Deciphering the Indus Valley Script: Hierarchical Clustering and Dependency Tree Analysis

R. Geetha Ramani, PhD
Professor
Department of Information Technology,
Madras Institute of Technology
Anna University,
Chennai 600044, India

Kishor S.
Department of
Information Technology,
Madras Institute of Technology
Anna University,
Chennai 600044, India

ABSTRACT
This study introduces a pioneering methodology aimed at deciphering the enigmatic Indus script, leveraging advanced computational linguistic techniques in conjunction with the Interactive Corpus of Indus Texts (ICIT). The approach utilizes hierarchical clustering and dependency tree construction to organize symbols based on visual similarities and uncover grammatical structures within sequences of Indus script symbols. Through meticulous analysis, nuanced insights into the distribution, significance, and syntactic relationships of symbols are revealed. Notably, the investigation identifies patterns in frequently occurring combinations and their positional arrangements. Additionally, the study explores network properties of symbols, elucidating their centrality and contextual versatility within the script's semantic framework. Employing probabilistic models, the research investigates the predictive potential of positional patterns in reconstructing missing symbols, offering promising avenues for future decryption endeavors. Furthermore, multi-sequence alignment is utilized to identify underlying grammar patterns within the language. This innovative approach, coupled with substantive findings, contributes significantly to ongoing decipherment endeavors, shedding new light on one of humanity's oldest linguistic mysteries.

Keywords
Indus script, Deciphering, Hierarchical Clustering, Grammar Patterns.

1. INTRODUCTION
The study of the Indus script, originating from the ancient Indus Valley Civilization, presents an enduring challenge in the field of linguistics and archaeology. Dating back to the 3rd millennium BCE, this script has intrigued scholars for generations with its elusive nature (Chakrabarti & Lal, 2007; Kenoyer, 2010). This research endeavors to delve into the mysteries of the Indus script through a multifaceted approach that combines hierarchical clustering and grammar analysis using dependency trees.

Hierarchical clustering serves as a foundational element of this approach, enabling the systematic categorization of Indus script symbols based on their visual characteristics. By grouping symbols into clusters, the aim is to uncover inherent patterns and structural regularities within the script. Additionally, grammar analysis employing dependency trees offers insights into the linguistic structures embedded within sequences of Indus script symbols. Through meticulous examination, the objective is to discern grammatical rules, syntactic constructions, and semantic associations, thereby shedding light on the script's purpose and meaning.

Drawing upon insights from various fields, such as linguistics, archaeology, and computer science, this interdisciplinary endeavor seeks to enrich our understanding of the cultural context and significance of the Indus script. Archaeological findings provide essential context regarding the script's origin, usage, and cultural practices, while advancements in computer science offer sophisticated analytical tools for deciphering complex patterns within the script.

The enduring mystery of the Indus script serves as a testament to the rich cultural heritage of the ancient Indus Valley Civilization. Despite centuries of scholarly inquiry, the script continues to defy complete decipherment. Nevertheless, it offers valuable insights into the social organization, economic activities, and religious beliefs of this ancient civilization. The ongoing pursuit of deciphering the Indus script holds immense potential to unlock a wealth of knowledge about the Indus Valley Civilization, its language, literature, and interactions with neighboring cultures, contributing to our understanding of ancient civilizations and human history as a whole.

2. RELATED WORK
This section outlines our comprehensive approach, integrating Hierarchical Clustering, Dependency Tree Analysis, and Symbolic Pattern Recognition to decode the Indus scripts. While Hierarchical Clustering and Dependency Tree Analysis unveil structural insights and grammatical rules, Symbolic Pattern Recognition offers a nuanced understanding of the symbolic representations within the scripts.

2.1 Hierarchical Clustering
In the study of deciphering the Indus Script, scholars analyze connections between symbols, language, and historical context. Each analysis highlights the need for thorough data and advanced methods to improve accuracy.

Harakawa, R., Ohtomo, K., Takahashi, K., Matsuda, Y., & Iwahashi, M. e**t.al,[1],** presents a hierarchical clustering method for natural disaster images. By displaying natural disaster images with similar categories, users are able to intuitively understand their surrounding environment and potential risks such as flood and sediment damage. Because there are unstructured natural disaster images without textual descriptions, a method that does not use textual features is necessary. The contribution of this study is to newly develop a method to hierarchically cluster natural disaster images using only visual features. Experimental results for images that are available from the research database show the superiority of our hierarchical clustering compared with flat clustering.

