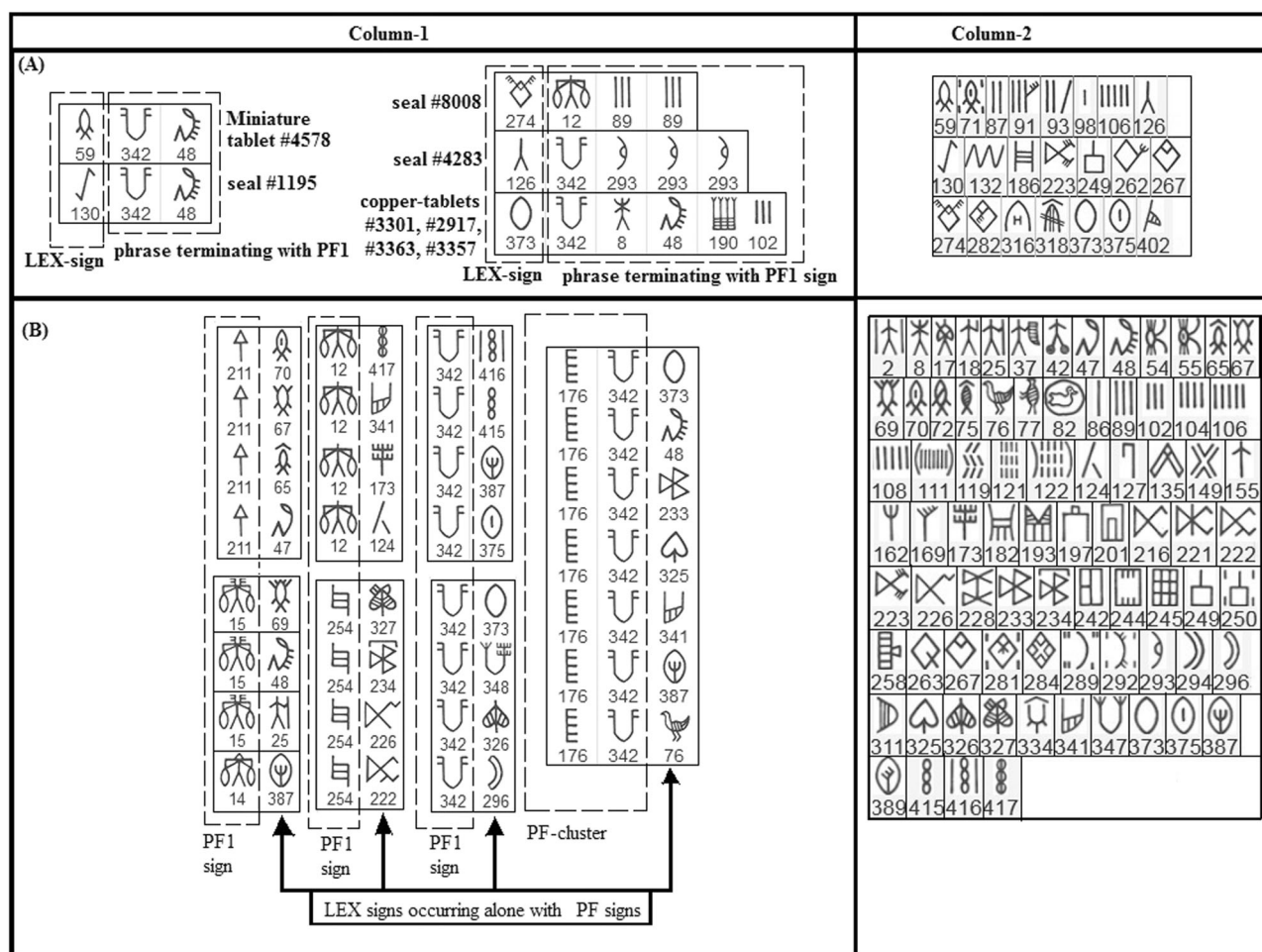
Fig. 11 Identification of logograms based on their associations with PF-signs. **a** Signs occurring alone after PFs, **b** Signs occurring alone with PFs. Column-1 contains representative examples of the patterns. Column-2 lists all the signs that have occurred in such patterns

It is noteworthy that the full-length stroked-jar CM-signs (⨃,⨃,⨃) are also classified as metrological signs in Section "Metrological signs (METs)", since in certain inscriptions their syntactic and contextual behaviors are clearly comparable to the behaviors of numerical signs (see Section "Numerical signs (NUMs)"). However, if we compare some similar DILs such as ⟦⟧, ⟦⟧, and ⟦⟧, it becomes evident that just like CMs ⟩ and ⟨, ⨃ too has connected the usual pre-connective sign ◇ with the usual post-connective phrase ⟦⟧. This implicit connective role must have been associated with the semantic role these metrological signs have played in such inscriptions. Figure 14a shows more examples where both single and composite CM-signs occur in similar inscriptional contexts, demonstrating their functional homogeneity.

Another notable feature of CM-signs is that they were syntactically bound to the pre-connective sequences, not the post-connective ones. As shown in Fig. 14b, whenever a composite inscription got split between two different lines, the "X-parts" followed by the CM-signs always remained in the first line, whereas the "Y-parts" got transferred to the second line. Thus, the pre-connective parts were obviously more strongly attached to the CM-signs, as otherwise at least in some instances the scribes would have kept the "CM Y" parts together.

Semantic completeness and distinct semantic scopes of the pre-connective and post-connective sequences. Generalizing the observations made in Section "Connective-morphemes (CMs) and composite inscriptions", I shall now propose that the pre-connective and post-connective constituents of any composite inscription-line constructed by using CM-signs must be semantically complete. So, any sign sequence that occurs as a complete pre-connective or post-connective constituent at least once, can be labeled as a semantically complete phrase with certainty.

Semantic completeness of pre-connective constituents: (i) Pre-connectives present in composite inscriptions of 650 inscribed objects contain only single signs. There are 92 such signs that have occurred alone as pre-connectives at least once. Now, since many of these pre-connective signs are already identified as lexemes using other criteria (e.g., signs ◇,⟨,⨉,⟩,⨅,⨆ etc., identified employing Criterion-1 (Fig. 5a); and signs ◇,⨉,⟩ etc., identified employing Criterion-6 (Fig. 11b), it is quite logical to expect that the other functionally homogeneous pre-connective single signs were also lexemes.

(ii) As shown in Fig. 15a, the pre-connective constituents of the obverse-side composite inscriptions of seals #5056 and #2626, also got repeated in their reverse bosses. Most possibly, these pre-connectives indicated the purpose or category of these seals and were present on the reverse boss as mnemonics or colophons so that the scribes could choose the right seal when several stamp-seals were kept in their stable upside-down position. Whatever be the reason for such occurrences, by separating out the pre-connective constituents on the reverse boss, Indus scribes have left an unmistakable clue about their semantic completeness.
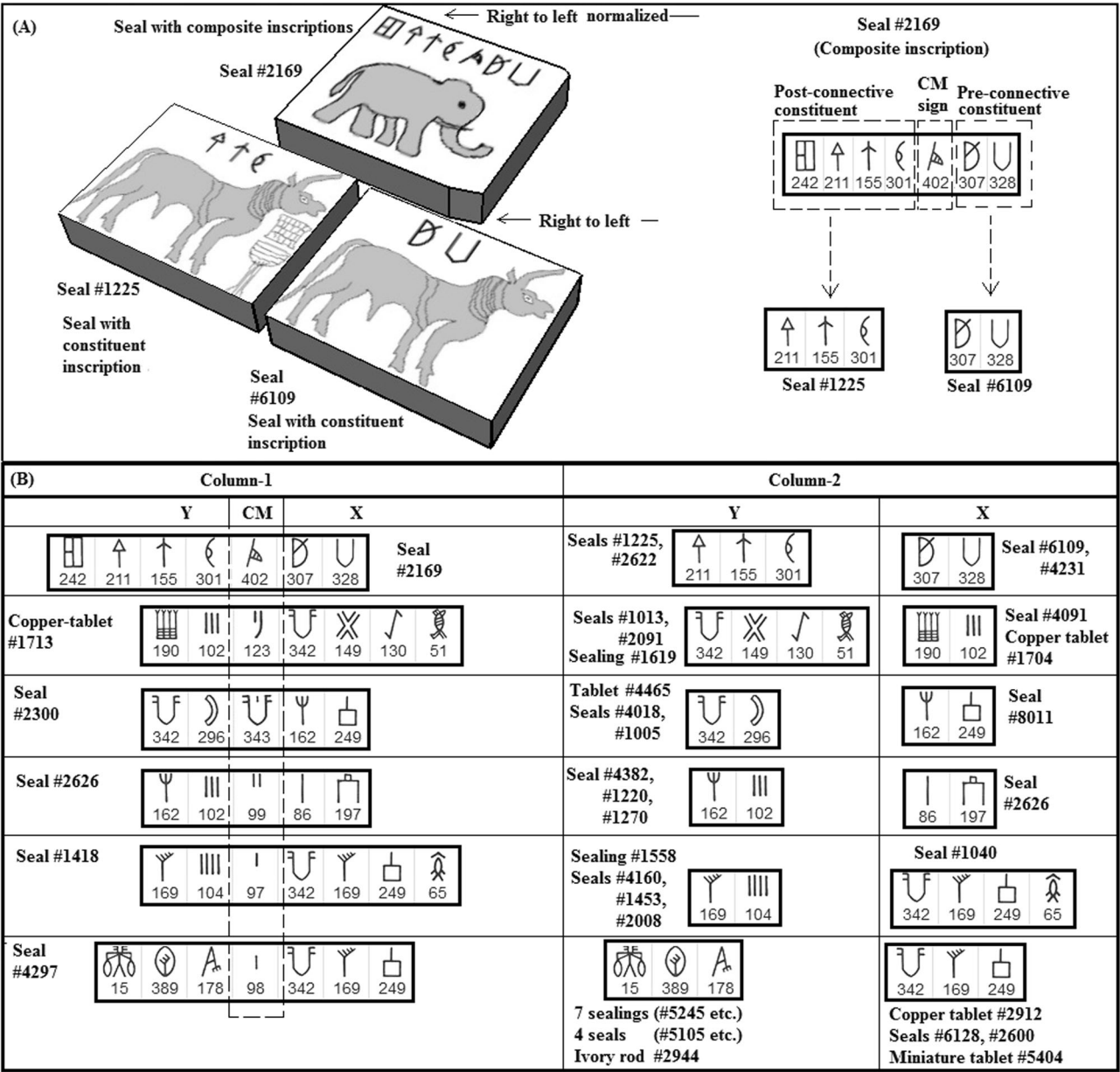
Fig. 12 CM-signs conjoining semantically complete constituents to form composite inscriptions. **a** Schematic diagrams of seals containing a composite inscription and its shorter constituents. **b** More examples of composite inscriptions (column-1) and their constituents (column-2)

(iii) Sometimes the sign-sequence occurring as the pre-connective of one inscription gets combined, to occur as a ligature in the pre-connective of some other inscription (e.g., compare the inscriptions [symbols], [symbols], [symbols], and [symbols] of seals #7107, #5091, #1048, and #2340, respectively). This pattern indicates the cohesive bonding present between the pre-connective sign-sequences that led to the formation of such ligature units.

(iv) Often, the pre-connective constituent of one inscription occurs in the post-connective part of another inscription (see Fig. 15b), proving that the pre-connectives were semantic units that occurred unaltered in different inscriptional contexts. Moreover, often the pre-connectives consisted of frequent collocations of IDF-80 (e.g., [symbols] (101 DILs), [symbols] (43 DILs) etc.).

All these patterns establish that pre-connective parts of composite inscriptions always held semantically complete phrases.

Semantic completeness of post-connective sequences: By searching IDF-80 programmatically, 135 distinct inscription-lines are found that have also occurred as post-connective constituents in 260 composite DILs. For example, the inscription-line [symbols], which occurs alone in 4 seals, one sealing and three pottery-shards, also occurs as post-connective in inscriptions [symbols] and [symbols]. This evidence alone satisfactorily indicates the semantic completeness of post-connective inscriptions.

Distinct semantic scopes: Sign usages of the pre-connectives visibly differ from that of the post-connective constituents. Signs that frequently occur in pre-connective positions (e.g., [symbols] etc.) occur with much less frequency as post-connectives. Moreover, when the signs [symbol] and [symbol] occur in the terminal positions of post-connective constituents (e.g., [symbols], [symbols], [symbols], [symbols], [symbols], [symbols] etc.) and non-composite inscription-lines (e.g., [symbols], [symbols], [symbols] etc.), they are mostly preceded either by stroke-signs, or by signs like [symbol], [symbol], or [symbol]. But in pre-
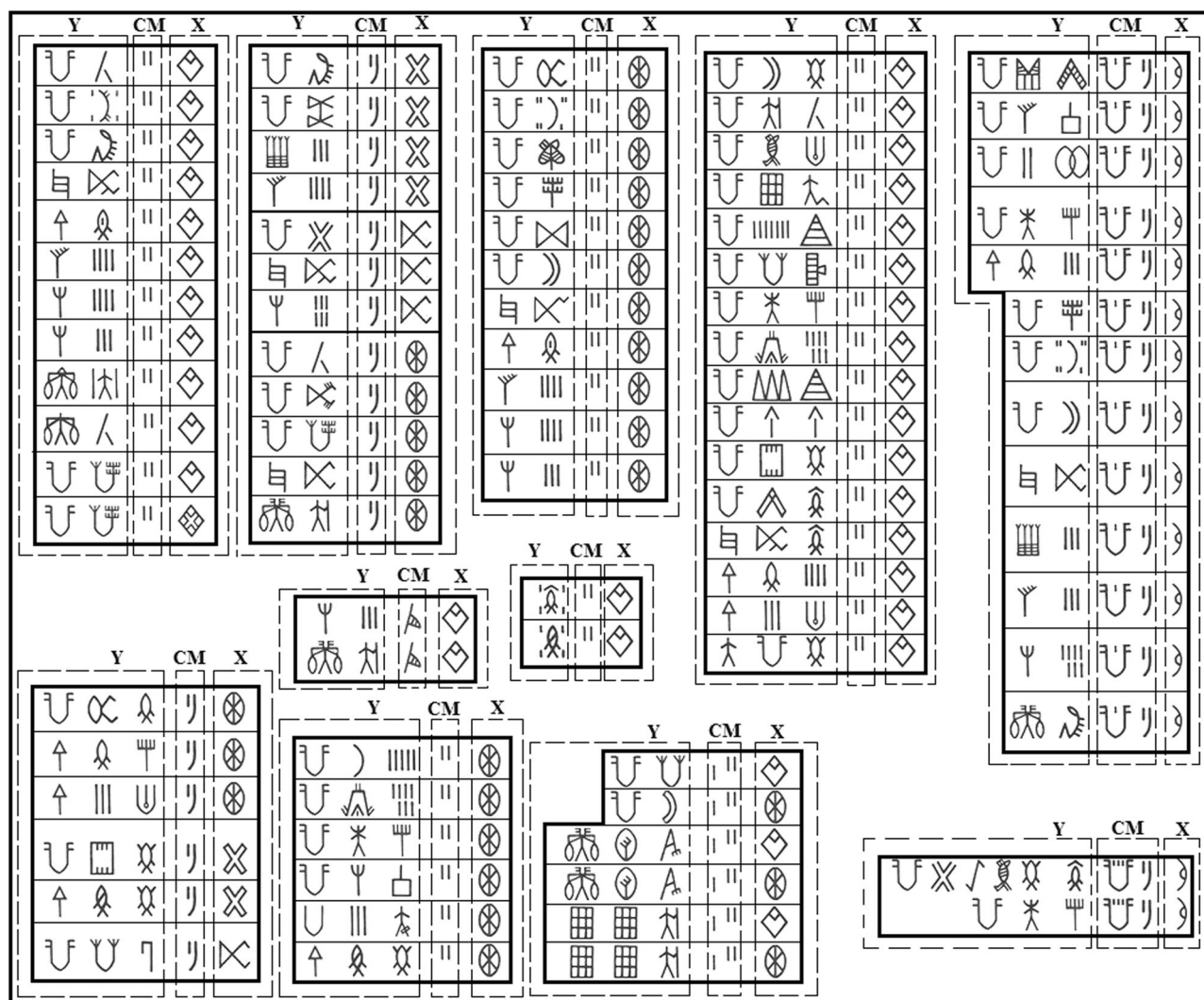
Fig. 13 More examples of composite inscriptions of the "X CM Y" pattern

connective positions (e.g., [signs], [signs], [signs] etc.), [sign] and [sign] generally follow a different set of signs ([signs], [sign], [sign] etc.), stroke-signs rarely preceding them in such contexts. Moreover, unlike the post-connective constituents that frequently end with PFs and ENCs (see Section "Encapsulated (ENC) signs"), typical pre-connective constituents rarely contain such terminal signs. Accordingly, the type of information content of the pre-connective and post-connective sequences undoubtedly differs.

Lexeme-signs based on the semantic completeness of pre-connective and post-connective constituents. **Criterion-7:** Since all pre-connective constituents are semantically complete phrases, a sign occurring alone as the pre-connective constituent of a composite inscription (see Fig. 16a-Column1), is identifiable as a logogram. Figure 16a-Column2 lists 92 such lexeme-signs.

Criterion-8: Since all post-connective constituents are semantically complete phrases, a sign occurring solitarily as a post-connective constituent (Fig. 16b-Column1), is likewise identifiable as a lexeme-sign. Figure 16b-Column2 lists 32 such lexeme-signs.

Criterion-9: This is a corollary of Criterion-6 and Criterion-8. Since post-connective constituents are semantically complete phrases, equivalent to semantically complete inscription-lines, the signs that occur alone with PF1-signs or PF-clusters in post-connective constituents (see Fig. 16c-Column1) are equivalent to the AWPF-lexemes (Criterion-6). Figure 16c-Column2 lists 52 such signs.

