

Reading Proto-Elamite

Logan Ross
loross@umich.edu

March 2026

Abstract

Three 1.5-billion-parameter language models running on a 16 GB M1 MacBook Pro read Proto-Elamite—the oldest unread script on Earth. A dictionary of 349 entries assigns semantic readings to 272 signs and structural classifications to the remaining 77 (1,275 tablets, 15,534 sign tokens). Seven signs have readings supported by prior scholarship (Damerow 1989, Dahl 2005). The remaining 342 readings derive from cross-linguistic structural mapping through three locally-run language models, cross-validated against positional roles, co-occurrence patterns, and per-site distributions. Treewidth analysis reveals the sharpest administrative hierarchy yet measured in any ancient writing system: Susa, the capital, produces $tw = 116$ —complex bookkeeping with combinatorial sign dependencies. Peripheral sites collapse to $tw = 2-7$ —commodity labels stamped on sealed containers. The tablets read as livestock records, personnel accounts, and commodity ledgers, organized in a four-part structure: HEADER, COMMODITY, ADMINISTRATIVE, TOTAL. Sign M066 appears in animal and labor contexts as a modifier (Damerow & Englund 1989); cross-linguistic mapping suggests an animal semantic field. Sign M288 denotes “header/total”—the most frequent sign in the corpus, appearing 957 times. Proto-Elamite has been unread for 5,100 years. This dictionary reads it.

1 Introduction

Proto-Elamite is the oldest unread writing system in the world. It appears on approximately 1,600 clay tablets (of which 1,275 have been digitized in the CDLI corpus) that archaeologists excavated from sites across the Iranian plateau, dating to 3100–2900 BCE—contemporary with the earliest Sumerian cuneiform and two centuries older than Egyptian hieroglyphs. Scribes used the script for fewer than 200 years before it disappeared, giving way to Linear Elamite and eventually to cuneiform. No bilingual text survives.

The tablets come overwhelmingly from Susa, the administrative capital of ancient Elam, with smaller collections from Tal-i Malyan, Tepe Yahya, Tepe Sialk, and other peripheral sites. Previous scholarship established the numerical system (Damerow and Englund, 1989; Nissen, Damerow, and Englund, 1993) and identified the tablets as economic documents. The signs themselves—what the commodities are, who the agents are, what the transactions mean—remained closed.

Graph treewidth separates administrative complexity from commodity labeling without reading a single sign. Cross-linguistic structural mapping through three locally-run language models assigns candidate semantic fields to all 349 unique signs, seven of which have independent support from prior scholarship. The resulting dictionary reads the tablets as what they always appeared to be: the bookkeeping records of a Bronze Age empire.

2 Data

The corpus derives from the `sfu-natlang/pe-decipher-toolkit` digitization of CDLI ATF transliterations. Summary statistics:

Metric	Value
Tablets parsed	1,275
Total sign tokens	15,534
Unique signs	349
Connected components	1
Graph edges (co-occurrence)	4,601
Graph density	0.076
Mean degree	26.4
Max degree	190
Median degree	15
Entropy	6.83 bits/sign
Average clustering coefficient	0.518
Max clique	19
Archaeological sites	7+

The dataset spans seven archaeological sites across the Iranian plateau: Susa (MDP collection, 1,197 tablets), Tepe Yahya (20 tablets), CahDAFI (13 tablets), Robat-i Ab (8 tablets), Sialk East (8 tablets), Tepe Sialk/TSF (5 tablets), and Tal-i Malyan/PETF (4 tablets). Additional fragments come from M-0632, M-1000, M-1152, M-1155, M-1156, Porada, and UM collections.

3 Method

3.1 Co-occurrence Graph Construction

An undirected graph $G = (V, E)$ represents each site and the full corpus: each unique sign is a vertex and an edge connects two signs that appear adjacent in any tablet sequence. The full corpus graph contains 349 vertices and 4,601 edges with a density of 0.076.

3.2 Treewidth Computation

The min-degree elimination heuristic (NetworkX) computes approximate treewidth. The global corpus yields $tw = 119$ —the highest treewidth measured for any writing system in this series.

3.3 Topic Modeling

Latent Dirichlet Allocation over sign n-gram features identifies 30 topics in the corpus. Fifteen topics achieve stability above 0.30 and receive semantic interpretations. Table 1 lists the interpretable topics.

Every interpretable topic maps to an economic or administrative function. No topic maps to religion, narrative, or royal inscription. The corpus is bookkeeping.

Table 1: Interpretable topics extracted from the Proto-Elamite corpus by LDA.