Gherbaoui, R., Benamrane, N., & Ouali, M. e**t.al,[2],** introduces a novel hierarchical clustering algorithm for color image segmentation. Traditional clustering methods often struggle with datasets featuring elliptical and chained shapes. Our algorithm overcomes these limitations by quantifying the degree of overlap between clusters, facilitating a more accurate merging process.

Liu, J., Wang, D., Yu, S., Li, X., Han, Z., & Tang, Y e**t.al,[3],** provide a comprehensive overview of image clustering. Specifically, we first discuss the applications of image clustering across various domains. Then, we summarize the common algorithms and propose a classification of image clustering. The existing methods are classified from four aspects: autoencoder based methods, subspace clustering, graph convolution network (GCN) based methods, and some other clustering methods. We introduce the main research contents and existing problems of various image clustering methods. We also introduce some recent methods and summarize the experimental results. Based on our taxonomy and analysis, creating and verifying new methods is more straightforward. Finally, we propose the future opportunities in this fast-developing field.

2.2 Dependency Tree Analysis

Yu, T., & Gao, S et.al,[4], aims to refine text content efficiently. Traditional methods often treat text as a sequence and employ encoders such as RNN to learn text semantic features. However, these approaches struggle to utilize the grammatical structure information implicit in the text, leading to lower summarization accuracy. In this paper, we propose a novel approach that utilizes dependency graphs to represent text and employs Gated Graph Neural Networks as encoders. This allows for easier learning of long-distance dependencies between nodes in the graph. We extract feature vectors from the text dependency graph based on the hidden states of nodes and subsequently utilize a Transformer decoder to generate text summarization. Experimental results demonstrate that our model outperforms selected comparison models in terms of performance.

Liu, G., Wang, K., Liu, W., & Cao, Y et.al,[5], deals with requirements described in natural language, traditional automated requirement checking models often face feasibility challenges. To address this issue and verify the consistency of information system requirements, this paper proposes a semantic model utilizing tree nodes representing natural language clauses. The model divides clauses into a representation of keyword sets using a seven-tuple approach. Additionally, the paper introduces a dependency tree model to enhance the characterization of relationships between syntactic structure and keywords, along with a dependency tagging algorithm and an algorithm for constructing and updating dependency parsing trees. Furthermore, a semantic similarity calculation method is proposed to determine similarity among sub-clause syntactic structures.

3. METHODOLOGY

The decipherment of the Indus script requires a meticulous and multi-faceted approach that combines traditional archaeological methods with cutting-edge computational techniques. In this section, we outline the methodologies utilized in our research, focusing on hierarchical clustering and dependency tree analysis. These methodologies offer complementary insights into the structural and semantic properties of the script, providing valuable avenues for unraveling its linguistic and cultural significance.

3.1 System Architecture

This section deals with the architecture diagrams of the proposed solutions and the explanation as to how they fit in our problem.

3.1.1 Hierarchical Clustering

The analysis commenced with the assembly of a dataset containing symbol sequences extracted from the Indus script corpus. Hierarchical clustering emerged as a foundational technique aimed at grouping similar symbols based on their structural characteristics. This method involved iteratively merging symbol clusters to construct a hierarchical tree-like structure, thus revealing inherent relationships and similarities among symbols.

Hierarchical clustering stands as a versatile tool extensively utilized in data analysis and pattern recognition. Its operation revolves around the successive merging of data point clusters, guided by pairwise similarity or dissimilarity metrics. This iterative process culminates in a hierarchical arrangement of clusters, depicted through a dendrogram, where proximity signifies similarity. In the context of symbol sequence analysis, hierarchical clustering facilitates the identification of coherent symbol groups sharing similar structural attributes.

Implementation of hierarchical clustering entailed the use of distance metrics like Euclidean distance or cosine similarity to quantify dissimilarity between symbols. This approach facilitated the measurement of structural disparities among symbol sequences and guided their clustering based on proximity. Additionally, linkage methods such as single, complete, or average linkage determined cluster merging strategies at each step.