Frequent pre-connective lexemes (PCLs). As discussed in Section "Lexeme-signs based on the semantic completeness of pre-connective and post-connective constituents", 92 distinct lexeme-signs have occurred alone as pre-connectives in the inscriptions of 650 objects. Now, among these signs, the most frequent 5 signs ([signs]) have occurred as pre-connectives in 500 objects. That only 5% of the 92 signs occurred as pre-connectives in more than 75% of such objects, irrefutably proves that certain signs were much more suitable to the semantic scope of pre-connective constituents than others. So, these 5 signs (Fig. 6f) are classified as pre-connective lexeme (henceforth PCL) signs, based on their startlingly strong liking for pre-connective positions (see Table 2). Although [sign] has a much lower percentage of pre-connective occurrences, its combinatorial patterns closely resemble the patterns of other PCLs in many inscriptions (see Fig. 17a).

Some other signs ([signs] and [sign]) also demonstrate, just as PCLs, very high percentages of pre-connective occurrences. But since each of them occurs in less than 10 DILs, they are excluded from the "frequent" PCL list.
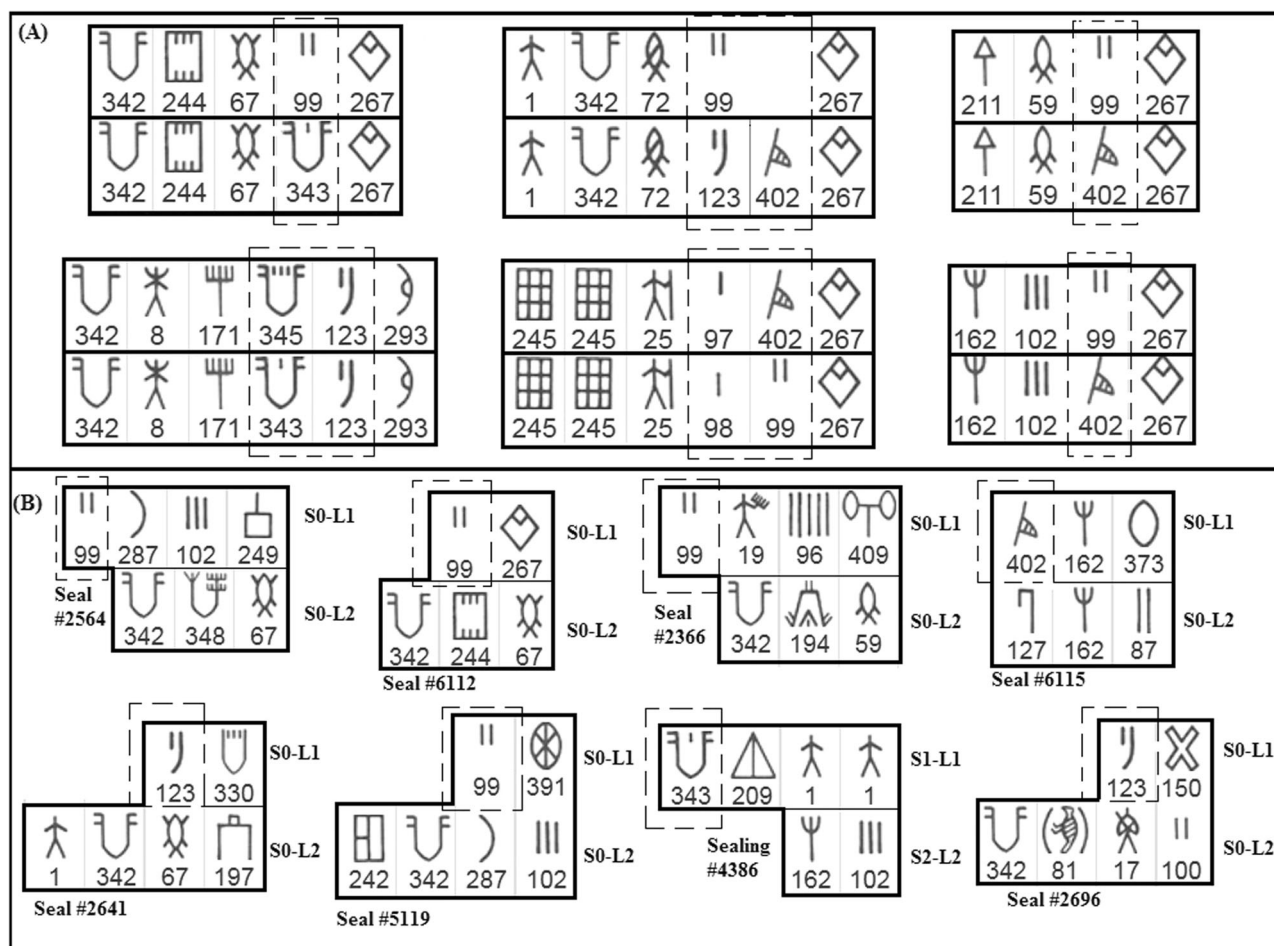
Fig. 14 Functional homogeneity of CMs **a**, and their patterns of occurrences in split-sequence inscriptions **b**

The examples in Fig. 17a demonstrate the functional homogeneity of PCL-signs. Moreover, even in non-pre-connective contexts, PCLs appear in mutually similar patterns (Fig. 17b, c).

Interestingly, as depicted in Table 3, certain PCL-signs reveal special affinity towards specific CM-signs (e.g., ⊠ occurs with ⟧ in 95% cases, whereas ⟧◇ never occurs as a construct in IDF-80).

Subordinating and coordinating roles of CM-signs. Reanalysing the "X CM Y" formats presented in Fig. 13, and searching the rest of IDF-80, we see that finding a "Y CM X" inscription such as ◇∥ᴗ⋏ or ⟩ᴗ⫴ᴍ⛏ remains extremely unlikely. This clearly indicates that the semantically distinct pre-connective and post-connective constituents mostly maintained a hierarchical relationship, akin to the subordinate and main clauses of natural language constructs, where cause-effect or purposive relationships are often expressed through the ordering of the subordinate and main clauses with respect to the purposive or causal conjunctions.

In Indus inscriptions, post-connectives must have functioned as principal clauses as they occur very frequently as independent inscription-lines (135 independent DILs have also functioned as post-connective constituents in 260 composite inscriptions). Contrastingly, despite being semantically complete, the pre-connectives generally do not occur as independent messages (e.g., ∥◇ occurs alone in only five objects, whereas in 277 objects it occurs in the pre-connective part), indicating that they were the subordinate clauses that provided some additional information about the post-connective main clause. As observed by linguists, generally subordinating conjunctions maintain "a fixed serial position in relation to their internal" arguments i.e., the subordinate clauses, "but not to their external argument" i.e., the principal clause (Blühdorn 2010). Thus, the abovementioned attachment of Indus CMs to the pre-connective sequences (Fig. 14b) reinforces my conjecture that the pre-connectives functioned as the subordinate clauses in Indus inscriptions.

Interestingly, some CM-signs (mostly ⟨ and ⟩) seem to have functioned also as coordinating conjunctions in certain composite inscriptions. There, the constituents occurring on either side of the CMs were similar in terms of their structure and sign content, and had similar levels of semantic independence (see Fig. 18b). For example, regarding the inscription ⌘⊕⛏∣ᴗᴛ⛏ of seal #4297, both the constituents ᴗᴛ⛏ and ⌘⊕⛏ have occurred as the complete message of different seals, sealings, tablets and ivory rods, and both have terminated with PF-signs, unlike the pre-connectives of typical subordinated clauses. Generally, in natural languages, unlike the relata of subordinating conjunction, "the relata of coordinators are typically of the same morpho-syntactic category" (Blühdorn, 2010). Thus, in the inscriptions of Fig. 18b, the CMs were arguably expressing some "and/or" kind of coordinating relationship, not any hierarchical subordinating relation. Interestingly, the typical PCL-signs (◇, ⊛, ⊠ etc.) seldom occur in such inscriptions, indicating that the subordinated composite inscriptions had a different semantic nature compared to the coordinated ones.

Crop-like signs (CROP-signs). There are certain phytomorphic signs (Ψ, ⴲ, ⴼ, and ▦) in IDF-80, which show a special tendency of following various kind of stroke-signs (see Fig. 19a). Moreover,
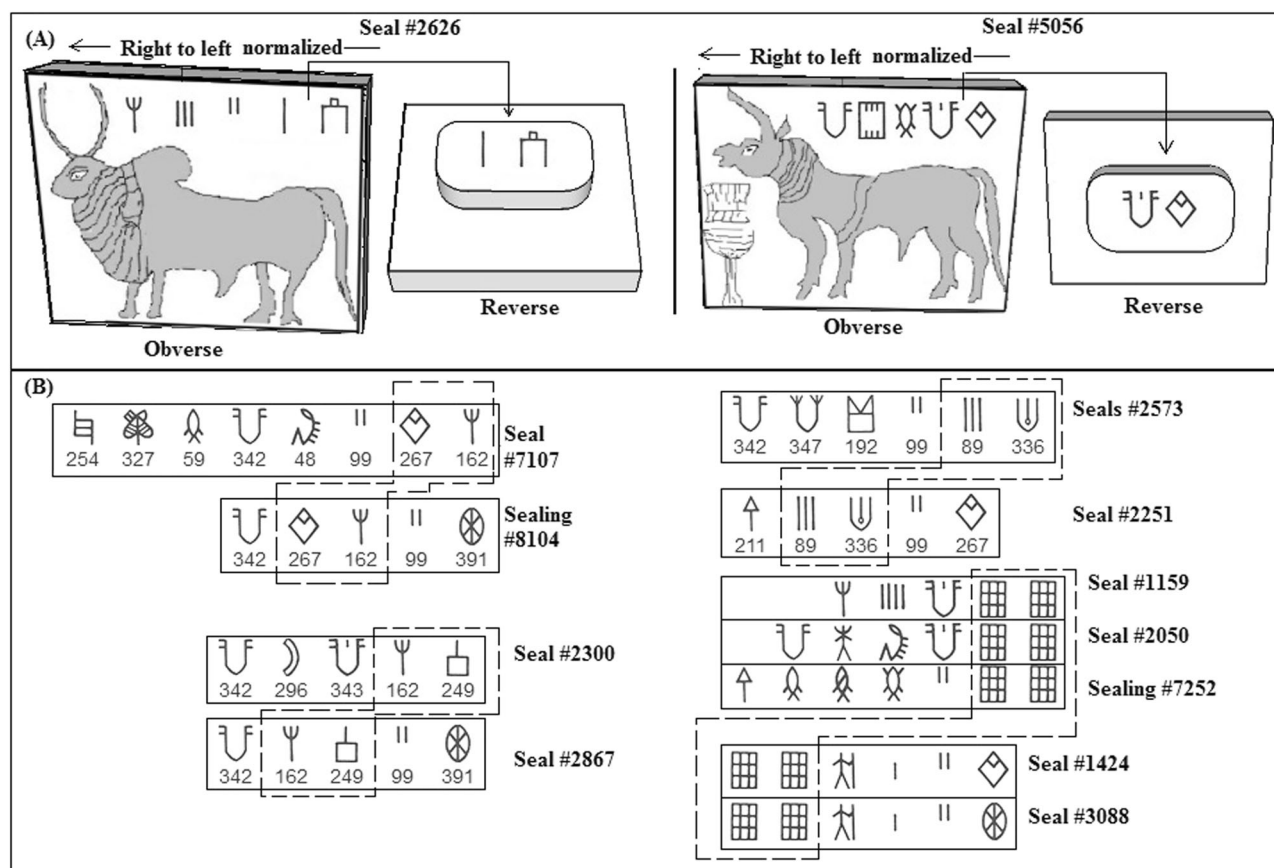
Fig. 15 Evidence of semantic completeness of pre-connective sequences. **a** Pre-connective part of the inscription on the obverse side being repeated on reverse boss; **b** Same bigrams appearing in pre-connective and post-connective parts of different inscriptions

when preceded by stroke-signs, they occur in the terminal positions of inscription-lines without any PF-signs following them (Fig. 19a). These phytomorphic signs are clustered together because of their functional homogeneity, and considering the close resemblance of their graphemes with the sheaves of grains or crops, their sign-class is named as CROP-signs (Fig. 6h). However, determination of whether these signs were semantically associated to grains or crops is not of concern here.

Three main features of CROP-signs are discussed below:

(i) Strong preference for specific preceding signs: Among the 319 DILs where CROP-signs occur, they generally follow a predictable set of signs. For example, CROP-sign ▓ occurs in 15 DILs, in 13 of which it follows sign ▓, and in 1 sign ▓, demonstrating a very strong affinity to three-stroked signs. Signs that most frequently precede CROP-signs are: Stroke-signs (100 DILs); sign ⊡ (75 DILs); signs ⊛ and ⊛ (26 DILs); and sign ⋀ (17 DILs). However, in pre-connective contexts (see Fig. 19c), CROP-signs generally follow a different set of signs (⊛, ⊛ etc.).

(ii) Terminal occurrences: CROP-signs have occurred at terminal positions without any posterior PF1-signs, in 161 DILs, i.e., 50% of their total occurrences. But whenever sign ⊡ precedes CROP-signs (see Fig. 19b), they are generally followed by PF1-signs (49 DILs).

(iii) Similarity between the ligatures: Graphically, signs ⊛ and ⊛ are ligatures of Ψ and Υ, made by putting their graphemes inside ovals. Interestingly, just like Ψ and Υ, their ligatures also demonstrate strong functional homogeneity by appearing in very similar inscriptions (e.g., compare the DILs ⊛ ⊛, ⊛ ⊛ and ⊎) ⊛, ⊎) ⊛).

Numerical signs (NUMs). There are 22 Indus signs, constituted by arranging long and short strokes in different horizontal and vertical patterns (see Fig. 6d). I identify these signs as numerical (NUM) signs based on their graphical appearances, combinatorial patterns and archeological contexts of usage. A vital clue regarding these stroke-signs' functionality is that a huge number of Indus pottery vessels and potsherds contain such signs in their body and rim inscriptions (see Fig. 1e), which according to Kenoyer "could be relating to accounting, such as the measure of the oil or grain placed in the jar prior to sealing it" (Kenoyer, 2006). Since several unbroken pottery vessels are found with only some stroke-signs inscribed on their body (see Fig. 1e), this research also argues that those strokes must have conveyed either the vessels' absolute capacity or the quantity of their content.

Interestingly, such stroke-signs also appear in typical patterns in Indus seals and tablets. For example, in each of the eight sets of 2-signs-long inscription-lines listed in Fig. 20, only the numbers of strokes constituting the stroke-signs vary, whereas their succeeding signs remain same. Now, repeating a sign n-times, to express the n-th multiple of its quantity, was a common technique used in the "cumulative-additive" and "multiplicative-additive" numerical systems of several ancient scripts, such as Egyptian Hieroglyphs, Aramaic, Proto-Elamite and Assyro-Babilonian (Chrisomalis, 2010; Nissen et al., 1993; Gardiner, 1969, p. 190–200). Many cubical dice of IVC (Dales, 1968, p. 14–23) also have used repetition of dots to express numerical values. Thus I propose that the repeated-stroke Indus signs also represented some quantification values, and their following signs represented the objects of that quantification. Since only a limited number of strokes have found use in the whole corpus, these signs
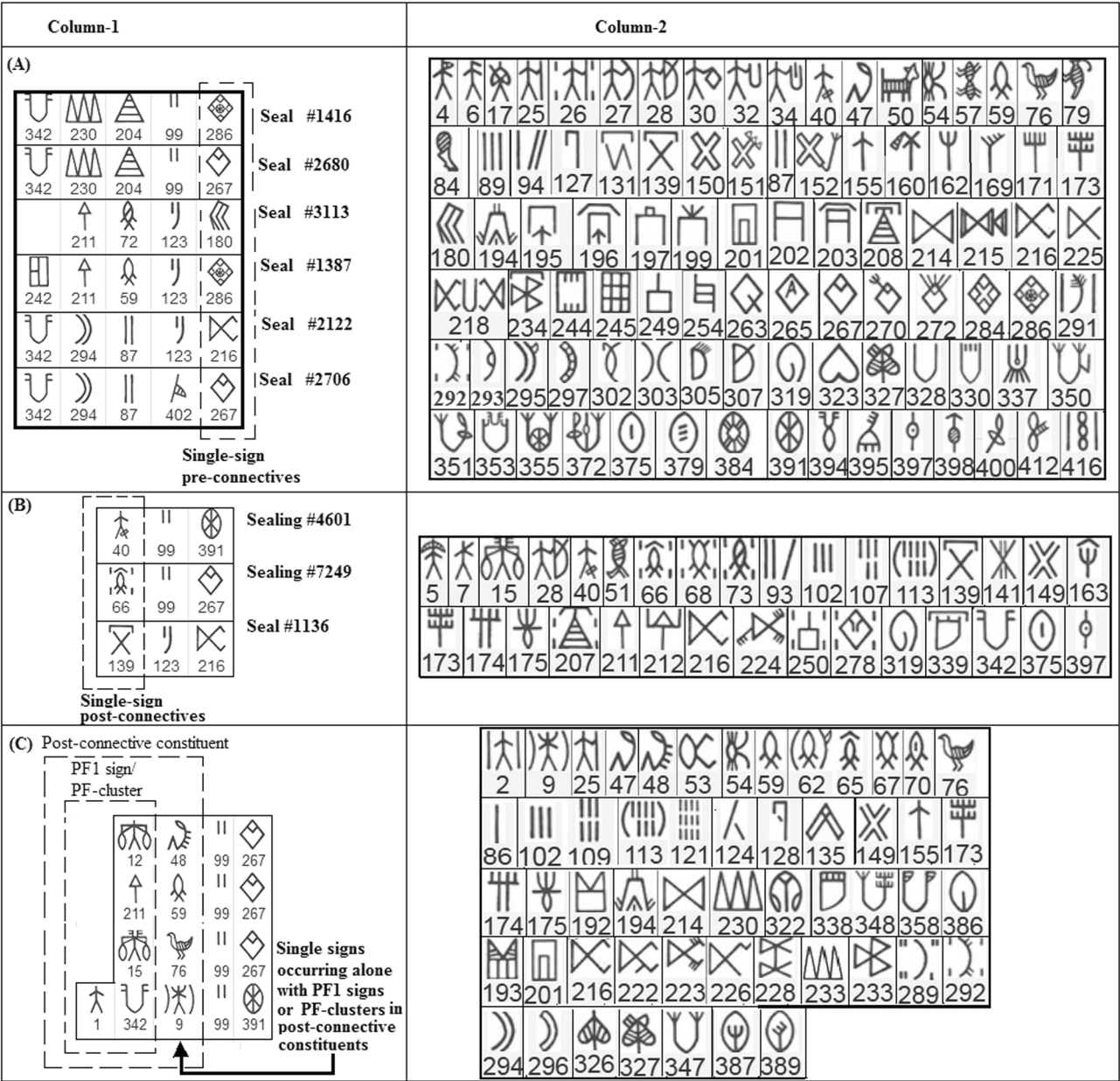
Fig. 16 Certain combinatorial patterns (Column-1) used to identify lexeme-signs (Column-2) of composite inscriptions. Signs occurring alone in pre-connectives **a** and post-connectives **b**. Signs occurring alone with phrase-finals in post-connectives **c**

possibly represented different denominations of some standard metrics used in ancient Indus economy. It is additionally contended here that the absolute quantities represented by the stroke-signs possibly varied depending on the types of objects they quantified. The proto-cuneiform numerals have demonstrated such context-dependent variances of the absolute value of the same numeral signs, depending on whether the quantified objects were animals, cereals or agricultural fields (Nissen et al., 1993, p. 131).