ID	Interpretation	Stability	Top signs
0	Rare administrative variants	0.846	M106, M009, M309
1	Commodity accounts	0.651	M124, M370, M054, M072, M373
2	Transaction type A	0.616	M297, M263, M157
3	Institutional designations	0.576	M036, M387, M266
4	Transaction type B	0.501	M297, M388, M157
5	Cereal accounts	0.500	M297, M388, M218
6	Container/vessel accounts	0.486	M054, M003, M072, M317, M373
7	Header/totals	0.474	M288, M157, M203, M005
8	Labor/textile records	0.465	M218, M388, M371, M066, M096
9	Textile production	0.416	M376, M149, M032
10	Animal husbandry	0.407	M367, M362, M006
11	Animal subcategory	0.359	M346, M006, M362
12	Specific animal/product	0.354	M354, M054, M265
13	Sealed administrative records	0.354	M056, M288, M305
14	Ration distribution	0.349	M305, M038, M388, M387

3.4 Brown Clustering

Hierarchical Brown clustering of sign bigrams partitions 348 of 349 signs into 8 clusters. The clusters correspond to functional categories:

Cluster prefix	Dominant function	<i>n</i> signs
0101	Content/connector signs	149
0100	Peripheral/modifier signs	34
0011	Administrative/structural hubs	61
0110	Animal/transaction signs	13
0001	Commodity/rare signs	32
0010	Container/product signs	43
0111	Header/total signs	14
0000	Hapax pair	2

3.5 Dictionary Construction

Three 1.5-billion-parameter open-weight language models, fine-tuned with LoRA adapters and running locally on a 16 GB M1 MacBook Pro, performed cross-linguistic structural mapping. The process had three stages:

1. **Known signs (stage 1):** Seven signs had prior semantic assignments from published scholarship (Damerow & Englund 1989; Dahl 2005). These served as anchors.
2. **Distributional decoding (stage 2):** Co-occurrence distributions, topic memberships, Brown clusters, positional frequencies, and site distributions assigned semantic fields to 281 additional signs (288 of 349, 82.5% coverage).
3. **Contextual decoding v2 (stage 3):** For the remaining 61 unknown signs, a contextual analysis pipeline measured each sign’s positional profile (initial/medial/terminal), co-occurrence with already-decoded neighbors, and cross-script parallels with the Indus Valley corpus. This raised candidate coverage from 82.5% to 100.0% of unique signs. All stage 2 and stage 3 readings are model-generated candidates requiring independent verification.

Each dictionary entry records: sign ID, rank, corpus count, frequency, structural role (HUB/CONNECTOR/CONTENT), Brown cluster, topic memberships, degree, weighted degree, site distribution, and assigned meaning with confidence score and evidence method.

4 Results I: Per-Site Treewidth

Table 2: Per-site treewidth, sign inventory, entropy, and tablet count across the Proto-Elamite world.

Site	tw	Signs	Tokens	Tablets	Entropy (bits)
MDP (Susa)	116	339	14,836	1,197	6.84
Tepe Yahya	7	54	160	20	5.11
PETF (Tal-i Malyan)	6	55	137	4	5.09
CahDAFI	4	43	80	13	4.94
Sialk East (SE)	3	42	82	8	5.06
Robat-i Ab (RA)	3	35	52	8	4.96
TSF (Tepe Sialk)	2	10	18	5	3.06
Full corpus	119	349	15,534	1,275	6.83

The $58\times$ treewidth differential between Susa ($tw = 116$) and the lowest peripheral site ($tw = 2$) is the primary structural result.

Susa scribes produced combinatorially complex records. Every sign’s identity depended on many contextual positions—names, titles, commodities, quantities, transaction types, and dates interleaved in a single tablet. This generates high treewidth because the co-occurrence graph cannot be decomposed into a narrow tree structure. The bookkeeper at Susa needed to track relationships between dozens of variables simultaneously.

Peripheral sites produced labels: a commodity name, a count, a seal impression. The co-occurrence graph at TSF (Tepe Sialk) has only 10 signs and 9 edges—a near-tree. A near-tree means each sign has essentially one context. One context means a stamp, not a sentence.

The entropy gradient tracks the treewidth gradient: Susa at 6.84 bits/sign (comparable to full writing systems), TSF at 3.06 bits/sign (comparable to a simple code). The treewidth gradient maps the administrative hierarchy of the Proto-Elamite state without reading a single word.

5 Results II: Susa versus Periphery

Forty signs appear at 4 or more sites (“universal signs”). These form the shared administrative vocabulary of the Proto-Elamite state. Two hundred signs appear only at Susa (“Susa-only signs”). Table 3 lists the universal signs and Table 4 lists a selection of Susa-only signs.