Post-completion, visualization of the dendrogram allowed for interpretation of the resulting hierarchical structure of symbol clusters. This visualization facilitated the identification of distinct clusters and hierarchical relationships among symbols, offering insights into their structural organization within the script.

Through the application of hierarchical clustering to the Indus script dataset, efforts were made to unveil underlying patterns and structural regularities embedded within the script. This methodological approach synergized with other techniques like dependency tree analysis, enriching the comprehension of the script's linguistic properties and cultural significance.

3.1.2 Dependency Tree

In the exploration of the Indus script, dependency trees were harnessed as a robust tool for dissecting syntactic structure and symbol relationships. Dependency trees provide a hierarchical representation of word or symbol relationships within a sequence, encapsulating dependencies such as subject-verb and modifier-noun relationships.

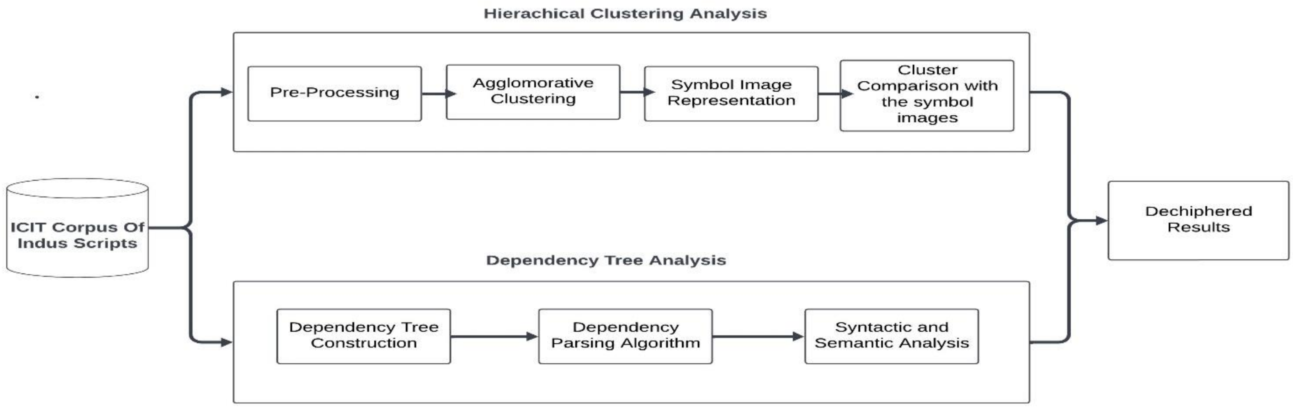


Figure 1: Overall System Architecture

Construction of dependency trees for Indus script symbols commenced with sequence segmentation followed by the adaptation of natural language processing techniques tailored for symbolic languages. Parsing these sequences facilitated the extraction of syntactic dependencies like modifiers, complements, and heads within each symbol sequence.

Upon identification, these dependency relationships were mapped onto a tree structure, with each symbol serving as a node and dependencies as directed edges. This visualization allowed for an in-depth analysis of syntactic relationships and hierarchical organization within symbol sequences.

Dependency trees furnished invaluable insights into the grammatical structure of the Indus script, unveiling patterns like subject-object relationships and verb agreements. Analysis of hierarchical symbol organization facilitated a deeper comprehension of underlying grammar and syntax encoded within the script.

Furthermore, dependency trees facilitated the development of computational models for symbol sequence analysis, enabling tasks such as sequence generation and translation. Leveraging syntactic information from dependency trees enhanced the capacity to interpret and generate meaningful symbol sequences, advancing efforts towards deciphering the Indus script.