Thus, based on the patterns of the inscription-lines listed in Figs. 20a, 14 stroke-signs (❘❘, ❚❚, ❚❚❚, ❚❚❚❚, ❚❚❚❚❚, ❛❘, ❘❘, ❘❘❘, ❘❘❘❘, ❘❘❘❘❘, ❘❘, ❚❚, ❚❚❚, ❚❚❚❚) are identified as NUM-signs, in the first iteration. In the graph of Fig. 20b, the nodes represent the NUM-signs. Here, whenever two NUM-signs occur preceding a common sign, corresponding NUM-sign-nodes get connected with an edge labeled with the common sign's serial-number. Since the resultant graph turns into a connected multigraph containing many cycles,

Table 2 The percentage of pre-connective occurrences of PCL signs

PCL sign	Count of DILs containing the PCL-sign	Count of DILs where the PCL occurs in pre-connective part	Percentage of pre-connective occurrences
◈	295 (11*)	256	87.46%
⊛	167 (6*)	123	75.45%
⟩	116 (5*)	76	63.79%
⊠	51	39	76.47%
⋈	72 (3*)	27	38.89%

The bracketed counts marked with "*" denote the count of instances containing "doubtfully read" PCL signs
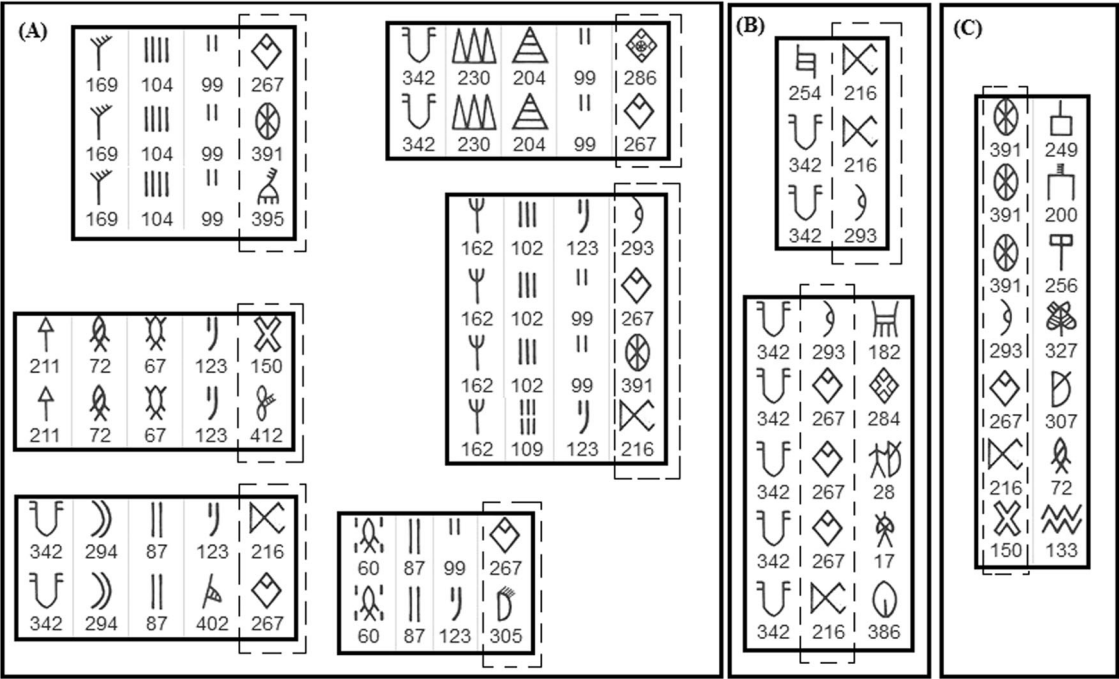
Fig. 17 Different combinatorial patterns of PCLs. **a** PCLs occurring in similar pre-connective contexts; **b** PCLs preceding PF1s; **c** PCLs occurring in the terminal positions

Table 3 Association between specific PCL and CM signs

PCL-sign	CMs that followed the PCL at least once	Count of DILs
◇	‖, ˡ‖, ⒜, ˌ , ˮ⒜, ˡ⒜, ˌ⒜, ᵁˮ⒤, ᵁ⒤	282
⊛	ˡ, ˌ, ᵎ, ᵎˡ, ˡᵎ, ˡᵎ	117
⎅	ˡ, ˌ, ᵎ, ᵎ, ᵁˮ, ⒜, ᵎ⒜, ˡᵎ, ˡ⒜, ᵁˮᵎ, ᵁˮᵎ, ᵁˮᵎ	63
⊠	ˡ, ˌ, ᵎ	37
⋈	ˡ, ˌ, ᵎ, ᵁˮ	22

it is clearly visible that all these NUM-signs shared very similar inscriptional contexts, and were functionally homogeneous.

The stroke-signs ‖/, ‖, ⦀⦀, ⦀⦀, ‖/, ⦀⦀, ⦀⦀, ⦀⦀ and ᴛ are also classified as NUM-signs, based on their graphical, combinatorial and contextual similarities compared to the 14 NUM-signs classified above (see examples in Fig. 21, and detailed discussion in Section-S3 of supplementary-file Supp1). Non-stroke signs ⊛, ⊛, ⎅, ⌒, ⌂ and ⌂ too are classified as NUM-signs, since, like stroke-numerals, they also demonstrate striking combinatorial and contextual similarities (see examples in Fig. 22, and detailed discussion in Section-S3 of supplementary-file Supp1).

Metrological signs (METs). Identification of NUM-signs facilitates detection of another class of signs, namely the metrological (MET) signs (Fig. 6g). Unlike the NUM-signs whose numerical values may change based on the objects of quantification, the MET-signs (△, ⊟, ∪, ∪, ∪, ∪, ∪, ∪) seem to represent different denominational values of fixed mensural standards represented through their basic graphemes. The identification process for each MET-sign is separately discussed below.

MET-signs ∪, ∪, and ∪: The 241 multi-sided sealings and tablets recorded in IDF-80, whose reverse sides typically contain one of the four inscriptions: ∪ˡ, ∪ˌ, ∪ᵎ and ∪⦀, provide the most crucial clue regarding the nature of the rimless jar-line sign ∪. Since these <NUM-∪> constructs contained only 4 variants of stroke-numerals, it becomes clear that they conveyed values related to some standardized quantification process which could assume mainly four numerical denominations. Since, unlike other quantified lexeme signs, ∪ rarely occurs in obverse inscriptions, it possibly represented a special standard of quantification. Interestingly, the obverse-sides of these multi-sided artifacts always contained usual semantically complete inscriptions. For example, inscription-line ᵁⓞⓞᵎ, which occurs alone in seals of Lothal and Mohenjo-daro, also appears on the obverse-side of eleven such multi-sided objects of Harappa. Since each of the reverse-side inscriptions ∪ˌ, ∪ᵎ and ∪⦀ has occurred at least once with ᵁⓞⓞᵎ (Fig. 23a), it is evident that the message of ᵁⓞⓞᵎ, was applicable to all the denominational quantities of the metrological-standard represented by sign ∪. Thus, ∪ possibly represented some standard equivalency, which was applicable to all such obverse-side messages. Since all but one of these multisided objects were found in Harappa, Harappa must have been the center of some administrative bureaucracy where this quantification and standardization process associated with sign ∪ was commonly practised.

Now, the rimless jar-like appearance of ∪ tempts its interpretation as an ancient volumetric unit. In fact, using the dimensions of some Indus vessels bearing the <NUM ∪> constructs, Wells (2015, p. 59–65) and Fuls (2010) have sought to deduce the volumetric value represented by ∪. Although presently available data is inadequate to validate their results, a very important graphemic clue corroborating this conjecture is obtained from certain bas-relief tablets of Mohenjo-daro, where the inscription-line ⒠⒤∪⦀ is positioned beside the iconography of "a sitting man making offering to a tree". Interestingly, as shown in Fig. 23b, "the offering vessel" held by the man in the
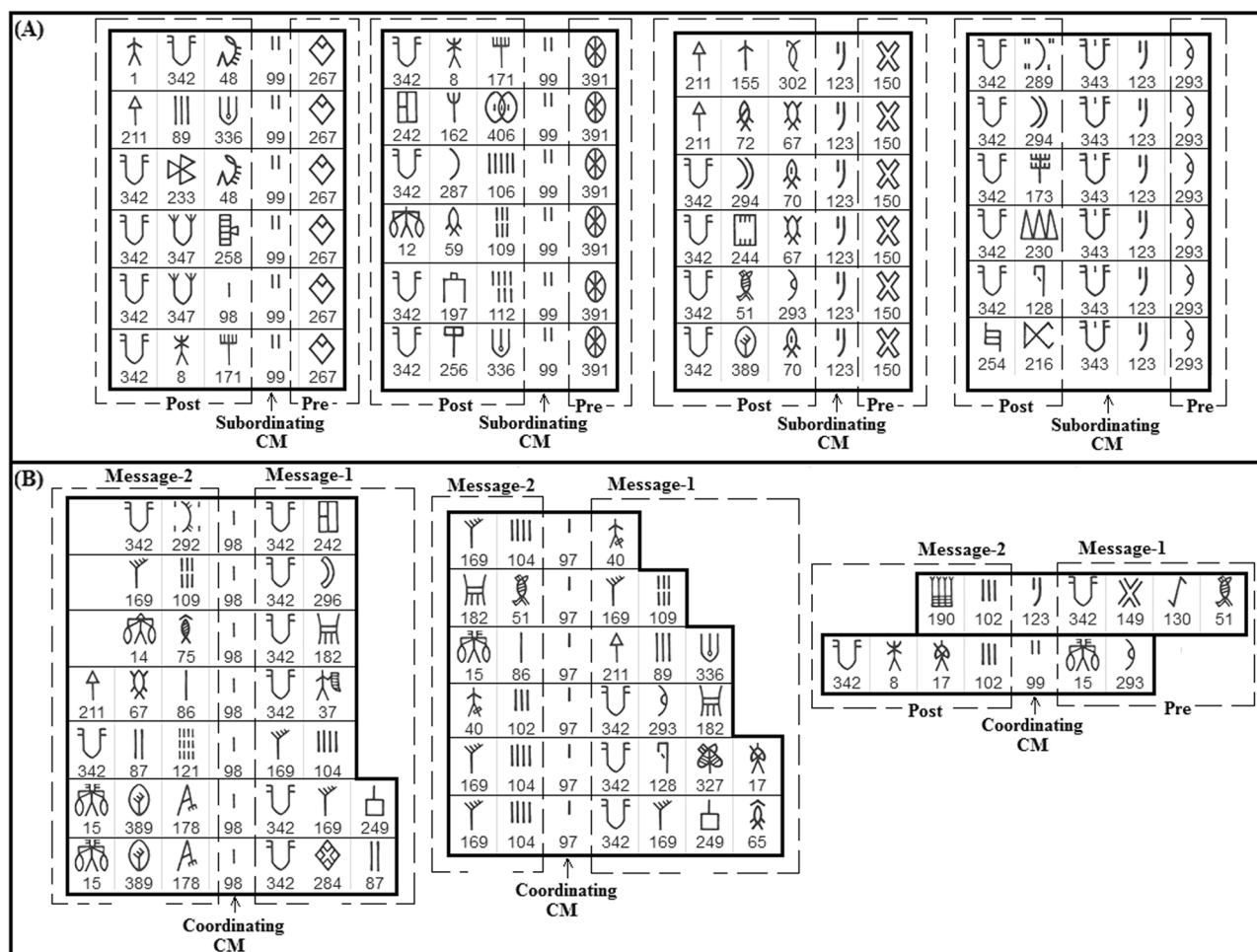
Fig. 18 Example inscriptions where CM-signs work as subordinating conjunctions **a** and coordinating conjunctions **b**

iconography is graphically identical to sign ⊔ (Parpola, 1994, p. 109; Wells, 2015, p. 56). Satisfaction about ⊔ graphically representing a vessel strengthens the interpretation that it functioned as a volumetric unit, especially since volumetric units were often represented by such symbols in ancient scripts. For example, the pictographic symbol of bevelled-rim bowls had entered the proto-Sumerian texts as an ideographic sign 'ninda' (initially ◊, then transformed to ◊) (Nissen, 2011, p. 70). Since bevelled-rim bowls were vessels of standardized quantities, possibly used in daily ration disbursement, the 'ninda' symbol, which was a standard measurement unit, also signified grain-ration as a derived meaning (Nissen, 2011, p. 70–71). The study accordingly argues that just as the 'ninda' sign, sign ⊔ also represented some standardized volumetric unit prevalent in ancient IVC.

Interestingly, in the Mesopotamian context, "A sign composed of the NINDA sign with vertical strokes above it appeared from the Uruk III (Jemdet Nasr) phase onwards, and this sign certainly signified 'ration' during the third millennium B.C. (Sumerian bur, Akkadian Naptānum)" (Millard, 1988, p. 53). Intriguingly, the descriptions of such stroked-◊ signs closely match the stroked-⊔ graphemes of signs ⊍ and ⊎, which have occurred in 15 inscribed objects of IDF-80. Thus based on their graphemic similarity, and shared numerical notation with ⊔, signs ⊍ and ⊎ are also identified as MET-signs.

METs ⊍, ⊎, ⊍: Although the stroked-jar signs ⊍, ⊎ and ⊍ are already classified as CM-signs, as discussed below, in some

inscriptions they have evidently pronounced metrological overtones as well.

(i) **Graphemic evidence:** Signs ⊍, ⊎, ⊍ and ⊍ (⊍ was discovered after the compilation of IDF-80, (Mahadevan, 1977:25)) are composed by putting different numbers of strokes inside the grapheme of the jar-like PF1-sign ⊔, clearly indicating the use of some shared numerical notation.

(ii) **Usage patterns similar to stroke-numerals:** As shown in Fig. 23c, signs ⊍, ⊎ and ⊍ often occur in similar inscriptional contexts as other stroke-numerals (e.g., |, ‖, ‖ etc.).