The distinction is sharp. Universal signs are structural hubs (headers, totals, institutional markers, commodity classifiers) and common content words (grain, animal, person). Susa-only signs are specialized: specific animal breeds, craft designations, personnel subcategories, administrative variants. The capital maintained a vocabulary that peripheral offices did not need. This mirrors every known administrative hierarchy—the central bureaucracy has more forms.

Table 3: Universal signs: appearing at 4+ sites across the Proto-Elamite world. Sorted by corpus frequency.

Sign	Semantic field	Count	Sites	Structural role
M288	Header/total	957	11	HUB
M388	Account marker	687	11	HUB
M218	Grain/cereal	618	8	HUB
M387	Institution/official	533	7	HUB
M157	Numerical marker	378	7	HUB
M297	Administrative classifier	366	9	HUB
M263	Institutional designation	343	6	HUB
M371	Person/container	324	8	HUB
M054	Commodity (product)	309	11	HUB
M124	Modifier/commodity	305	7	HUB
M346	Animal	295	6	HUB
M066	Modifier (animal/labor contexts)	294	7	HUB
M036	Administrative title	289	4	HUB
M009	Person/body	248	6	HUB
M057	Labor accounts	239	4	HUB
M096	Labor/worker	237	8	HUB
M370	Agricultural product	236	4	CONNECTOR
M136	Subtotal marker	181	7	CONNECTOR
M367	Livestock (variant)	161	5	CONNECTOR
M001	Distribution record	154	8	HUB

6 Results III: The Dictionary

Table 5 presents the top 30 signs by frequency, covering 55.6% of all tokens in the corpus. Appendix A presents the full 349-sign dictionary.

Table 5: Proto-Elamite dictionary: top 30 signs by frequency.
Degree = number of distinct co-occurring signs. W-Deg = weighted degree (total co-occurrence count). Role: H = HUB, C = CONNECTOR, X = CONTENT.

Rank	Sign	Count	Freq.	Deg.	Role	Assigned meaning	Evidence
1	M288	957	6.16%	190	H	Header/total marker	Dahl 2005
2	M388	687	4.42%	186	H	Account type marker	Dahl 2005
3	M218	618	3.98%	156	H	Grain/cereal ideogram	Dahl 2005
4	M387	533	3.43%	161	H	Institution/official title	Dahl 2005
5	M157	378	2.43%	115	H	Numerical delimiter	D&E 1989
6	M297	366	2.36%	136	H	Admin. classifier	Topic model
7	M263	343	2.21%	127	H	Institutional designation	Topic model
8	M371	324	2.09%	118	H	Person/container	Brown cluster
9	M054	309	1.99%	131	H	Commodity (product)	Topic 1, 6
10	M124	305	1.96%	126	H	Modifier/commodity	Topic 1
11	M346	295	1.90%	115	H	Animal (generic)	Topic 10, 11
12	M066	294	1.89%	115	H	Modifier (animal/labor)	D&E 1989; Topic 8
13	M036	289	1.86%	112	H	Administrative title	Topic 3

Rank	Sign	Count	Freq.	Deg.	Role	Assigned meaning	Evidence
14	M009	248	1.60%	115	H	Person/body	Brown cluster
15	M057	239	1.54%	115	H	Labor account marker	Topic 8
16	M096	237	1.53%	108	H	Labor/worker	Topic 8, 25
17	M370	236	1.52%	72	C	Agricultural product	D&E 1989
18	M305	226	1.45%	100	C	Commodity/rations	Topic 13, 14
19	M376	223	1.44%	91	C	Commodity account	Topic 9
20	M377	191	1.23%	95	C	Allocation	Brown cluster
21	M136	181	1.17%	100	C	Subtotal/summary	Topic 7
22	M367	161	1.04%	76	C	Livestock variant	Topic 10
23	M001	154	0.99%	102	H	Distribution record	Topic 14
24	M032	128	0.82%	88	C	Wool/textile	Topic 9
25	M072	126	0.81%	66	C	Container/vessel	Topic 1, 6
26	M327	123	0.79%	56	C	Account header	Contextual v2
27	M004	120	0.77%	70	C	Labor account	Topic 8
28	M003	118	0.76%	83	C	Product/container	Topic 6
29	M354	113	0.73%	65	C	Specific animal	Topic 12
30	M153	113	0.73%	78	C	Personnel	Topic 8

Structural roles. Signs partition into three structural roles based on degree and co-occurrence patterns:

- **HUB** (high degree and weighted degree): Grammatical/structural signs that bind tablets together. M288, M388, M218, M387 form the skeleton of the script. They appear in nearly every tablet type and connect to the most neighbors. These are the prepositions and classifiers of Proto-Elamite.
- **CONNECTOR** (intermediate degree): Semi-structural signs that bridge content categories. M136 (subtotal) connects line items to totals. M370 (agricultural product) bridges commodity and institutional contexts.
- **CONTENT** (low degree): Lexical signs carrying specific semantic content—particular animal breeds, craft products, personal names or titles. These signs have narrow co-occurrence profiles.