3.2 Algorithm

3.2.1 Algorithm for Hierarchical clustering

Hierarchical Clustering Algorithm:
1. Data Acquisition and Preprocessing:
1.1 Obtain the symbol sequences from the dataset.
1.2 Preprocess the symbol sequences to extract individual symbols and their relative positioning within each sequence.
2. Distance Metric Calculation:
2.1 Define a distance metric to quantify the dissimilarity between pairs of symbol sequences.
2.2 Calculate the pairwise distance matrix based on the chosen distance metric.
3. Hierarchical Clustering:
3.1 Initialize each symbol sequence as a separate cluster.
3.2 Compute the pairwise dissimilarity between clusters using the distance matrix.
3.3 Merge the two closest clusters into a new cluster.
3.4 Update the distance matrix to reflect the dissimilarities between the new cluster and the remaining clusters.
3.5 Repeat steps 3.2-3.4 until all symbols are grouped into a single cluster or until a stopping criterion is met.
4. Visualization:
4.1 Visualize the hierarchical clustering process using a dendrogram to illustrate the merging of clusters at each step.
4.2 Explore the dendrogram to identify clusters of symbols with similar structural characteristics.
5. Interpretation:
5.1 Analyze the resulting clusters to uncover underlying patterns and structural regularities within the symbol sequences.
5.2 Interpret the hierarchical structure of the dendrogram to understand the relationships and similarities among symbol clusters.
5.3 Assess the quality of the clustering solution based on domain-specific criteria and expert knowledge.
6. Post-processing:
6.1 Apply post-processing techniques, such as pruning or refining the cluster assignments, to improve the quality of the clustering solution.
6.2 Evaluate the effectiveness of the hierarchical clustering algorithm in capturing meaningful patterns and relationships within the symbol sequences.
7. Validation:
7.1 Validate the clustering results using domain-specific knowledge or external validation measures.
7.2 Assess the robustness of the hierarchical clustering algorithm to variations in input parameters or dataset characteristics.
8. Optimization:
8.1 Explore strategies to optimize the hierarchical clustering algorithm, such as tuning the distance metric or adjusting the stopping criterion.
8.2 Iterate on the algorithm and visualization techniques to enhance the interpretability and utility of the clustering results.

Figure 2: Algorithm for Hierarchical Clustering

3.2.2 Algorithm for Dependency tree analysis

Dependency Tree Construction Algorithm for Indus Script Analysis:
1. Data Preparation:
1.1 Segment the Indus script corpus into individual symbol sequences.
1.2 Tokenize the symbol sequences to obtain the basic units (symbols) for dependency analysis.
2. Dependency Parsing:
2.1 Adapt dependency parsing algorithms for symbolic languages to analyze the syntactic structure of the Indus script symbol sequences.
2.2 Implement techniques such as transition-based parsing, graph-based parsing, or neural network-based parsing to extract dependency relationships between symbols.
2.3 Identify dependencies such as modifier-noun, verb-object, or subject-verb relationships within the symbol sequences.
3. Dependency Tree Construction:
3.1 Represent the dependencies as a directed graph, where each symbol is a node, and the dependencies are directed edges.
3.2 Construct the dependency tree by selecting a root node and organizing the remaining nodes based on their dependency relationships.
3.3 Assign labels to the edges indicating the type of dependency (e.g., modifier, object, verb).
3.4 Ensure that the dependency tree is acyclic and rooted, with a clear hierarchical structure.
4. Visualization:
4.1 Visualize the dependency tree using graph visualization techniques tailored for symbolic languages.
4.2 Include labels for nodes and edges to facilitate interpretation of the syntactic relationships within the Indus script.
5. Analysis:
5.1 Analyze the hierarchical structure of the dependency tree to identify syntactic patterns and relationships within the Indus script.
5.2 Explore the dependency tree to understand the syntactic roles of symbols and the overall grammatical structure of the script.
5.3 Extract relevant features from the dependency tree for downstream tasks such as decipherment, interpretation, or cultural analysis.
6. Interpretation:
6.1 Interpret the dependency relationships within the tree to gain insights into the syntactic properties of the Indus script.
6.2 Identify common patterns or motifs in the dependency structures and their implications for linguistic and cultural analysis.
7. Validation:
7.1 Validate the dependency tree construction process using expert annotations or linguistic theories specific to the Indus script.
7.2 Assess the accuracy and completeness of the dependency tree representations relative to established linguistic frameworks and domain-specific knowledge.
8. Optimization:
8.1 Explore techniques to optimize the efficiency and accuracy of dependency parsing for the Indus script, considering the unique characteristics of symbolic languages.
8.2 Fine-tune the parameters of the dependency parsing algorithm to improve its performance on the Indus script corpus and enhance the quality of the dependency tree representations.