(iii) **Archeological evidence:** Numerous pre-firing inscriptions and post-firing graffiti, often coexisting in the rim and body of the pottery vessels used for trading commodities in IVC, consisted of both inscriptions and simple tally marks used for accounting (Kenoyer, 2006). In my view, the frequent presence of <NUM-CROP> constructs in numerous pottery shards (see CISI) indicate that the CROP-signs possibly represented a special type of commodity, whose standard quantification values were expressed through their accompanying numerals. So, when the stroked-jar signs appear in the rim and body inscriptions together with stroke-numerals, it strongly indicates that they also represented some standard quantification system used in IVC. For example, the inscriptions ⊍‖ and ‖⊮ were found, respectively, in the rim and body of a large-sized jar (#2931, CISI #M-2062) of Mohenjo-daro (Fig. 23d). Notably, two types of quantifier-quantified constructs have coexisted in this jar: i)The ligature ⊮ made of NUM-sign ‖ and CROP-sign ⊻, that possibly quantified the jar's content; and ii) the stroked-jar sign ⊍ following the NUM-sign ‖
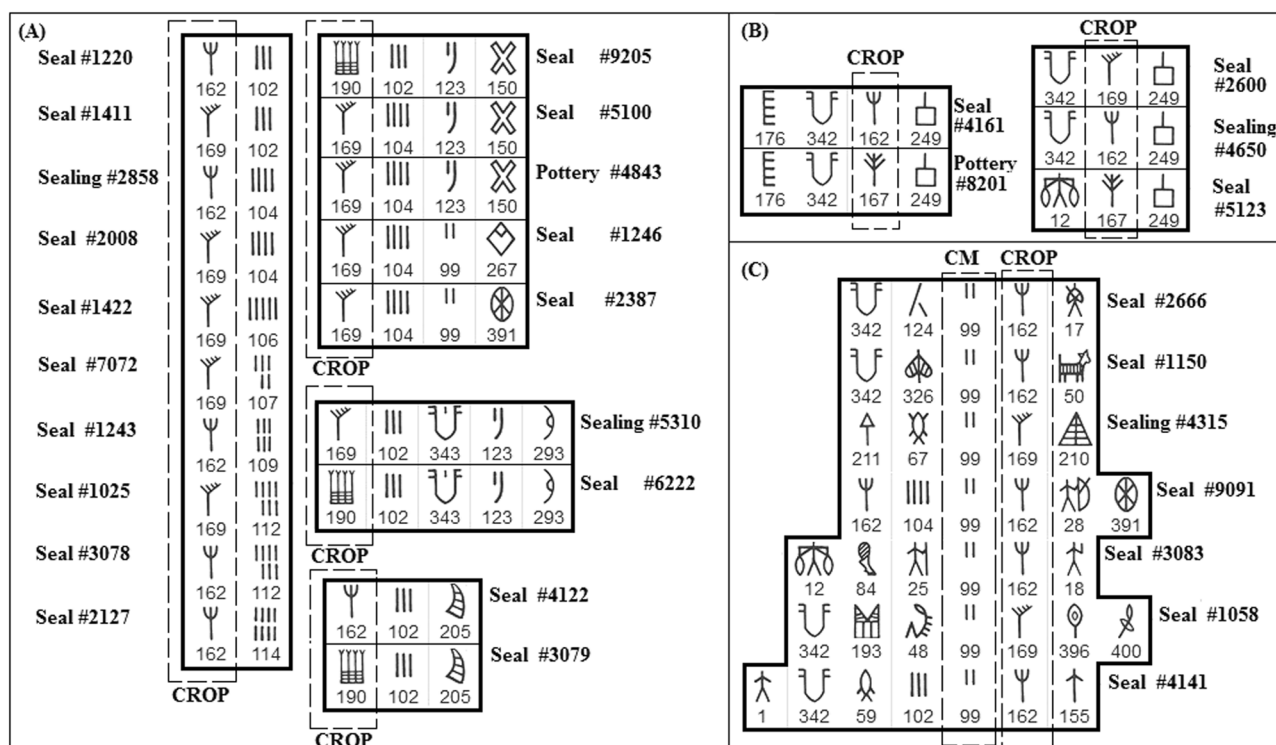
Fig. 19 Certain characteristic combinatorial patterns of CROP-signs. **a** CROP-signs occurring in terminal positions preceded by stroke-numerals; **b** CROP-signs preceded by sign ⊡; **c** Pre-connective occurrences of CROP-signs
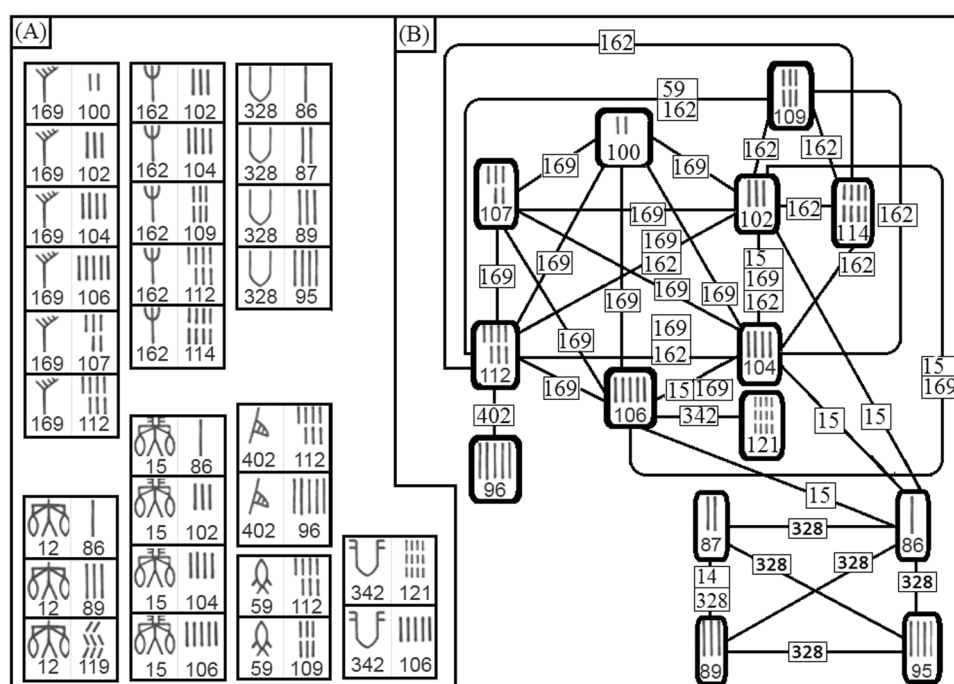

Fig. 20 Certain usage patterns of numerical signs. **a** Inscriptions where a common sign follows different stroke-signs. **b** Graphical representation of common inscriptional contexts shared by numerical signs

that probably represented the absolute volume of the jar. Similarly in the inscription found on jar #2936 (CISI #M-2061), the stroked-jar sign precedes a CROP-sign, resembling the <NUM-sign CROP-sign> construct of jar #2931. Interestingly, in a parallel pattern, the co-occurrences of the 'ninda'-sign (𝖣) with the phytomorphic grain-unit "SE" (the crop-sign of barley ⤳) in

proto-cuneiform tablets, have helped Sumerologists to find the rough value of the grain-based numerical system (Nissen et. al., 1993).

All such direct and indirect evidence corroborates my claim about the metrological functionality of the stroked-jar signs.

MET Sign ⊡ Sign ⊡ shows a very strong association with CROP-signs (Fig. 19b). Of 299 DILs where CROP-signs occur,
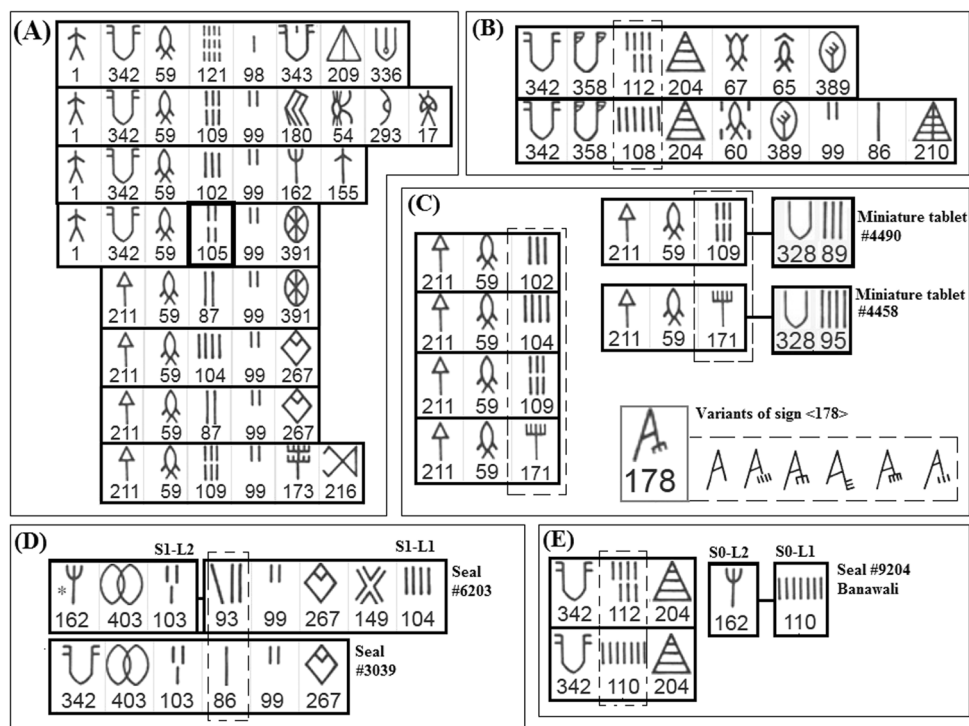
Fig. 21 Stroke-signs sharing common inscriptional contexts with other stroke-numerals. Examples relating to sign ⫴ **a**, ⫼ **b**, ⵋ **c**, ⫽ **d**, and ⫼ **e**
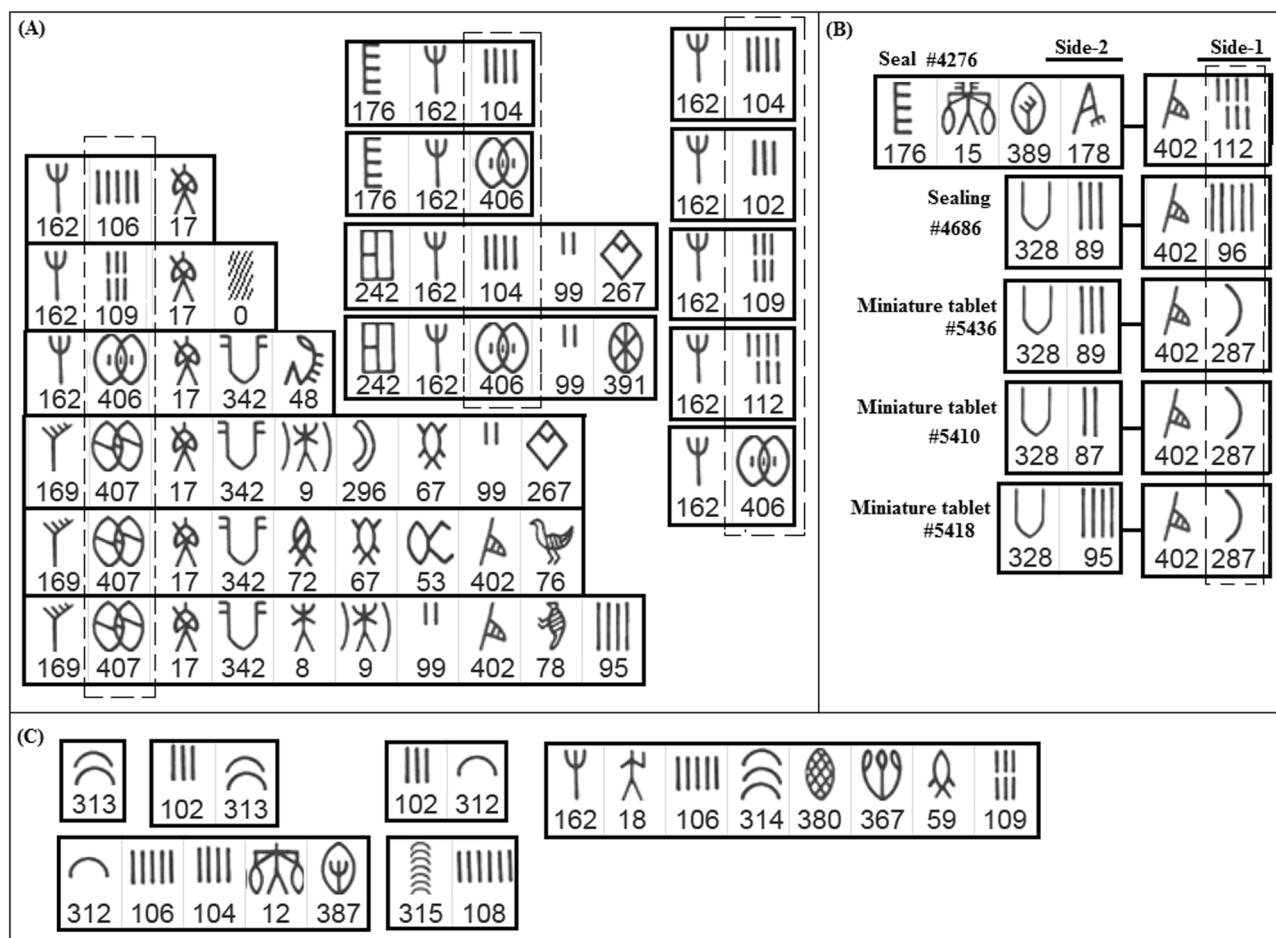
Fig. 22 Usage patterns of signs which are not stroke-signs but behave like numerals. Patterns for signs: ⊘⊘ and ⊛⊘ in **a**, sign ⟩ in **b**, and signs ⌒, ⩘ and ⩘ in **c**
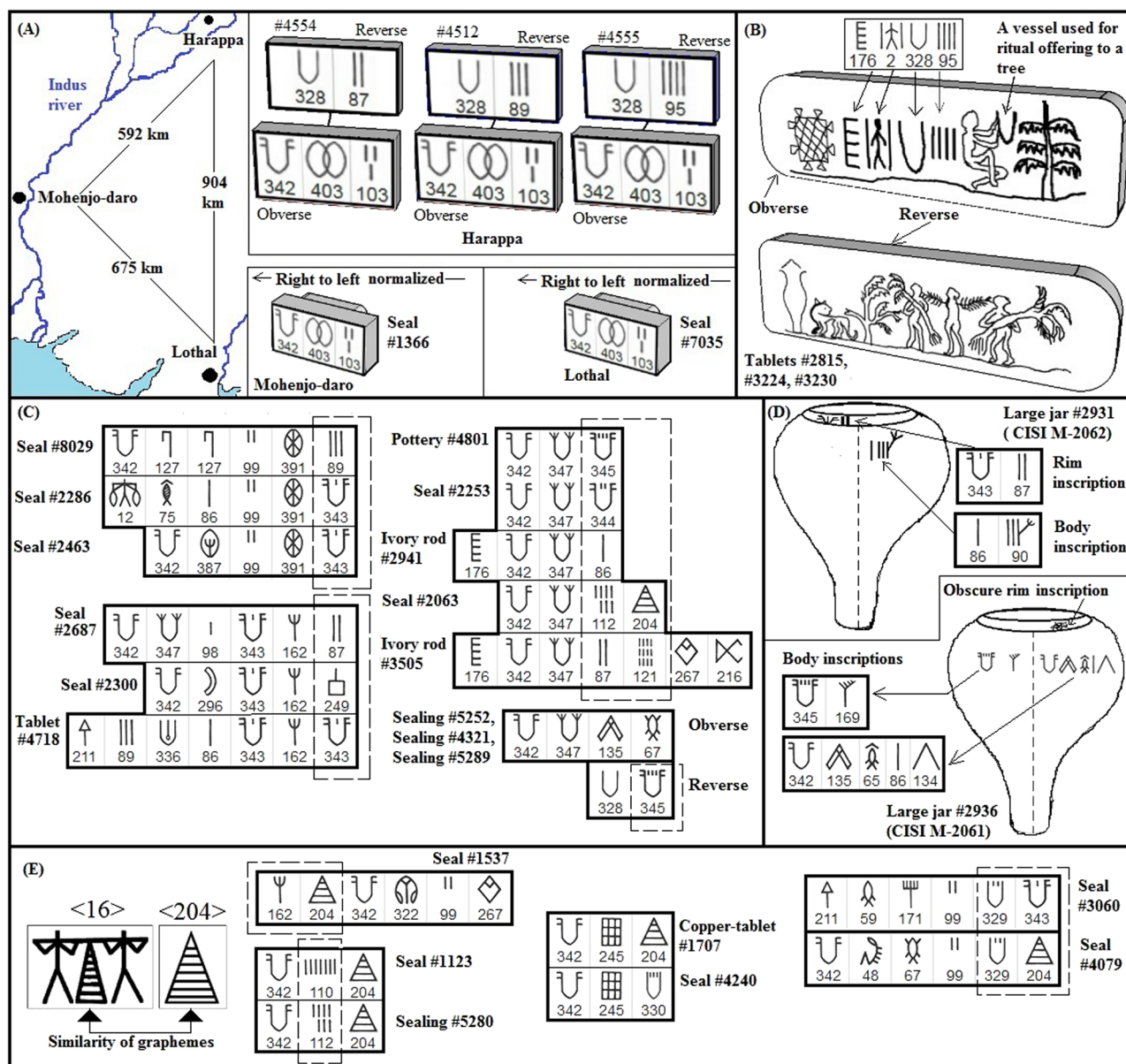
Fig. 23 Certain typical occurrences of MET-signs. **a** Occurrences of ⊍ on reverse-side inscriptions; **b** Graphical similarity between ⊍ and a ritual vessel iconography; **c** Certain numeral-like occurrences of stroked-jar signs; **d** Occurrences of stroked-jar signs in pottery graffiti. **e** Occurrences of signs △, ⊍, and ⊍

112 contain <NUM-CROP> constructs; whereas in 73 DILs sign 凸 precedes the CROP-signs. Intriguingly, when preceded by NUM-signs, CROP-signs mostly occur in the terminal-positions, without any PF1-signs following them. But, when CROP-signs are preceded by 凸, in 87% cases some PF1-sign follows the CROP-signs (e.g., ⊍Ψ凸, ⊍Ψ凸, ⊗Ψ凸 etc.), indicating that in such contexts sign 凸 semantically replaces the NUM-signs, and in doing so it needs PF1-signs to follow the constructs. Thus, through sharing a contrastive context with NUMs, sign 凸 indicates that it is functionally connected to some quantification and metrology.

Some other inscriptions also suggest that sign 凸 could be functionally replace the numeral quantifiers. For example, though sign ⊍ collocates with NUM-sign ⦀ in 101 DILs, when preceded by sign 凸 ⊍ is never followed by a numeral,. Similarly, in sealing #7280 of Lothal, 凸 has preceded ⊍ in a position that is predominantly occupied by stroke-numerals.

MET-sign △: The sign 衯, graphically comprising two anthropomorphic figures carrying a triangular object suspended from a shoulder pole, provides an interesting graphical clue about the nature of sign △. Since the grapheme of △ closely resembles the suspended object of 衯 (see Fig. 23e), sign △ seems to have represented something that had to be carried, possibly a symbol of some weight-based system. Combinatorially, in seal #1537, sign △ precedes CROP-sign Ψ just like the stroke-numerals. It also often occurs adjacent to the stroke-numerals, and MET-signs ⊍ and ⊍ (Fig. 23e). Moreover, in certain seals, △ is found in similar positions as MET-sign ⊍ (Fig. 23e). Considering all these, △ is classified as a metrological sign.