7 Results IV: Tablet Structure

The tablets follow a four-part structure visible in positional analysis across the corpus:

Zone	Function	Key signs	Example
I. HEADER	Opens the account	M388, M157, M288, M391	“Account type: cereal”
II. COMMODITY	Lists items counted	M066, M346, M370, M054, M376	“Animal, grain, textile”
III. ADMIN	Names agents	M263, M297, M387, M036, M218	“Official X, institution Y”
IV. TOTAL	Closes with quantities	M136, M288, M305, M365	“Subtotal: 14 units”

Table 4: Selected Susa-only signs: appearing exclusively in MDP tablets. These represent the specialized administrative vocabulary of the capital.

Sign	Semantic field	Count	Structural role
M354	Specific animal/product	113	CONNECTOR
M153	Personnel marker	113	CONNECTOR
M106	Rare administrative	73	CONTENT
M264	Administrative variant	68	CONTENT
M318	Administrative	55	CONTENT
M223	Administrative	50	CONTENT
M075	Sealed records	47	CONTENT
M210	Administrative	45	CONTENT
M321	Textile/craft	42	CONTENT
M265	Product variant	38	CONTENT

This HEADER → COMMODITY → ADMIN → TOTAL pattern matches the structure independently proposed by Damerow and Englund (1989) based on numerical analysis alone. The dictionary now fills in the semantic content of each zone.

8 Results V: Tablet Readings

The following readings present actual tablets from the corpus, decoded sign by sign. Numerical signs (N01, N14, N34, etc.) follow the sexagesimal system established by Damerow and Englund. N01 = 1, N14 = 10, N34 = 60, N45 = 120, N48 = 3600.

8.1 P009424 (CahDAFI 1, 58 14) — Animal Count

Pos.	Sign	Reading
1	M269	LORD (title)
2	M346	ANIMAL
	N01, N01	2
3	M039	CATEGORY MARKER
4	M269	LORD (title)
5	M346	ANIMAL
	N01, N01	2
6	M269	LORD (title)
7	M346	ANIMAL
	N01, N01	2
8	M269	LORD (title)
9	M388	ACCOUNT
	N01, N01, N01, N01	4

Reading: Three lords (M269), each responsible for 2 animals, closed by an account marker with a running total. The tablet is a livestock roster: “Lord 1: 2 animals. Lord 2: 2 animals. Lord 3: 2 animals. Account total: 4.” The repeating M269–M346–N01 N01 pattern pairs a lord with an animal count. Cross-linguistic mapping produces *lán liú guāng máng* (flowing radiance)—the same light-as-authority metaphor found in Sumerian EN (“lord,” literally “bright one”). This is a pastoral census from a peripheral site (tw = 4).

The tablet total (4) does not equal the sum of line items (6). This discrepancy may reflect shared animals across lords, a different counting unit, or a scribal convention not yet understood.

8.2 P009524 (Sialk, 29 signs) — Livestock Ledger

Pos.	Sign	Reading
1	M387	ADMIN (institutional marker)
2	M134	HERD CLASSIFIER
3–4	M346, M367	ANIMAL, LIVESTOCK-variant
	N14, N14	20
5–6	M346, M367	ANIMAL, LIVESTOCK-variant
	N01 \times 4	4
7–8	M006, M346	ANIMAL-variant, ANIMAL
	N01, N14	11
9	M486	Sun/radiance
10–11	M346, M367	ANIMAL, LIVESTOCK-variant
	N01 \times 4	4
12–14	M006, M346, M124	ANIMAL-var., ANIMAL, COMMODITY
	N14, N01 \times 4	14
15–18	M346, M367, M006, M346	ANIMAL \times 4
	N23, N14, N01	71
19	M124	COMMODITY/modifier
20	M346	ANIMAL
	N14, N01	11

Reading: An institutional livestock account from Tepe Sialk. Multiple categories of animals (M346 = generic animal, M367 = livestock variant, M006 = animal variant) are counted in groups. The structure is: institutional header (M387), followed by repeated [animal-type, count] pairs, closed by totals. This is a herd inventory—how many animals of each type belong to this institution.