Figure 3: Algorithm for Dependency Tree

4. RESULTS AND INTERPRETATION

This section consists of the results that are obtained from executing our methodologies and what we infer from it

4.1 Hierarchical Clustering:

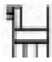
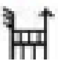
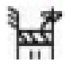
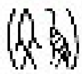
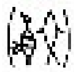
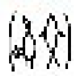
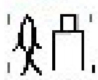
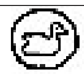
1.Animal Symbols

- Contextual Variance: Symbols on pottery reflect public, regional ideologies, while those on seals indicate more exclusive, possibly elite, or ritualistic contexts.
- Horn Characteristics: Horn shapes and sizes offer clues to the animal's identity; large, sweeping horns

suggest wild water buffalo, while smaller, curly horns may indicate domestic varieties.

- Regional Depictions: Comparing symbols across sites reveals regional variations, such as horn designs or curvature, aiding in understanding the animal's significance within local cultural contexts.

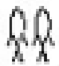
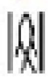
Table 1. List of signs which has resemblance to animals.

Serial No.	Sign No	Description	Design
1	255	Fish with stripes	
2	256	Striped fish on a hook	
3	257		
4	260	Fowl	
5	261	Two Fowls	
6	262	Duck	
7	263		
8	264		
9	265		
10	266		
11	267		
12	268		
13	269		
14	270		
15	271		
16	272		
17	340	Dog	
18	341		
19	342	Horse	
20	343		
21	344	Goat	
22	345		
23	346		
24	347		
25	348	Tiger/Lion	
26	940		
27	942		
28	943		
29	944		
30	952		

2. Fish Symbols

- Parpola associates the 'dot + fish' sign with the Harappan proto-form of Durga, the goddess of fertility and victory in war.
- He suggests that Rohini (Aldeberan) is the star most likely associated with this goddess, based on his exploration of Hindu mythology.
- The association between Rohini (the 'red star') and the carp ('rohita', meaning 'the red fish') forms the basis of identification for the 'fish' sign.
- Parpola interprets the mark inside the 'fish' sign as the auspicious red tilaka mark worn by women, equating it with Dravidian 'pottu' (meaning '(red) dot, drop').
- However, linguistic challenges arise, as 'pottu meen' is not attested in Dravidian with the meaning 'star Rohini', and 'pottu' itself does not inherently mean 'red' in Dravidian languages.
- Various interpretations exist regarding the significance of the 'fish' symbol, including its representation as a trade item quantity, a weight standard, or a religious motif.
- Scholars have explored the 'fish' symbol's potential as an ideogram, suggesting it may have held religious or political significance within the Indus Valley society.
- The symbol's association with the Matsya kingdom, both in literary and archaeological contexts, suggests a possible connection between the 'fish' symbol and dynastic or territorial identities.
- The prevalence of fish symbolism extends beyond the Indus Valley civilization, with examples found in historic kingdoms like the Pāṇḍyas, indicating a continued cultural significance across different time periods and regions in South Asia

Table 2. List of signs which has resemblance to fish.

Serial No.	Sign No	Description	Design	Intended Meaning
1	219	Two fish		
2	220	Fish		Fish or Star
3	221	Fish in Brackets		
4	222	Fish in Brackets with a horn		Related to moon
5	223			
6	224			
7	225			
8	226			
9	227			
10	228	Fish with rays spreading out		Related to Sun
11	232			
12	233	Dividing Fish		Sunset
13	234			
14	235	Fish+roof		Black star which is Saturn
15	236	Fish with hat and four strokes		
16	240	Fish+rays		Cat Fish
17	241			
18	242			
19	243			
20	244	Fish + Wings		Bird

3.Agriculture symbols

- Food crops: Harappans cultivated bread wheat, barley, sesame, peas, melons, date-palm, species of Brassica. Cotton was an important crop and the centre of origin of Gossypium arboretum lies in the

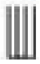
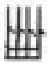
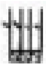
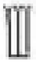
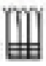
Indus Valley.

- Wheat: Among the cereals, wheat and barley were recorded from Harappa and Mohenjo-daro. Helbaek has pointed out that wheat and barley have been cultivated together from the very beginning of village- farming in western Asia.

- Barley: All barleys, wild and cultivated, belong to the same potentially interfertile population and are grouped under one species, Hordeum vulgare.

- Rice: Rice is native to South and East Asia- including Indus region and the Ganges valley. Cultivation of rice involved several different centres of domestication

Table 3. List of signs related to agriculture.