Nature of the quantifier signs and their relationship with pre-connectives and phrase-finals. As discussed before, the numerical signs of ISC used such a restrictive number of strokes in such restricted patterns, that they were surely incapable of representing
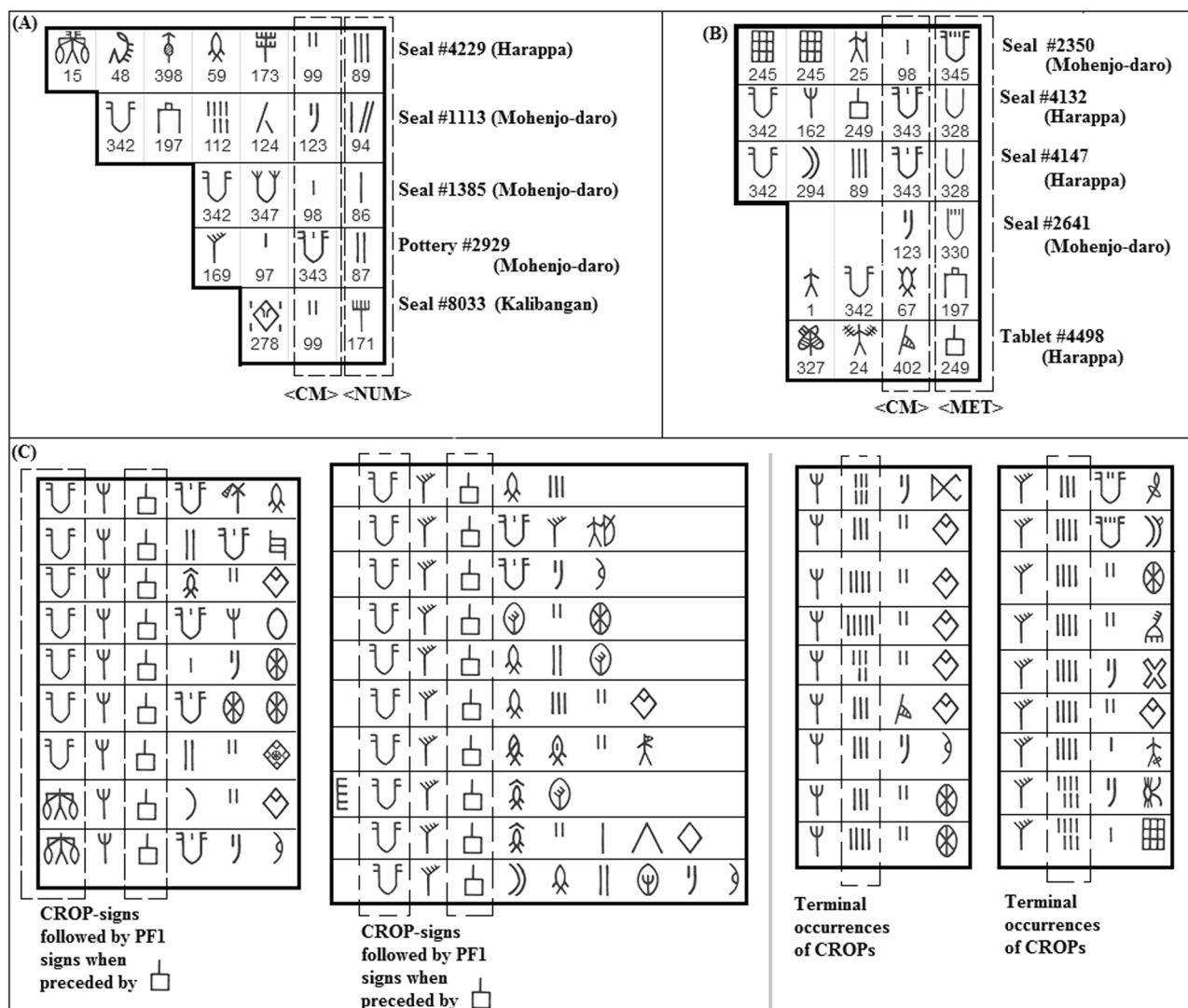
Fig. 24 Connection between NUM, MET, PCL and PF1-signs. **a** NUM-signs in pre-connective positions; **b** MET-signs in pre-connective positions; **c** PF1-signs following <⬚ CROP> constructs

the ad hoc quantities used in daily commercial transactions. In fact, their pre-designed usages in seals and pre-firing pottery tally marks clearly indicate that they represented various standardized quantities used in IVC's economy, just like the restrictive range of numerals found in modern tax-tokens, currency-coins and measuring cans. Below, I shall discuss the intriguing relationship shared by the quantifier signs and the pre-connective and PPFs.

(i) Numerical pre-connectives and metonyms: Revealingly, the NUM-signs and MET-signs have appeared alone in the pre-connective positions of some composite inscriptions, found in seals excavated from several different locations of IVC (see Fig. 24a, b). Now, NUM-signs and MET-signs are characteristically attributive lexemes, generally used for quantifying their left-adjacent substantives. Since pre-connective signs are supposed to be substantive lexemes, the use of quantifier-signs in such positions is apparently unexpected. A possible explanation of these usages is that certain numerical and metrological values were so closely associated with certain commercial or economic processes that they were used as the metonyms of those processes, and hence could play the roles of substantives or nouns. A good Indic example of such metonymy is that the fraction "ṣaḍbhāga", meaning one-sixth, also signified the royal tax, because the rate of

that tax was traditionally fixed to be one-sixth of the produce (Thapar, 2015).

Semantic connection between PF1s and METs: As discussed before, while preceding the CROP-signs of ISC, the contrastive patterns demonstrated by the MET-sign ⬚ and the stroke-numerals (Fig. 24c), strongly suggest that they signified different modes of standardized quantification, and the PF1-signs were applicable in only one of these modes, indicating an indirect semantic link between PF1-signs and metrology. Moreover, the MET-signs ⑃, ⑁ and ⑂ use the graphemes of the most frequent PF1-sign ⑃. Now, in logographic scripts, the choice of similar graphemes for different logograms strongly indicates some underlying semantic connection between them. Thus, though the phrase-level role of PF1s suggests that they were not metrological qualifiers themselves, they surely had some strong connection with quantification and metrology.

Encapsulated (ENC) signs. Nineteen signs of ISC are constituted by enclosing graphemes of certain other signs inside typical 4-stroke circum-graphs (see Fig. 25a). This special graphemic feature helps to cluster these signs into a special sign-class, called ENCs. Since ENCs could occur alone in complete inscription-
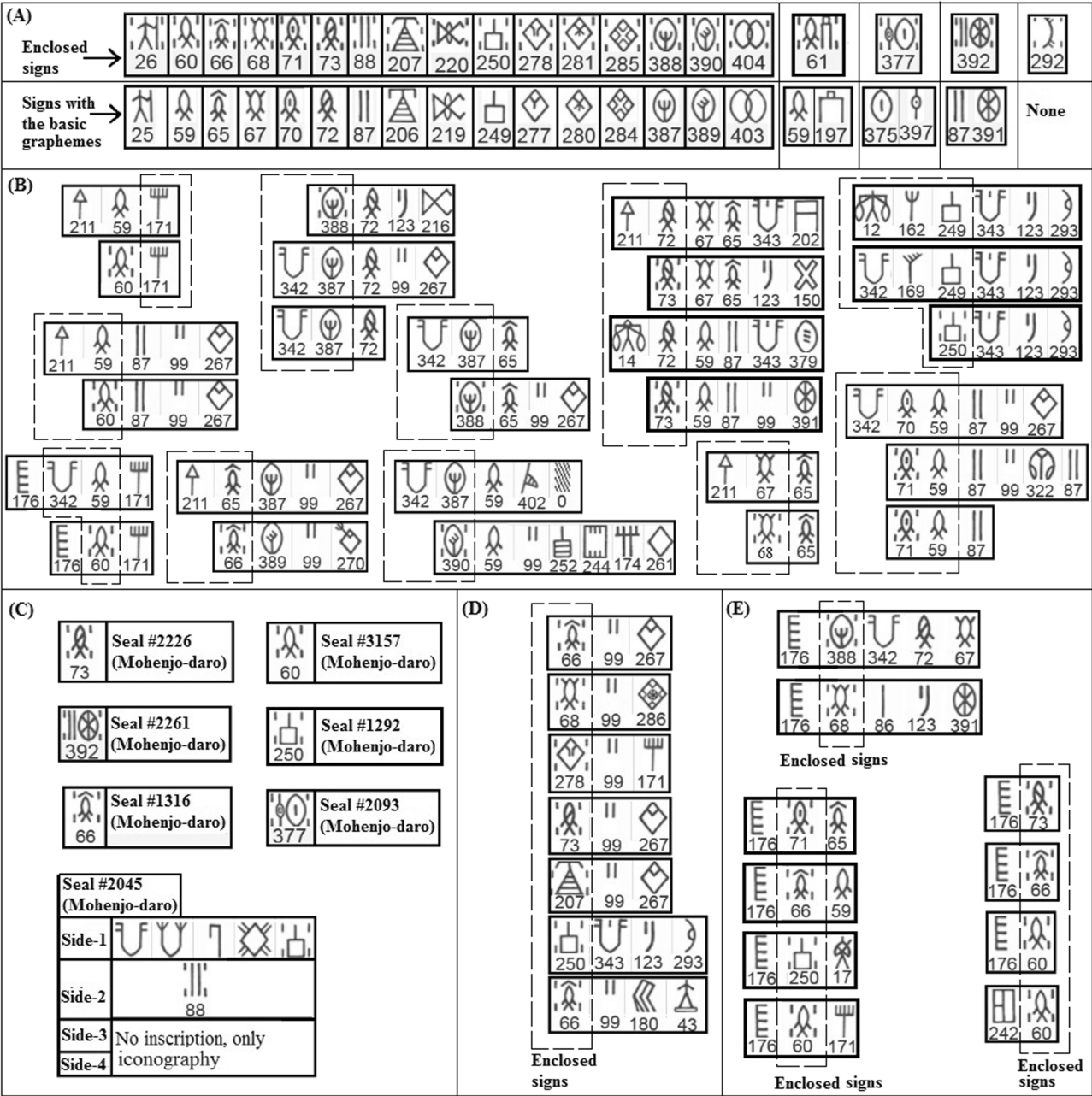
Fig. 25 Graphical and combinatorial features of Enclosed signs—**a** List of ENC-signs and their basic graphemes; **b** ENC-signs replacing their basic signs and PF1s in similar inscriptional contexts; **c** ENC-signs occurring alone in artifacts; **d** ENC-signs occurring alone as post-connectives; **e** ENC-signs getting directly followed by PF2-signs

lines (Fig. 25c), and in post-connective constituents (Fig. 25d), they were indubitably logograms (Criterion-1, Criterion-8).

A scrutiny of the combinatorial patterns of ENCs reveals that they very often substitute the constructs made by their basic signs and PF1-signs, in identical inscriptional contexts (Fig. 25b). For example, comparing the DILs ⟦𝑋⟧ ‖ ″ ◇ and ⟦↑ 𝑋 ‖ ″ ◇⟧, or ⟦𝑋⟧ 𝑇 and ⟦↑ 𝑋 𝑇⟧, we find that ⟦𝑋⟧ replaces the construct made by its basic sign 𝑋 and PF1-sign ↑, in otherwise identical contexts. Interestingly, just like PF1s, ENCs too frequently occur in the terminal positions (Fig. 25d), often preceding PF2-signs (Fig. 25e). Moreover, ENCs are rarely followed by PF1-signs (only in 4 of 110 objects). Thus, the circum-graph ⟦ ⟧ of ENCs must have played a similar semantic role as the PF1-signs. On the other hand, the ENCs completely retained the semantic content of their

basic signs, as the frequent collocations formed by their basic signs (e.g., 𝑋 ‖ and 𝑋 𝑇), are also maintained in their enclosed forms (e.g., ⟦𝑋 𝑇⟧ and ⟦𝑋 ‖⟧).

Unclassified Indus signs and their probability of being lexemes. Excluding the 254 lexemes and functional morphemes identified above, 163 unclassified signs remain (sorted and grouped in Fig. 26 according to their frequency of occurrences in DILs). Now, 133 (81%) of these unclassified signs occur in only one to five DILs, whereas 44% occur in just one DIL. This study argues that having occurred in very few of the artifacts excavated till now, these signs possibly missed the isolable positions used as the criteria for identifying lexemes. However, if we consider the startling graphical (Fig. 27a) and combinatorial similarities (Fig.
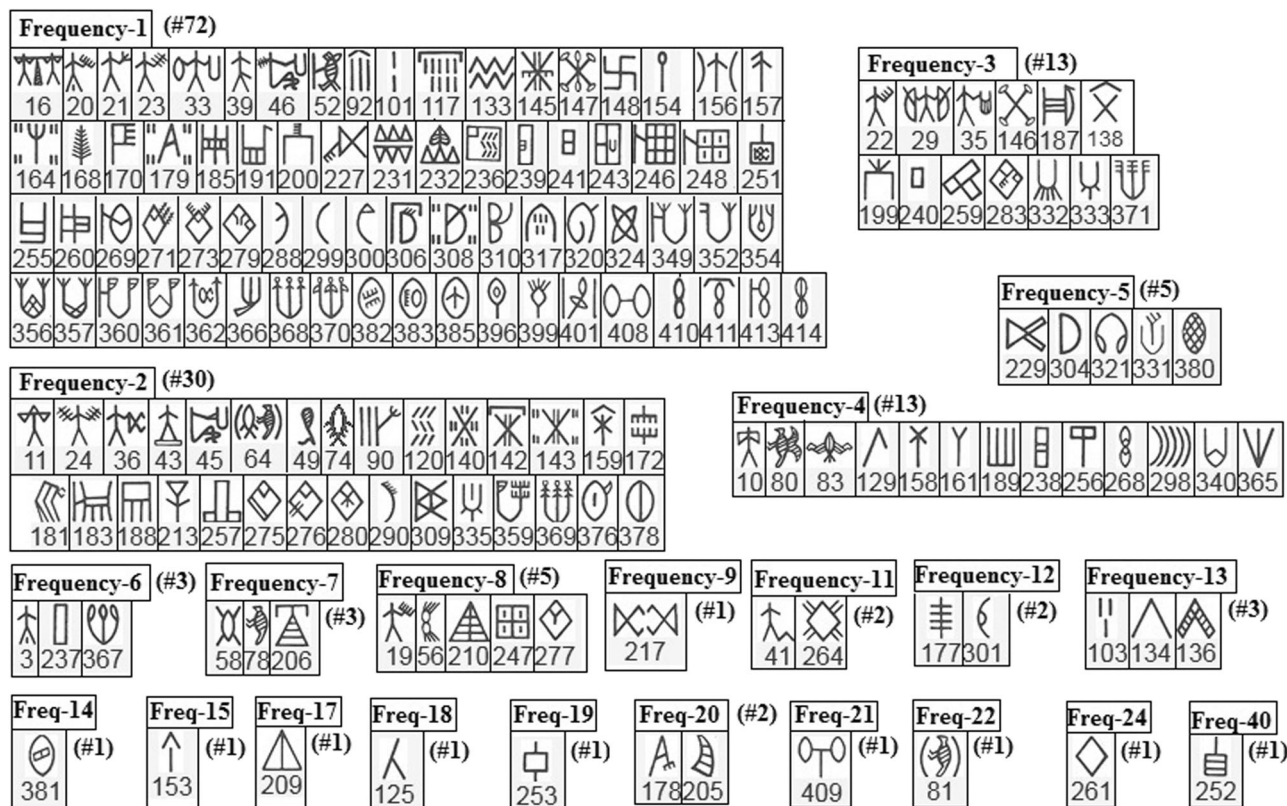
Fig. 26 Unclassified Indus signs grouped by their frequency of occurrences in distinct inscription-lines

27b) shared by these signs with their classified counterparts, it becomes evident that a large number of these signs were undoubtedly lexemes too.

Generative processes used to create new Indus signs. One of the arguments made by Farmer et al. (2004), to deny ISC the status of a "genuine script" is that in genuine scripts "the percentages of singletons and other rare signs" are expected to "drop as new examples of those signs showed up in new inscriptions". But, in ISC, "those percentages appear to be rising instead over time, suggesting that at least some Indus symbols were invented 'on the fly' only to be abandoned after being used once or a handful of times" (Farmer et al., 2004). Despite the incisive novelty of their observation, a different perspective regarding this is provided below, that might induce second thoughts about their conclusion.

First, notwithstanding the finding of different infrequent signs in new excavations, "the new signs have more often been ligatures of two or more signs already known as separate graphemes than entirely new signs" (Parpola, 1994, p. 79). For example, the one-timer compound sign ⬚ comprises the graphemes of lexemes ⬚ and ⬚. Figure 27a shows more such examples. As already discussed by Parpola (1994, p. 79–80), ISC had used certain standard methods of making more complex and compound graphemes from basic graphemes. Some of these methods were: adding modifiers such as ⬚ or ⬚ to basic signs (⬚, ⬚, ⬚, ⬚, ⬚, ⬚ and ⬚, ⬚, ⬚, ⬚, ⬚ etc.); joining graphemes of certain basic signs to the hand(s) of an anthropomorphic sign (⬚⬚⬚⬚⬚⬚⬚⬚⬚⬚⬚⬚); making mirror-image sign pairs (⬚, ⬚, ⬚); constructing ENC-signs by putting basic signs inside a 4-stroke circum-graph ⬚ (⬚, ⬚, ⬚, ⬚ etc.); inserting graphemes of basic signs inside an oval ⬚ (⬚, ⬚, ⬚, ⬚); or simply conjoining graphemes of two or more basic signs

(⬚⬚⬚⬚⬚⬚⬚). Now, the use of generative rules to construct new phonetic, semantic or grammatical units by reusing existing ones is a universal characteristic of linguistic systems, goaded by the need to strike a balance between "economy of derivation" and "economy of representation" (Chomsky, 1995). Because ISC also uses such generative modifiers (⬚, ⬚ etc.), many new signs could possibly be constructed and deciphered by Indus people without much ado, and even if those signs are infrequent in the artifacts excavated yet, this detracts little from ISC's status of a "genuine script". Rather such generative patterns strongly indicate the linguistic nature of the Indus signs.

Secondly, as Wells (2011, p. 74–75) shows, signs like PF1 ⬚ had assumed certain allographic variations in different stratigraphic layers of Indus valley locations. But the sign's functional features remained same in all these variations. This type of data proves that Indus phrases have used the key functional signs in the same way for quite a long period. On the other hand, the general substantive lexemes that mainly function as content-morphemes carrying information, possibly represented an open class of lexemes with potential to subsume some new member-signs with further excavations of Indus artifacts. This is possibly because, if some of the Indus signs represented certain commodities used in IVC, a new sign would then be needed whenever a new type of commodity is used.