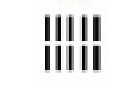
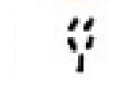
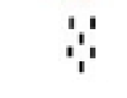
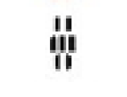
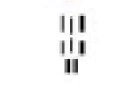
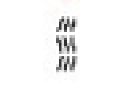
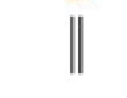
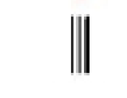
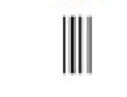
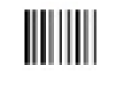
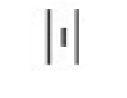
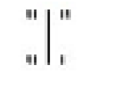
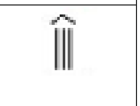
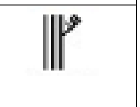
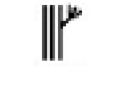
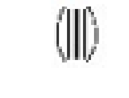
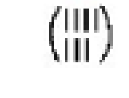
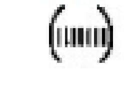
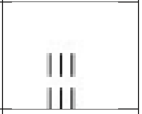
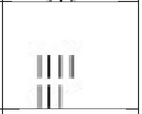
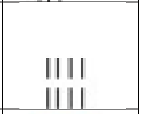
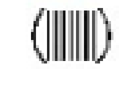
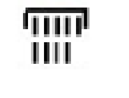
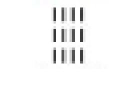
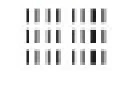
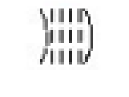
Serial No.	Sign No	Description	Design
1	420	Rice	
2	421	Wheat	
3	422	Barley	
4	423		
5	424		
6	425		
7	426		
8	427		
9	428		
10	429		
11	430		

4. Numeral signs

- Origin of Number Symbols: The paper explores the origin of number symbols, tracing them back to the Indus Valley Civilization, with evidence from excavations at Mohen-jo-daro and Harappa post-1920 challenging previous perceptions about the oldest civilization.
- Significance of Indus Valley Numerals: The paper argues that the invention of numeral symbols in the Indus Valley Civilization laid the groundwork for the discovery of zero, indicating India's early contribution to the development of numerical systems.

Table 4. List of signs which are possible numerals

Serial No.	Sign No	Design
1	1	
2	2	
3	3	
4	4	
5	5	
6	6	
7	7	
8	12	
9	13	
10	14	
11	15	
12	19	

13	20	
14	25	
15	26	
16	27	
17	28	
18	29	
19	31	
20	32	
21	33	
22	34	
23	36	
24	37	
25	39	
26	41	
27	42	
28	44	
29	45	
30	46	
31	47	
32	48	
33	49	
34	16	
35	17	
36	18	
37	50	
38	51	
39	55	
40	56	
41	57	

5. Crop and road symbols

- Pictorial Interpretations: Basic signs like "share (as grain)" (Sign 137) and "crop" (Sign 162) are analyzed, with variants and compound signs providing clues to their meanings. For example, Sign 137 depicts stalks laden with grain arranged in X-like form to denote "share (as grain)."

- Modifiers and Compound Signs: Elements like "sky," "one-eighth," and "roof" modify basic signs, yielding compound signs indicating different aspects of agricultural distribution, such as "god's share of grain/crop" or "upper (landlord's) share of grain/crop."

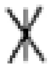
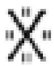
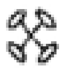
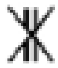
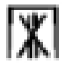
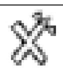
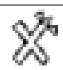
- Bi-lingual Parallels: Bi-lingual parallels from Dravidian and Indo-Aryan languages, such as Tamil and Malayalam, support the proposed ideographic identification of the signs, indicating cultural continuity.

- Survivals in Later Periods: Some agricultural signs from the Indus script are found as isolated symbols in pottery graffiti from later Chalcolithic and Megalithic periods, suggesting continuity in agricultural practices.

- Agricultural Implements: The paper also identifies pictorial signs representing agricultural implements, such as harrows, and suggests their symbolic use in the script, although not in their literal sense.

- Grid Analysis: The grid of related signs presented in the paper facilitates analysis, enabling the identification of rare signs and predicting potential new discoveries, contributing to ongoing research on the Indus script.