Collocations and repeated sign sequences

Collocations of ISC. In the Oxford dictionary (2016), the term 'collocation' is defined as "The habitual juxtaposition of a particular word with another word or words with a frequency greater than chance". Interestingly, in Indus inscriptions too, certain signs co-occur adjacently, maintaining a specific order, in far-greater-than-chance frequencies. Since Indus signs do not

Fig. 27 Similarities between certain uncategorized Indus signs and certain Indus logograms. **a** Graphical similarities. **b** Combinatorial similarities

demonstrate statistically significant correlations beyond the bigram level (see Section-S4 of supplementary-file Supp1), this study concentrates mostly on the bigrams of ISC.

Identifying the "true collocations" of ISC is an important task. For, when "reading" the inscriptions, one must know which parts of them should be "read" together as smaller semantic units. Here "true collocations" refer to the frequently co-occurring sign-sequences 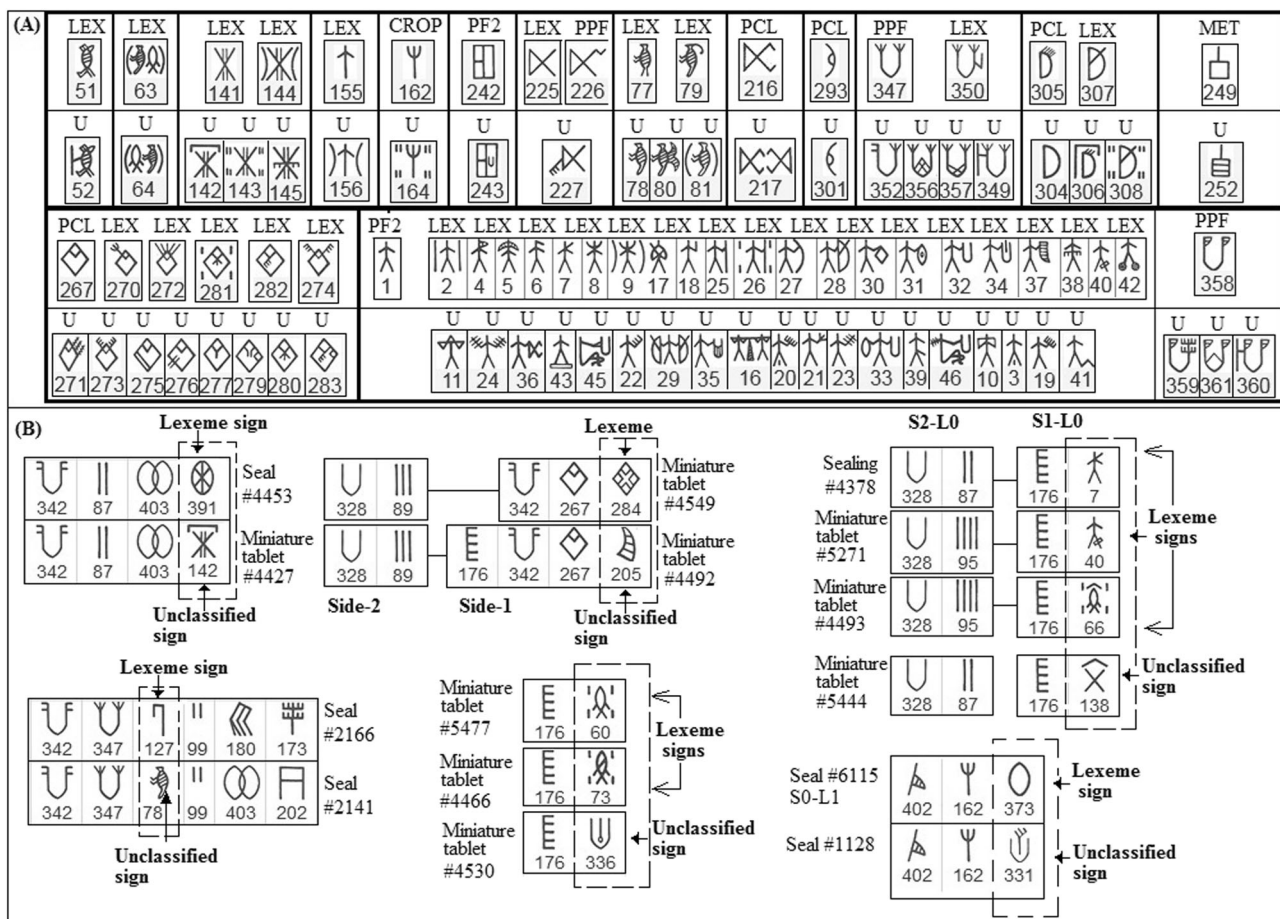of ISC that are really connected through some semantic relation. This study tries to exclude the "false collocations", where the adjacent signs do not really share any semantic link. For example, all of the following bigrams, i.e., ⬚⬚ (11 DILs), ⬚⬚ (10 DILs), and ⬚⬚ (19 DILs) are "false collocations", as in each of their occurrences, they were part of the trigrams ⬚⬚⬚, ⬚⬚⬚ and ⬚⬚⬚, where the bigram ⬚⬚ (101 DILs) was the "true collocation". Although, in this scenario, comparison of the relative bigram frequencies of a sequence could help in identifying the "true collocations", that method fails in certain cases. For example, the bigrams ⬚⬚ and ⬚⬚ have occurred in 39 DILs and 11 DILs, respectively. Thus, an unsupervised algorithm, modeled to consider relative bigram frequencies, would segment the inscription-line ⬚⬚⬚⬚ as ⬚⬚⬚⬚, which would be patently wrong. Being a PF1-sign, ⬚ is not an integral part of its preceding phrase. Contrarily, the lexeme-sign ⬚ has a pro-nounced affinity towards NUM-signs (⬚ is preceded by NUM-signs in 109 DILs, and ⬚⬚ is the eighth most frequent bigram of IDF-80). Thus knowledge of the Indus sign-classes helps us identify ⬚⬚ as the "true collocation", ignoring the misleading bigram frequencies.

Even among the "true collocations", this study further distinguishes between "general collocations", where certain signs frequently co-occur due to the semantic relationships shared by their sign-classes, and the "fixed collocations", where the co-occurrences are driven by the semantic bonds between the individual signs. For example, the frequent <PCL CM> constructs (⬚, ⬚, ⬚ etc.) or <PF1 PF2> constructs (⬚, ⬚, ⬚ etc.), are "general collocations", in which knowledge of the semantic affinity between the sign-classes (e.g., the affinity between PCLs and CM-signs) and knowledge of some such constructs, may help one guess more such constructs rightly. For example, by replacing sign ⬚ of ⬚ with another PCL-sign ⬚, one can surmise the existence of ⬚ (227 DILs). Contrarily, ⬚ (101 DILs), is a "fixed collocation", as here ⬚ occurs exclusively with a specific NUM-sign ⬚. Thus, replacing the ⬚ of ⬚ by any other stroke-numeral does not generate any such collocation that actually occurs.

With "general collocations" (<PCL CM>, <PF1 PF2>, <PPF PF1>, etc.) already explored, this section focuses on the "fixed collocations" present in ISC. To find these, I exclude all the bigrams that contain any PF1s or CM-signs, since such signs have certain fixed phrase-structural roles to play, and hence do not usually form fixed pairs with specific individual lexemes.

Unfortunately, as discussed above, no straightforward formula exists that can exclude all the "false collocations" found while parsing the inscriptions. However, this study proposes that, if the same sign-sequence occurs as a semantic unit in different inscriptional-contexts in different DILs, one can confidently identify it as a "true collocation", even if such a sequence occurs

Fig. 28 Bigram collocations occurring in different inscriptional contexts. Instances where the same collocation occurs: as the only inscription-content in inscribed objects (Column-1), alone with phrase-finals (Column-2), in pre-connective (Column-3) and post-connective constituents (Column-4), or as part of a longer sign-sequence in non-composite inscriptions (Column-5)

Fig. 29 A list of bigrams which occurred in at least 10 distinct inscription-lines. The count of DILs for each bigram and whether they are True (T) collocations or False (F) collocations is mentioned above each entry

in only 3 to 4 DILs. The possible "inscriptional contexts" are: (i) occurrence alone in an inscription (Fig. 28 column-1); (ii) occurrence alone with only PF-signs in an inscription (Fig. 28 column-2); (iii) occurrence in both pre-connective (Fig. 28 column-3) and post-connective constituents (Fig. 28 column-4) of different inscriptions; and (iv) occurrence as part of a non-composite inscription-line containing other sign-sequences (Fig. 28 column-5). A strong example of such a "true collocation" is the

moderately frequent bigram ⩕⩔ (12 DILs), which has occurred in all such different inscriptional contexts (see Fig. 28).

Figure 29 lists the 55 most frequent bigrams (excluding the ones containing PF1s and CMs) of ISC, sorted by their frequency in DILs. The 'F' and 'T' marks in the figure signify "False collocations" and "True collocations", respectively. However, among the infrequent bigrams (less than 10 DILs) not listed in Fig. 29, some are arguably "true fixed collocations", as they meet
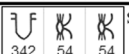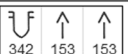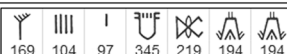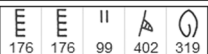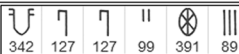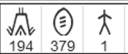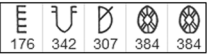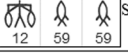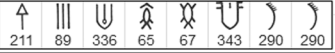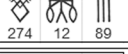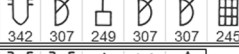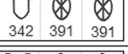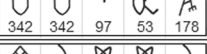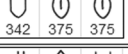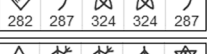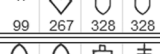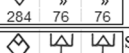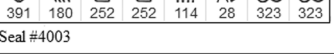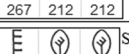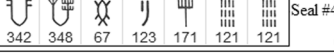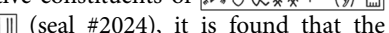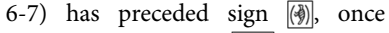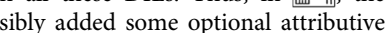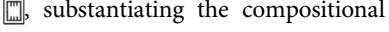
Fig. 30 A list of duplicated sign-sequences, with their count of occurrences and example inscriptions

the abovementioned criterion of occurrence as a unit in different inscriptional contexts (e.g., ⬚⬚ (6 DILs)).
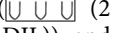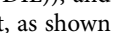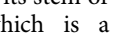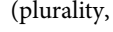
Compositional collocations and genitive constructs. In natural languages, compositional collocations are defined as collocations, whose meanings can be derived by combining the meanings of their shorter constituents. As analyzed below, many of the collocations of ISC are compositional in nature.

For example, comparing the occurrences of the collocation ⬚∥ (11 DILs) and its constituent sign ⬚, this study finds that their inscriptional contexts are sometimes very similar. For instance, comparing the pre-connective constituents of ⬚ (seal #1306) and ⬚ (seal #2024), it is found that the lexeme-sign ⬚ (Criteria 6-7) has preceded sign ⬚, once individually, and once as part of the collocation ⬚∥. Similarly, in inscriptions ⬚ (seal #6123) and ⬚ (seal #2137), the individual sign ⬚ and its collocation ⬚∥ precedes the CROP-sign ⬚ in similar contexts. So, the meaning of ⬚ must have been independently applicable in all these DILs. Thus, in ⬚∥, the collocate NUM-sign ∥ possibly added some optional attributive detail to the meaning of ⬚, substantiating the compositional nature of the collocation.

Moreover, many of the Indus collocations are constructs where NUM-signs or MET-signs precede certain CROP-signs (e.g., ⬚∥, ⬚∥, ⬚∥, ⬚, ⬚, etc.) or certain specific lexemes (e.g., ⬚∥, ⬚, ⬚, etc.). Since NUM-signs and MET-signs are numerical qualifiers that quantify their adjacent lexeme-signs in some way (Mahadevan, 1986), such collocations can be described as qualifier-qualified constructs, which are inherently compositional in nature.

However, distinguishing between the collocates as "qualifiers" and "qualifieds", may not always be so straightforward. For example, for collocation ⬚⬚, both collocates ⬚ and ⬚ have individually occurred as AWPF-lexemes in different inscriptions (e.g., ⬚⬚, ⬚⬚, ⬚⬚, ⬚⬚, etc.). Moreover, sign ⬚ has also occasionally functioned as a PCL sign, making determination of which one of them was attributive in nature quite difficult. I suggest that such collocations were possibly genitive constructs where one nominal sign qualifies the other, depending on their sequence.

Duplicated, triplicated and quadruplicated signs of ISC. The repeated sign-sequences of Indus inscriptions often demonstrate quite different combinatorial patterns compared to their non-repeated counterparts, which suggests that such sign-repetition was possibly some morphological tool used to introduce certain semantic changes. For example, unlike the PF2-sign ⬚, which typically follows PF1s, ENCs or CROP-signs in phrase-final positions, ⬚⬚ could occur as a complete post-connective constituent (⬚⬚ in seal #2347). Similarly, although sign ⬚ seldom occurs in phrase-final positions, ⬚⬚ frequently assume the role of PF1-signs (compare DILs like ⬚⬚⬚ and ⬚⬚, or constructs like ⬚⬚, ⬚⬚, and ⬚⬚⬚).

IDF-80 records only four triplicated sequences (⬚⬚⬚ (2 DILs), ⬚⬚⬚ (1 DIL), ⬚⬚⬚ (1 DIL), and ⬚⬚⬚ (1 DIL)), and one quadruplicated sequence (⬚⬚⬚⬚ (2 DILs)). But, as shown in Fig. 30, duplicated sign-sequences are quite frequent.

Repetition of an entire word, or partial repetition of its stem or root, is linguistically known as reduplication, which is a morphological device often used to denote "number (plurality,
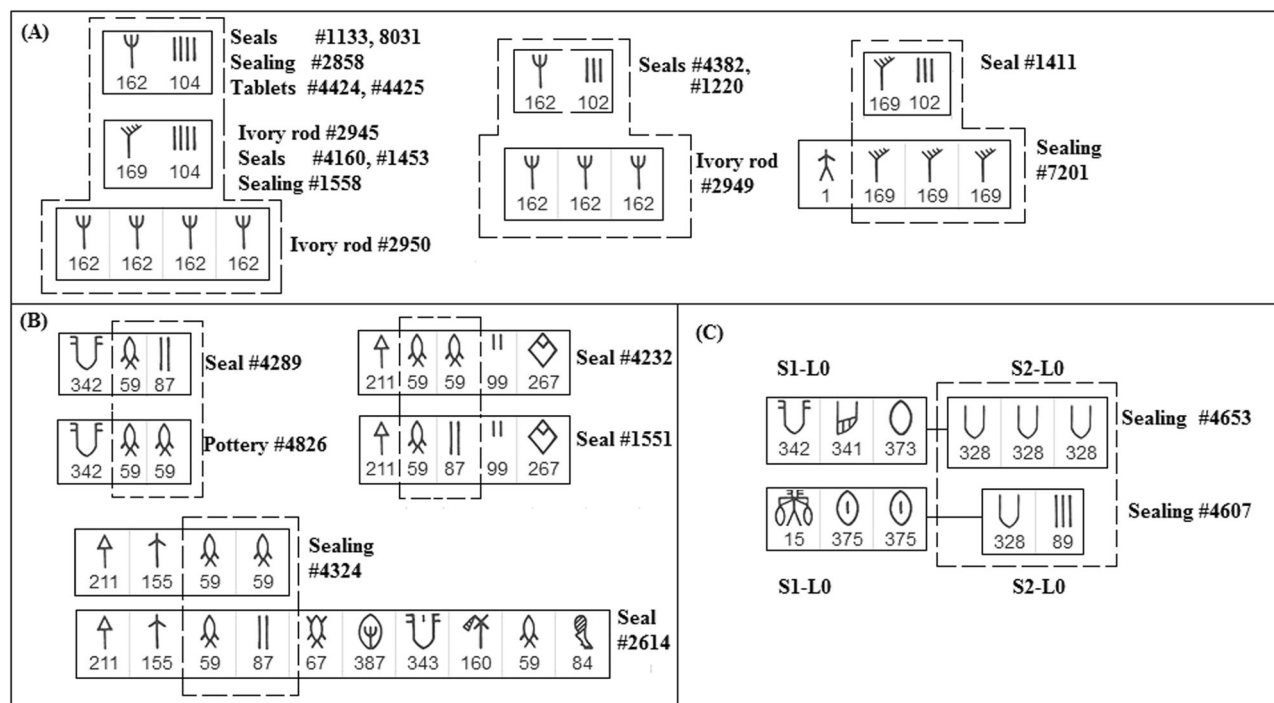
Fig. 31 Repeated-sign sequences and their corresponding <NUM LEXEME> constructs occurring in similar contexts. **a** Repeated CROP-signs; **b** Repeated fish-like signs; **c** Repeated rimless-jar signs

distribution, collectivity), distribution of an argument, tense, aspect (continued or repeated occurrence; completion; inchoativity), attenuation, intensity, transitivity (valence, object defocusing), or reciprocity" (Rubino, 2013). Several languages of Indian subcontinent use various forms of reduplication (Mohan, 2008). At this stage, it is difficult to identify all the functionalities served by the duplicated signs of ISC. But certain signs, particularly those which often get preceded by numerals, share, when repeated, intriguingly similar inscriptional contexts with their corresponding <NUM SIGN> constructs (see Fig. 31). For example, [sign] (where [sign] follows the two-stroked numeral [sign]) and [sign], share very similar inscriptional contexts (Fig. 31b), indicating that in such constructs, duplication possibly has served a similar quantifying role as the NUM-sign [sign] (Fig. 31a-c provide more such examples).

Visualizing the phrase-structures of Indus inscriptions. Having identified the lexemes, their sign-classes, and the collocations, we can now analyze how the inter-related sign-classes contributed in the process of making meanings through the inscriptions. I shall first describe a glossing method that helps visualize the phrase-structures, and then formulate certain rules to dissect the inscriptions into different semantic segments.