Table 5. List of signs which are related to crops and roads/stars.

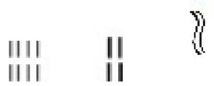
Serial No.	Sign No	Description	Design	Intendent Meaning
1	645	Two intersecting lines		Divide/Share related to Grain
2	646	Two intersecting lines with two dots		
3	647			
4	678	Two intersecting lines with hat		Hat means "upper" and this means landlord's share
5	679	Two intersecting lines with upper bracket		Sky modifier(upper bracket) and divide grains means "gods" share as temple offerings
6	680	645 + 4 pairs of short stokes		One eight share of the crop
7	681			
8	683			
9	684			
10	685	645+ two strokes		Share of crop
11	686			Variant of share of crop
12	687			
13	688			
14	689			
15	690			Streets
16	692			
17	693			
18	694			
19	690			
20	692			
21	693			
22	694			

6.Enclosed Signs

In the Indus script, some symbols are enclosed within others. Their purpose remains debated, possibly indicating grammatical nuances, numerical values, or social hierarchies within the ancient civilization's writing system. These nested symbols suggest a complex encoding of information, yet their precise function remains enigmatic.

Modifiers: - Enclosing Symbols:

"032" "900" "904" "906" "034"



"018" " 014" "927"

Table 6: Most frequently occurring enclosed symbols

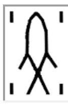
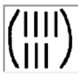
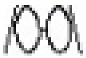
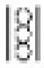
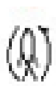
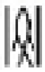
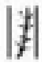
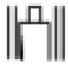
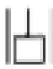
Sign No	Image	Frequency
226		41
236		26
97		20
48		18

Table 7: Enclosed symbols

592	
688	
689	
793	
942	
943	
944	
945	
946	
947	
241	
265	
267	
47	
48	
49	
50	
57	
73	
85	
97	
101	
804	
807	
822	
842	
843	
854	
855	
870	
878	
895	
910	
912	
926	
940	

144	
170	
217	
221	
222	
223	
224	
225	
226	
232	
234	
236	

269	
270	
271	
281	
375	
393	
412	
466	
562	
577	
591	

4.2 Dependency Tree Analysis

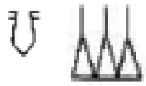
1. "740" and "002" frequently appear together, indicating a consistent pairing and "740" is often followed by "002" in sequential order, suggesting a dependency between these numbers.
2. "520 " "220" "233" "240" implies a hierarchical relationship among the numbers.
3. "705" often appears alongside "233", suggesting an action-object association.
4. The temporal ordering of numbers is evident in sequences where "920" follows "740" and "002".
5. "032" frequently appears between "740" and "002", contributing to syntactic coherence.
6. Some sequences follow agent-action structures, with "100" (agent) preceding "415" (action).
7. Object-attribute pairings are observed, with "415" (object) followed by "002" (attribute).
8. Certain sequences exhibit sequential clusters such as "740" "100" "415", indicating a clustering tendency within the language.
9. "032" often serves as a transitional phrase between numerical elements, suggesting a structural cue for transitions or shifts in meaning.
10. Semantic groupings are apparent, such as "740" "176" and "740" "690" "435" "255", possibly denoting specific categories or concepts within the language.

11. "061" occasionally follows "740", potentially serving as a temporal marker or modifier within the language.

12. Certain symbols, like "740" and "002", exhibit a high frequency of co-occurrence, indicating a strong semantic or grammatical association.

13. Some sequences, such as "740" "900" "850", suggest conditional structures or clauses, possibly indicating conditional relationships or dependencies.



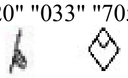
14. Symmetrical patterns, like

"740" "460" "503" followed by

"031" "002" "900", hint at symbolic symmetry or parallel structures within the language.

15. Sequences like

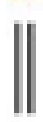
"520" "033" "705" "803" followed by



"001" "368" "861" suggest predetermined sequences or predefined syntactic patterns.

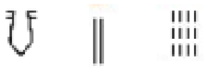
16. Certain sequences imply structural hierarchies, with

numbers like "740" often preceding more



detailed or subordinate elements like "032"

or "002".



17. Sequences like "740"

"032" "055" followed by "001" "407" "004" hint at semantic expansion or elaboration, with subsequent elements building upon preceding ones.