Glossing Indus inscriptions. To visualize the formulaic structures of Indus inscriptions, I programmatically parse the DILs of ISC (only completely undamaged inscription-lines), and apply a sign-by-sign glossing method to replace each sign, using the abbreviated name (PF1, PCL, etc.) of its sign-class. The lexeme-signs that are not categorized into any functional sign-class are simply glossed as LEX. Figure 32a demonstrates the step-by-step glossing procedure. Figure 32b shows three inscription-lines getting glossed through different steps. Now, CM-signs having been glossed before MET-signs, some of the metrological occurrences of the polyvalent signs [sign], [sign] and [sign] get wrongly glossed as CMs. Luckily, however, the inscriptional contexts can help to easily

distinguish their metrological occurrences from the connective ones.

Interestingly, using the procedure depicted in Fig. 32a, many distinct inscription-lines get glossed into same pattern groups (Fig. 33). For example, 57 DILs get glossed as <LEX PF1> s ([sign], [sign] etc.), 39 DILs as <LEX LEX PF1> s (e.g., [sign]), 37 DILs as <PCL CM LEX PF1> s ([sign], [sign] etc.) and 31 DILs as <PCL CM LEX LEX PF1> (e.g., [sign]). This proves that despite containing different signs, such DILs shared very similar structures.

Inscription segmentation techniques. Certain inscription-segmentation techniques naturally emerge from the results of the structural analysis. Each of the techniques is explained step-by-step, through segmenting a 14-signs-long inscription-line (seal #2654), which is one of the two longest inscription-lines recorded in IDF-80.

(A) Segmentation-Step1, using PFs: Since the PF1-signs and PF-clusters denote the syntactic and semantic boundaries of semantically complete phrases, they can be used to identify the shorter semantically complete messages (if any) present in an inscription-line.

Example: [inscription sequence] =>

Here using the PF1-signs [sign] and [sign], this longer inscription gets segmented into two semantically complete phrases.

(B) Segmentation-Step2, using CMs: If a CM-sign is present in an inscription-line, then the inscription-content present on either side of it can be separated out as pre-connective and post-connective constituents.

Example: [inscription sequence] =>

[segmented inscription]
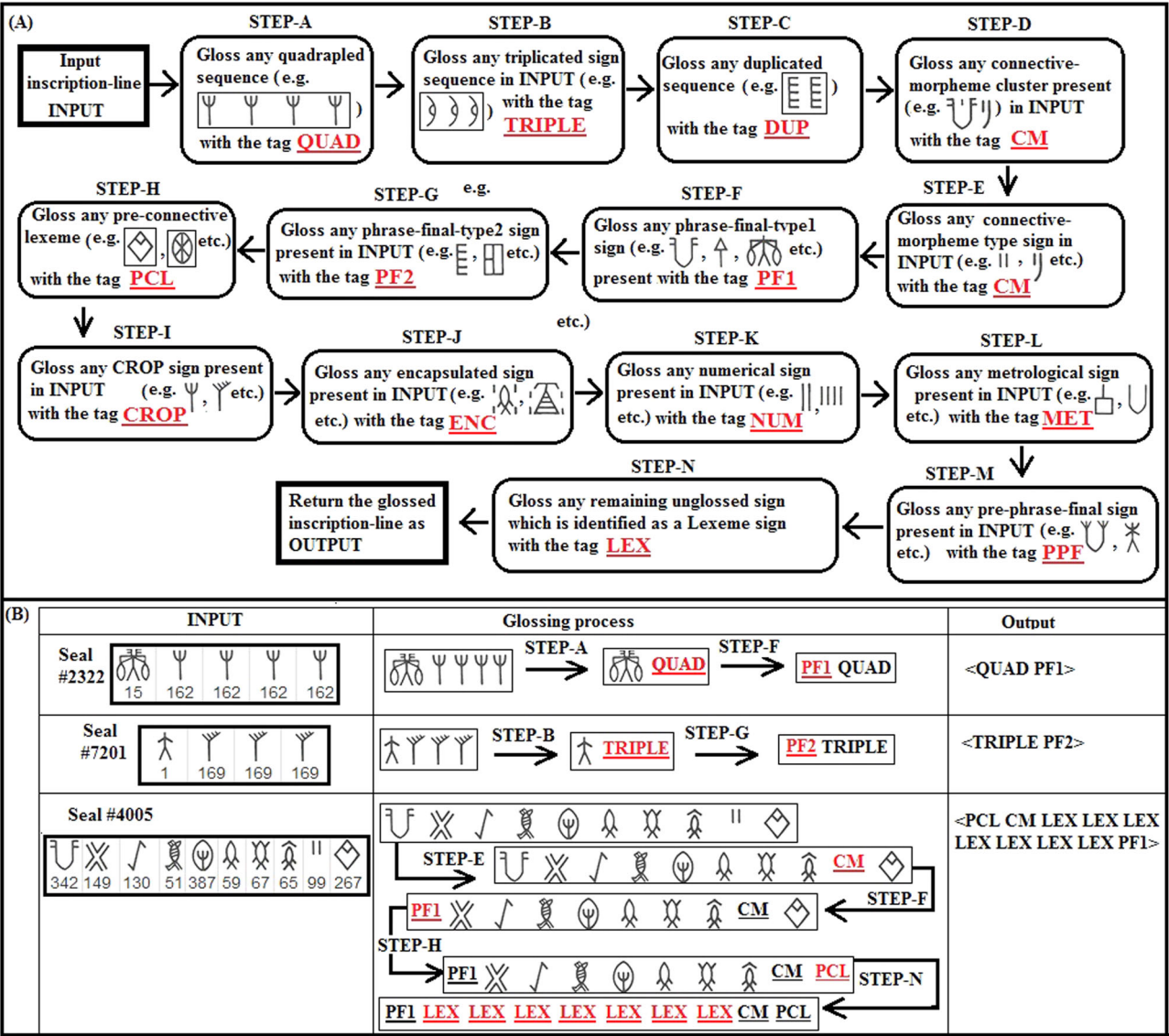
Fig. 32 Glossing Indus inscriptions. **a** A sequence chart of the glossing algorithm. **b** Examples of the glossed outputs at different stages. The newly glossed results of each step are highlighted in red

Here, the second semantic constituent is a composite inscription that got segmented into pre-connective and post-connective parts.

(C) Segmentation-Step3, collocations and repeated sequences: The collocations and repeated-signs present in the inscription-lines (if any) should be identified to mark the smaller semantic segments of the message.

Example:
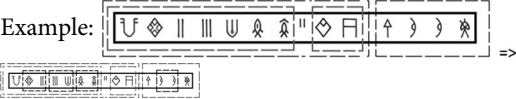
There are three bigram collocations and one duplicated sign-sequence in the above message.

(D) Glossing-STEP-4, glossing individual signs: Each individual sign of the inscription-line can be glossed with the abbreviated name of its sign-class to visualize the formulaic structure of each inscriptional segment.

Example:
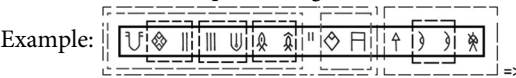
 Fig. 34 demonstrates the segmentation process of another long inscription.

Compositional semantics of Indus inscriptions. In linguistics, compositional semantics explores the ways through which the meaning of a phrase, a sentence, or a longer constituent is built using the meanings of its smaller semantic units. A demonstration of how it is possible to get a basic idea about the compositional semantics of the Indus inscriptions without inferring the meaning of even a single Indus sign follows.

The characteristic brevity of Indus inscriptions (around 70% of the 2409 DILs contain only one to five signs) is often reported as a major problem that incommodes the decipherment process. However, certain relatively short inscriptions can be extremely useful in understanding the compositional semantics of the longer ones. For example, if an inscription-line L1 occurs in Seal1, and consists of the sign-sequence <X Y Z>, and a longer inscription-line L2 occurs in Seal2, comprising the sequence <A B X Y Z>, then the part <X Y Z> in L2 can be safely separated out

Fig. 33 Representative examples of different patterns of glossed inscription-lines

as a semantically complete unit, as it was capable of conveying Seal1's complete message.

The examples in Fig. 35 bring out the above pattern. For example, all the <LEX PF1> constructs listed in Fig. 35a (↑ 🐾, ↑ 🐾, ⋃ 🌿, and ⋃ 🌀), are either complete inscription-lines or complete post-connective constituents, and their meanings must have been mainly informed by the meanings of the lexemes 🐾, 🐾, 🌿, and 🌀, since the accompanying PF1-signs convey meanings of much more generic scopes applicable to a large number of other inscriptions (see Criterion-6). For the convenience of discussion, let us name the information-contents of 🌀, 🐾, 🐾, and 🌿 as INF-A, INF-B, INF-C, and INF-D, respectively. Now, each of the semantically complete <LEX LEX PF1> constructs (⋃ 🌿🐾, 🖿 🌿🐾 and ↑ 🐾🐾) of the inscriptions listed in Fig. 35b, is made of pair-wise combinations of signs 🐾, 🐾 and 🌿, followed by a PF1-sign. Therefore, their meanings can be represented as INF-B and

INF-C (↑ 🐾🐾), INF-C and INF-D (🖿 🌿🐾), and INF-B and INF-D (⋃ 🌿🐾), other than the generic meanings of their respective PF1-signs. Similarly, the meaning of the <LEX LEX LEX LEX PF1> inscription ⋃ 🌿🐾🐾🌀 comprises INF-A, INF-B, INF-C, and INF-D, supplemented by the generic meaning of the most frequent PF1-sign ⋃ (see Fig. 35c). Since in each of the inscriptions the lexemes were used without any additional change to their graphemes, and no other sign occurred in between them, their juxtaposing technique possibly was an agglutinative one. All this evidence clearly indicates that often the longer inscriptions were simply composed of the information-units used in different smaller inscriptions.

The semantic compositionality of the 13-signs-long inscription-line ⋃ ⋈ ∫ 🐾 🐾 ‖ ⋈ 🐾 " Y 🕴 ▥ ‖ can be analyzed by applying this same principle. As already demonstrated in Fig. 4, its post-connective constituent is made of two bigram

Fig. 34 Segmentation tree of a 13 signs-long inscription **a** and its final glossed form **b**

collocations (⊠∥ and ⊠⊠) and one trigram collocation (⊠ʃ⊠), each of which has occurred in other inscriptions as their main semantic content (see Fig. 36). Similarly, the pre-connective constituent of ⊔⊠ʃ⊠⊠∥⊠⊠"Ψ⊠⊡∥ consists of two bigram collocations (Ψ⊠ and ⊡∥), each of which has typically occurred as semantic units in pre-connective parts of other inscriptions. Moreover, both Ψ⊠ and ⊡∥ occur together in the pre-connective part of inscription ⊠⊠⊠"Ψ⊠⊡∥⊠ in seal #2018. Thus even the message of the second longest inscription-line of IDF-80 is merely composed of many shorter messages. Analyzing many such examples, a generalization can be safely ventured that the longer Indus inscriptions were structurally no different than the shorter inscriptions, as they just contained more units of information, not different types of information.

We could, thus, get a good idea of the semantic composition-ality of Indus inscriptions without ascribing any meaning or sound to its constituent signs.

Co-occurrence restriction patterns maintained in Indus phrases. "Co-occurrence restrictions", both syntactic and semantic, in the context of lexical affinity, lexical repulsion, and grammaticality, is an oft-quoted term in linguistics (Cruse, 1986). Interestingly Indus sign-classes too evince various forms of co-occurrence restrictions.

For example, in IDF-80:

- Not a single inscription-line contains more than one PPF-sign.
- The PF1-signs rarely occur adjacent to each other. Only 9 DILs contain <PF1 PF1> constructs such as ⊠⊔, ⊠⊔, and ⊟⊔. Moreover, multiple non-adjacent PF1-signs seldom occur inside the same semantic unit.
- Only 8 DILs contain <PF2 PF2> sequences (⊠目, 目⊠, ⊡⊠, 目⊠). Moreover, multiple non-adjacent PF2-signs rarely occur in an inscription-line.
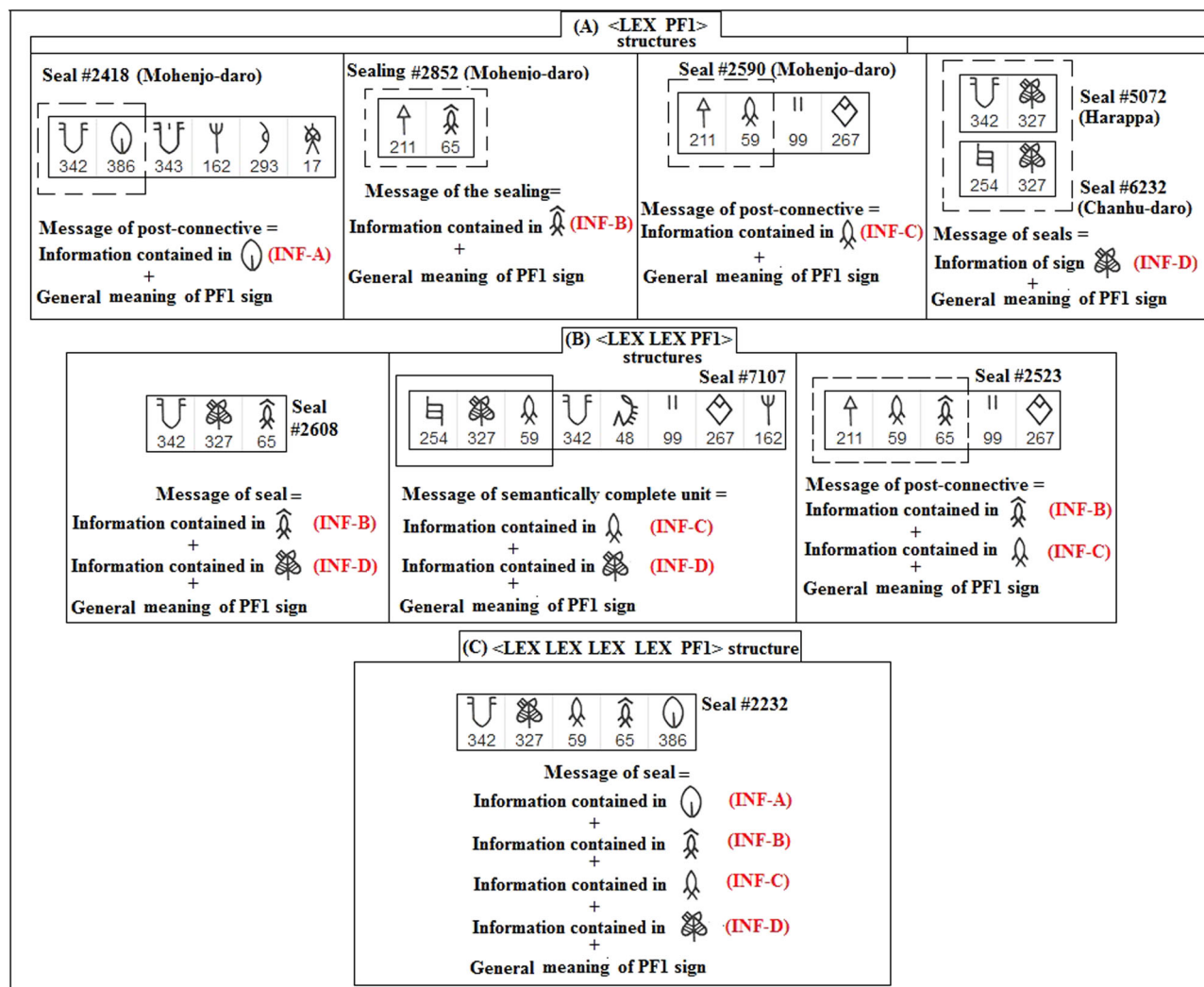
Fig. 35 Demonstrating how the meanings of certain longer Indus inscriptions **b**, **c** were made of informational units present in smaller inscriptions **a**

- Only 1 DIL contains more than one ENCs (⬚), and that too in separate semantic units.

These patterns very strongly suggest that a single Indus message could logically contain maximum one value from the semantic scope of certain sign-classes. This scenario is comparable to that of the stamps of Fig. 2a, where each stamp could contain only one monetary unit ("Anna", "Rupee" etc.). However, occasionally multiple units might co-occur to express mixed values like "1 Rupee 25 Paise", comparable to the infrequent occurrences of <PF1 PF1> and <PF2 PF2> constructs.

One of the interesting co-occurrence restriction patterns existing between different sign-classes is that, not a single inscription exists where a PPF-sign has preceded an ENC-sign. As ENCs contain the semantics of PF1-signs too (see Section "Encapsulated (ENC) signs"), the way PF2-signs follow ENCs (Fig. 25e), PPF-signs should expectedly precede them in some inscriptions. Therefore this complete absence of <PPF ENC> constructs indicates that ENCs included the semantic role of PPF-signs too, rendering their presence redundant.

These co-occurrence restriction patterns may prove to be crucial clues for understanding certain semantic aspects of Indus inscriptions.

The logographic nature of Indus inscriptions. Since 254 lexeme signs are already identified, and the high probability that many unclassified signs were lexemes is already discussed, adding another section discussing the logographic nature of ISC may seem apparently unnecessary. Yet, since many scholars continue to believe that a significant number of Indus signs have functioned as phonograms, this point needs pressing from diverse perspectives.

Co-occurrence restriction patterns reject the phonogram hypothesis. A very compelling, nearly unassailable proof of the logographic nature of Indus inscriptions comes from the co-occurrence restriction patterns maintained in them. Various natural languages across the world use different co-occurrence restriction patterns in their phonological, as well as grammatical and lexical constructs. But phonological co-occurrence restrictions, being mostly based on "'articulatory economy', 'auditory contrast' and 'articulatory-acoustic stability'", are completely different in nature from their semantic counterparts, which are in turn influenced by the "physical and physiological properties of the speech production and perception systems" (Solé, 1999). This is why phonological co-occurrence restrictions pertain to the

Fig. 36 Compositional semantics of a 13-signs-long inscription-line

locales of syllables, morphemes and small words only, seldom operating in larger domains of phrases or sentences. Often phonemes that cannot co-occur in a syllable can appear in the root and suffix of a polysyllabic word (MacEachern, 1999, p. 28). Contrastively, semantic co-occurrence restrictions, originating in needs of logical compatibility between different linguistic elements, operate at the levels of collocations, phrases, sentences and even discourse (Cruse, 1986, p. 103–104, p. 277–279).