5. CONCLUSION AND FUTURE WORK

The combination of hierarchical clustering and dependency tree analysis has provided valuable insights into the structural organization and syntactic relationships within the Indus script. Our interdisciplinary approach has revealed hierarchical patterns and dependencies among symbols, shedding light on the script's linguistic properties and cultural significance. By leveraging hierarchical clustering to identify groups of symbols with similar structural characteristics and constructing dependency trees to analyze their syntactic relationships, we have advanced our understanding of the script's underlying grammar and syntax.

Moving forward, future research endeavors could explore innovative avenues to enhance our decipherment efforts and broaden our understanding of the Indus script. Integration of advanced natural language processing techniques, such as semantic analysis and deep learning, could facilitate the extraction of meaning and context from symbols, enabling more nuanced interpretations of the script's content. Additionally, multimodal data fusion techniques could be employed to combine textual and image data, providing richer contextual information for decipherment and analysis.

Interactive visualization tools could play a crucial role in facilitating exploration and analysis of the script network and decipherment results. By developing user-friendly visualization platforms, researchers and the public alike can engage with the data in an intuitive and interactive manner, fostering collaboration and knowledge sharing in the field of ancient script decipherment.

Furthermore, incorporating historical and archaeological contextualization into decipherment efforts could provide valuable insights into the cultural and historical context of the Indus script. By integrating archaeological findings and historical records, researchers can contextualize the script within the broader framework of the Indus Valley Civilization, enriching our understanding of its origins, usage, and cultural significance.

By embracing these advancements and interdisciplinary approaches, we can deepen our understanding of the Indus script's grammar and semantics, unlocking valuable insights into one of humanity's earliest writing systems and contributing to the broader study of ancient civilizations and linguistic diversity.

6. REFERENCES

[1] Tsutomu Ito, Ryosuke Harakawa, Masahiro Iwahashi, "Word Clustering Using Graphical Lasso-Guided PCA for Trend Analysis of COVID-19", *2021 IEEE 10th Global Conference on Consumer Electronics (GCCE)*, pp.200-201, 2021

[2] R. Gherbaoui, N. Benamrane and M. Ouali, "A New Similarity Measure and Hierarchical Clustering Approach to Color Image Segmentation," *2023 IEEE 24th International Conference on Information Reuse and Integration for Data Science (IRI)*, Bellevue, WA, USA, 2023, pp. 34-39, doi: 10.1109/IRI58017.2023.00014.

[3] J. Liu, D. Wang, S. Yu, X. Li, Z. Han and Y. Tang, "A Survey of Image Clustering: Taxonomy and Recent Methods," *2021 IEEE International Conference on Real-*

time Computing and Robotics (RCAR), Xining, China, 2021, pp. 375-380, doi: 10.1109/RCAR52367.2021.9517087.

[4] T. Yu and S. Gao, "Abstractive Text Summarization With Semantic Dependency Graph," *2023 5th International Academic Exchange Conference on Science and Technology Innovation (IAECST)*, Guangzhou, China, 2023, pp. 567-570, doi: 10.1109/IAECST60924.2023.10502964.

[5] G. Liu, K. Wang, W. Liu and Y. Cao, "The Construction and Measure Method of Dependency Parsing Tree Model," *2019 IEEE 10th International Conference on Software Engineering and Service Science (ICSESS)*, Beijing, China, 2019, pp. 1-4, doi: 10.1109/ICSESS47205.2019.9040816.

[6] Jelenik F (1997) Statistical Methods for Speech Recognition (MIT Press, Cambridge, MA).

[7] Manning C, Schu¨ tze H (1999) Foundations of Statistical Natural Language Processing (MIT Press, Cambridge, MA).

[8] Yadav N, et al. (2009) Statistical analysis of the Indus script using n-grams. arxiv: 0901.3017 (available at http://arxiv.org/abs/0901.3017).

[9] Drake AW (1967) Fundamentals of Applied Probability Theory (McGraw–Hill, New York).

[10] Rabiner LR (1989) A tutorial on Hidden Markov Models and selected applications in speech recognition. Proc IEEE 77:257–286. 13. Mahadevan I (1977) The Indus Script: Texts, Concordance, and Tables (Memoirs of Archaeological Survey of India, New Delhi, India)