The co-occurrence restriction patterns of ISC are intriguing. For example, while the CROP-signs and individually occur in 162 and 137 DILs, respectively, they have not co-existed in the same DIL even once. Moreover, certain signs such as and have occurred exclusively with and its ligature , in artifacts found from various Indus locations. Since and occur in inscriptions found from similar stratigraphic levels of same Indus locations (Mahadevan, 1977), they were contemporary signs. Thus, such special affinity between , and proves that and must have been separate signs, not allographs. Then, if CROP-signs were phonograms, what stopped them from co-occurring in different positions of the same inscription even once? Similarly, if the PPF-signs occurring in 309 DILs were phonograms, why did not a single DIL contain multiple PPFs? Phonological co-occurrence restrictions might restrict adjacency, but cannot operate in inscriptional domain for longer inscriptions.

The co-occurrence restrictions between signs that can sometimes occur adjacently are even more revelatory. For example, among the 326 DILs containing PF2s, only 10 contain two PF2-signs, among which 6 DILs contain <PF2 PF2> constructs. Similarly while many <CM CM> constructs exist, only two DILs contain both of and . Finally, among the 1338 DILs containing PF1s,

only 41 have two PF1s, while just one DIL contains three PF1s. The general lexeme-signs also demonstrate the same occurrence-patterns as the functional sign-classes. For example, among the 288 DILs where occurs, only 11 DILs contain two non-adjacent signs, while 7 have its reduplicated form . None of the other fish-like signs (, , , etc.) occur more than once in the same DIL, whereas they often co-occur adjacent to each other. Similarly, PCL-sign occurs in 167 DILs. But excluding the 5 DILs with , only 3 DILs contain more than one sign. Thus, neither the PCLs and Fish-signs, nor the PF1s, PF2s, CROP-signs, ENCs, CMs, and PPFs were phonograms, since they all demonstrated inscription-level co-occurrence restriction patterns, while adjacency was occasionally allowed for them, which evidently violates the rules of phonological co-occurrence restriction. Analyzing such "low sign-repetition rates in individual inscriptions", Farmer et al. (2004) have earlier argued that "little if any sound encoding existed" in Indus inscriptions.

The longer inscriptions make the phonogram hypothesis about PF1s appear even more absurd. For example, analyzing the long seal-inscription (#1087), and the shorter seal-inscriptions, (#8001), (#2549), (#4289), (#4285), (#2269), and (#3228), we find that these shorter inscriptions are formed using the marked smaller constituents of . Now, if these constituents were phonetically constructed, it is startling to see that each short inscription needed the help of the supposed PF1 phonograms to complete their supposed word-sounds. Then how is it possible that such a long inscription needed the PF1-sound only at the end, nowhere before? No natural language can suffer such skewed sounds for their words.

The triplicated and quadruplicated sequences of ISC (⊔⊔⊔, ⋏⋎⋎⋎, ⋎⋎⋎⋎, etc.), further buttress my arguments against the phonogram-hypothesis, as Indic languages seldom contain triplicated or quadrupled phonemes to form a word.

Countering the hypothesis that logograms and phonograms co-existed in ISC. Some scholars argue athwart all this evidence that ISC has used a mixed system of writing, where both logograms and phonograms co-existed (Wells, 2015, p. 53, p. 71). Truly, certain ancient scripts (e.g., Egyptian and Maya Hieroglyphs) have used both phonetic signs and lexemes in the same texts. But, the coexistence of such signs always maintained specific pre-defined rules in such scripts. For example, the Egyptian Hieroglyphs used primarily phonetic spellings, comprising mostly consonant signs, while their accompanying logograms/ideograms functioned as semantic complements that distinguished between homophonous sounds. On the other hand, Maya Hieroglyphs primarily used logograms, which were either followed or preceded by phonetic complements to indicate the pronunciations of the words (Mora-Marín, 2008, p. 195–213). Thus, if Indus Inscriptions contained mixed writing, we should have seen two conspicuously different sign-classes representing the logograms and phonograms. Now, as discussed before, since the PF-signs are syntactically detachable from their preceding sequences, and demonstrate a distinct phrase-level role, they show the most conspicuous combinatorial contrast compared to other signs. Thus, if the PFs are phonograms, their preceding sign-sequences must be logograms and vice versa. But as aforestated, PFs were not phonograms (Section "Co-occurrence restriction patterns reject the phonogram hypothesis"), but logograms (Section "Phrase-final (PF) signs and their subcategories"). Thus, the sign-sequences preceding the PFs should complementarily be phonograms. But, as shown in Section "Co-occurrence restriction patterns reject the phonogram hypothesis", the CROP-signs, CMs, PPFs, ENCs, PCLs and many other general signs (e.g., the fish-like signs) are also not phonograms. So, the inscriptions that are constituted mostly with these signs cannot be instances of mixed writing.

Strict positional preferences and co-occurrence preference patterns disprove the phonogram hypothesis. For a proof by contradiction, let us first assume that Indus inscriptions were phonetically written. Next, let us analyze the most frequent 30 signs, each of which has occurred in more than 80 DILs. These 30 signs (sorted below in decreasing order of frequency) have collectively occurred in 87% of the 2409 DILs.

Among these signs, the PF-signs ⊔, ⋏, ⋏, and 𝔼, and the PPF-sign ⊍ are reputed for their rigid preferences for terminal and pre-phrase-final positions. The CM-signs ⟋, ⟍, ⋏, and ⊍ are mainly located in between two semantically complete constituents. The PCLs ◇, ⊛ and ⟩ dominantly occur in the initial positions of pre-connective constituents. CROP-signs ⋎ and ⋎ mostly occur glued to specific NUM-signs and MET-signs such as ⫴, ⊔ etc. The other frequent signs, such as ⊍, ⋏, ⋏, ⋈, ⋎, ⊤ etc. very often occur as part of fixed collocations (e.g., ⫴⊍, ⋏⫴, ⋏⊤, ⋈⋈, ⟋⋎). Therefore, if these signs are to spell out words phonetically, words of a grievously restricted phonetic range would result, which is simply unacceptable for the words of any natural language. Thus the hypothesis that Indus inscriptions were phonetically constructed cannot be accepted, at least not for most of the inscriptions excavated till date.

How the logogram model explains the structural peculiarities of ISC. It can do it for all the structural features seen in ISC, as argued below.

Co-occurrence restrictions: As established in Section "Contextualising the formalized data-carriers of IVC", the semantic domain of Indus inscriptions was dominantly associated with some formalized data-carriers and metrological devices that were used in specific commercial processes of IVC. Now, the inscription-level co-occurrence restrictions demonstrated between the members of the functional sign-classes of ISC suggest that the nature of the inscriptions' messages was probably such that, generally only one member of a certain sign-class (e.g., PPFs) could apply to one message. This pattern is characteristic to the texts of various modern data-carriers and metrological devices. For example, a modern container that measures some liquid will only contain metrological units like litre/milliliter, whereas a dry measure weight will be inscribed with units like pound, gram etc.

Some of the reduplicated sign-sequences possibly were special morphological units, whose meanings were derived from the meaning of the basic signs using some specific rules. Thus sign-adjacency was permitted for certain sign-classes, whereas inscription-level co-occurrences were prohibited for them.

Co-occurrence preferences: The co-occurrence preferences of specific collocates quite obviously indicate that certain lexemes were semantically more compatible to each other in the semantic scope of the inscriptions. As discussed before, Indus collocations were compositional in nature, signifying that certain attributive lexemes were more applicable to certain substantive lexemes, leading to the formation of fixed collocations.

Compositionality of longer inscriptions: The longer inscriptions found in certain formalized data-carriers are expected to demonstrate semantic compositionality by getting constructed with semantic units present in smaller inscriptions present in other data-carriers (see Fig. 35). For example, the texts used in different ration-tokens of a country can be (i) "Meat" (ii) "Fats" (iii) "Cheeses" (iv) "Sugar" (v) "Fish" etc. Now, another token may contain a longer text such as "Meat, Fats, Fish, and Cheeses", which basically conjoins some related rationed items that may individually occur in other tokens.

Positional preferences: The positional preferences of certain signs might be a simple document-specific format. For example, in the stamps shown in Fig. 2a, the phrases like "Share transfer" or "India Non Judicial", occur in the bottommost parts and denote the type of the stamps. This positional preference is just document-specific, not bound by any linguistic rules. Similarly, the positional preferences of PF1s and PF2s might have been a document-specific convention maintained in Indus seals and tablets.

Order of signs: Unlike certain possible document-specific formats, the syntactic orders maintained in the bigram collocations (e.g., the <NUM CROP> constructs) seem to be influenced by linguistic rules. For example, the languages that use prenominal adjectives generally place qualifying morphemes before the qualified morphemes. Interestingly, prenominal adjectives and <qualifier qualified> constructs find use in most of the Indic languages.

Conclusion

The most important contribution of this study should possibly be that, if a researcher agrees with its results, s/he would no longer try to treat Indus signs as phonograms in order to spell out words. Moreover, since the inscribed objects are identified as formalized data-carriers, in which linguistic syntaxes and document-specific syntaxes can play equally important roles, a researcher would not have to explain each syntactic feature from a linguistic and

grammatical aspect. The focus of the future semantic analysis should be on understanding the semantic role of each of the functional sign-classes and the reason behind their inter-dependence. For example, the reasons behind the relatedness of the MET-signs and PF1s, the substantive-type occurrences of the NUM-signs, the semantic relationship between pre-connective and post-connective constituents, the very restrictive usages of the NUM-signs and the high probability of the <NUM CROP> collocations, should be seriously explored in any future semantic analysis. Often in logographic writing systems, the grapheme chosen for a logogram resembles the real world objects which symbolize the semantic concept of the logogram (though over time those graphemes might go beyond recognition by getting more stylized and abstract). Since many Indus signs are quite pictorial in nature (e.g., some versions of sign ▦ clearly demonstrate a man bearing loads on a shoulder yoke (see CISI seal H-1046)), it might be possible to trace some signs back to the concepts/objects they symbolized. Since the archeological evidence strongly suggests that ISC was used in some highly standardized socio-economic activity of ancient Indus life, one should explore the functionality of the sign-classes and investigate whether most of the graphemes used for the logo-grams of a functional sign-class are related to some particular socio-economic symbolic dealing. In this context, the historical evidence extracted from the earliest available literatures of ancient India should be thoroughly analyzed. Additionally, the archeological and historical evidence obtained from the civilizations which were ancient trade partners of Indus valley (e.g., the ancient Mesopotamian civilization), and the archeological, linguistic and historical evidence found from the civilizations of the Bactria–Margiana Archeological Complex which were related to Indus civilization in several interesting ways, should also be consulted.

Data availability

The digitized corpus of Indus inscriptions (Mahadevan, 1977), enhanced in 1980, is maintained by the IndusResearch Centre of Roja Muthiah Research Library Trust (http://rmrl.in), and is available to any *bona fide* scholar upon request.

References

Bhan K K (2011) Review of Prehistoric Cultures of Gujarat and Need to Develop Accurate Settlement Gazetteer. In: Uno T (ed) Changing Perspective of Japan in South Asia in the New Asian Era: The State of Japanese Studies in India and other SAARC countries. International Research Center for Japanese Studies: Kyoto, p 325–354

Blühdorn H (2010) A semantic typology of sentence connectives. In: Harden T, Hentschel E (eds) 40 jahre partikelforschung. Stauffenburg, Tübingen, p 215–231

Bonta S (2010) The Indus Valley script: a new interpretation. Pennsylvania State University, Altoona

Chrisomalis S (2010) Numerical notation: A comparative history. Cambridge University Press, New York

Chomsky N (1995) The minimalist program. The MIT Press: Cambridge Massachusetts: London, England

Cruse DA (1986) Lexical semantics. Cambridge University Press, Cambridge

Dales GF (1968) Of dice and men. J Am Orient Soc 88(1):14–23

Oxford Dictionaries (2016) English oxford living dictionaries. https://en.oxforddictionaries.com

Farmer S, Sproat R, Witzel M (2004) The collapse of the Indus script thesis: the myth of a literate Harappan civilization. Electron J Vedic Stud 11(2):19–57

Floridi L (2015) Semantic conceptions of information. First published October 5, 2005; substantive revision January 7, 2015. https://seop.illc.uva.nl/entries/information-semantic/

Fuls A (2010) Entwicklung einer geographisch-epigraphischen Datenbank der Indusschrift. In: Weisbrich S and Kaden R (eds) Entwicklerforum Geodäsie und Geoinformationstechnik 2010. Shaker Verlag: Aachen, p 29–45

Gardiner SA (1969) Egyptian grammar: Being an introduction to the study of hieroglyphs. 3rd revised ed. Griffith Institute, Ashmolean Museum, Oxford

Hunter GR (1934) The script of Harappa and Mohenjodaro and its connection with other scripts. Kegan Paul, Trench, Trubner & Co, London

Joshi JP, Parpola A (eds) (1987) Corpus of Indus seals and inscriptions. 1. Collections in India. Memoirs of the Archaeological Survey of India, No. 86. Suomalainen Tiedeakatemia, Helsinki

Kenoyer JM (2005) Steatite and faience manufacturing at Harappa: new evidence from Mound E excavations 2000-2001. Mus J 3(4):43–56

Kenoyer JM (2006) The origin, context and function of the Indus script: Recent insights from Harappa. In: Osada T, Hase N(eds) Proceedings of the pre-symposium and the 7th ESCA Harvard-Kyoto Roundtable. Research Institute for Humanity and Nature, Kyoto, p 9–27

Kenoyer JM (2010) Measuring the Harappan world: insights into the Indus order and cosmology. In: Iain Morley, Colin Renfrew (eds) The archaeology of measurement: comprehending heaven, earth and time in ancient societies. Cambridge University Press: Cambridge, New York, p 106–22

Kober AE (1948) The Minoan scripts: fact and theory. Am J Archaeol 52(1):82–103

Koskenniemi K, Parpola A (1982) A Concordance to the Texts in the Indus Script. Department of Asian and African Studies, University of Helsinki, Research Reports, no. 3, Helsinki

MacEachern MR (1999) Laryngeal Cooccurrence Restrictions. Garland Publishing: New York & London

Mackay E (1931) Seals, seal impressions, and copper tablets, with tabulation. In: Marshall JH (ed) Mohenjo-daro and the Indus Civilization. Arthur Probsthain, London, p 370–405

Mahadevan I (1977) The Indus script: texts, concordance and tables. Memoirs of the Archaeological Survey of India No. 77

Mahadevan I (1978) Recent advances in the study of the Indus script. Puratattva 9:34–42

Mahadevan I (1986) Towards a grammar of the Indus texts: 'intelligible to the eye, if not to the ears'. Tamil Civiliz 4(3–4):15–30

Mahadevan I (2014) Dravidian proof of the Indus script via the Rig Veda: a case study. Bulletin of the Indus Research Centre, Roja Muthiah Research Library, Chennai

Millard AR (1988) The bevelled-rim bowls: their purpose and significance. Iraq 50:49–57

Mohan S (2008) Echo Word Formation In Indian Languages: A Typological Study. Bulletin of the Deccan College Research Institute 68:329–339

Mora-Marín DF (2008) Full phonetic complementation, semantic classifiers, and semantic determinatives in ancient Mayan hieroglyphic writing. Anc Mesoam 19(2):195–213

Nissen HJ, Damerow P, Englund RK (1993) Archaic bookkeeping: early writing and techniques of economic administration in the ancient Near East. (trans: Larsen P). University of Chicago Press, London

Nissen HJ (2011) The early history of the ancient Near East, 9000-2000 BC. (trans: Lutzeier E, Northcott K J). University of Chicago Press, Chicago, London

Parpola A (1994) Deciphering the Indus script. Cambridge University Press: New York

Parpola A, Pande BM, Koskikallio P (eds) (2010). Corpus of Indus seals and inscriptions. 3.1. Supplement to Mohenjo-daro and Harappa. Memories of the Archaeological Survey of India, Vol. 96. Suomalainen Tiedeakatemia, Helsinki

Possehl GL (1999) Indus age: the beginnings. University of Pennsylvania Press, Philadelphia

Possehl GL (2002) The Indus civilization: a contemporary perspective. AltaMira Press, Walnut Creek

Rubino C (2013) Reduplication. In: Haspelmath M, Dryer MS, Gil D, Comrie B (eds) The world atlas of language structures online. Max Planck Institute for Evolutionary Anthropology, Leipzig

Shah SGM, Parpola A (eds) (1991) Corpus of Indus seals and inscriptions. 2. Collections in Pakistan. Memoirs of the Department of Archaeology and Museum, Governinent of Pakistan, Vol. 5. Suomalainen Tiedeakatemia, Helsinki

Shannon CE (1948) A mathematical theory of communication. Bell Syst Tech J 27:379–423

Solé MJ (1999) The phonetic basis of phonological structure: the role of aerodynamic factors. In: Proceedings of the Ist Congress of Experimental Phonetics. Tarragona, Spain, p 77–94

Thapar R (2015) The Penguin history of early India: from the origins to AD 1300. Penguin, London

Ventris M, Chadwick J (1953) Documents in Mycenaean Greek. Cambridge University Press: Cambridge

Wells BK, Fuls A (2006) Online Indus Writing Database: Berlin. http://caddy.igg. tu-berlin.de/indus/welcome.htm. Accessed Aug 2016

Wells BK (2011) Epigraphic approaches to Indus writing. Oxbow Books, Oxford, Oakville

Wells BK (2015) Archaeology and epigraphy of Indus writing. Archaeopress, Oxford

Yadav N (2013) Sensitivity of Indus script to site and type of object. Scripta 5:67–103

Additional information

The online version of this article (https://doi.org/10.1057/s41599-019-0274-1) contains supplementary material, which is available to authorized users.

Competing interests: The author declares no competing interests.

Reprints and permission information is available online at http://www.nature.com/reprints

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.