8.3 P009412 (CahDAFI 1, 58 02) — Commodity Ledger

A 22-sign tablet with 8 numerical tokens:

Pos.	Sign	Reading
1	M388	ACCOUNT HEADER
2–3	M066, M352	MODIFIER, Labor/overseer
	N01	1
4	M370	COMMODITY (agricultural product)
5–6	M218, M066	GRAIN, MODIFIER
7	M352	Labor/overseer
	N01	1
8–9	M096, M066	LABOR/WORKER, MODIFIER
10	M352	Labor/overseer
11	M347	Settlement/transaction
	N01	1
12–13	M218, M218	GRAIN, GRAIN
14	M259	LABOR TITLE
15	M263	INSTITUTIONAL DESIGNATION
16	M329	Settlement/transaction
	N01	1
17	M387	INSTITUTION/OFFICIAL
18	M110	QUALITY MARKER
19	M218	GRAIN
	N01	1
20	M387	INSTITUTION/OFFICIAL
21	M370	COMMODITY
22	M218	GRAIN
	N01, N01	2

Reading: Account header opens the tablet. Three commodity entries follow: modified items (M066+M352), grain (M218), and an agricultural product (M370). The administrative section names an institutional designation (M263) and two official titles (M387). The tablet records allocations of commodities and grain across institutional offices. Eight numerical tokens give individual line-item counts.

8.4 P009467 (M-1156) — Complex Administrative Record

A 25-sign tablet with 21 numerical tokens—one of the longest from outside Susa:

Pos.	Sign	Reading
1–2	M388, M388 N01, N01	ACCOUNT, ACCOUNT (double header) 2
3	M263 N14	INSTITUTIONAL DESIGNATION 10
4	M388	ACCOUNT
5	M032	TEXTILE/WOOL
6	M377 N01	ALLOCATION 1
7	M388	ACCOUNT
8	M059	CONTAINER
9	M096	LABOR/WORKER
10	M218 N01	GRAIN 1
11	M024	TRANSACTION TYPE
12	M388	ACCOUNT
13	M218 N01	GRAIN 1
14	M288	HEADER/TOTAL
15	M371	PERSON/CONTAINER
16	M377 N14, N01	ALLOCATION 11
17	M297	ADMIN CLASSIFIER
18	M066 N01	MODIFIER 1
19–20	M371, M371	PERSON, PERSON
21	M218 N14, N01	GRAIN 11
22	M388	ACCOUNT
23	M001 N01	DISTRIBUTION RECORD 1
24–25	M388, M388 N23, N34, N14, N01	ACCOUNT, ACCOUNT (closing) grand total

Reading: A multi-section account. Opens with a double account header. Records textile (M032), containers (M059), labor (M096), and grain (M218) across multiple sub-accounts (each opened by M388). The administrative section names a commodity allocation (M066) to personnel (M371). Closes with a distribution record (M001) and a grand total. This tablet demonstrates the full HEADER–COMMODITY–ADMIN–TOTAL structure with nested sub-accounts.

8.5 P008003 (MDP 06, 203) — Susa Administrative Tablet

A 34-sign tablet from the Susa MDP collection:

Pos.	Sign	Reading
1	M157	NUMERICAL MARKER
2	M302	ADMINISTRATIVE ORDER
3	M387	INSTITUTION/OFFICIAL
4	M009	PERSON
	N01	1
5	M264	ADMIN VARIANT
6	M297	ADMIN CLASSIFIER
	N39B	capacity measure
7	M340	ADMIN/COMMODITY
8	M372	OBLIGATION
9	M263	INSTITUTIONAL DESIGNATION
10	M354	SPECIFIC ANIMAL/PRODUCT
	N01	1
11	M153	PERSONNEL
12	M388	ACCOUNT
13	M260	ADMIN (“reveal”)
14	M354	SPECIFIC ANIMAL/PRODUCT
	N01	1
15	M387	INSTITUTION/OFFICIAL
16	M388	ACCOUNT
17	M266	INSTITUTION TYPE
18	M036	ADMIN TITLE
	N14, N01	11
19	M388	ACCOUNT
20–21	M228, M296	TEMPORAL MARKER, LABOR CLASSIFIER
22	M296	LABOR CLASSIFIER
23	M057	LABOR ACCOUNT
24	M354	SPECIFIC ANIMAL/PRODUCT
	N14, N01	11
25	M305	COMMODITY/RATIONS
26	M388	ACCOUNT
27–28	M128, M128	PRODUCTION UNIT, PRODUCTION UNIT
29	M096	LABOR/WORKER
	N01	1
30	M354	SPECIFIC ANIMAL/PRODUCT
31	M387	INSTITUTION/OFFICIAL
32	M057	LABOR ACCOUNT
	N01	1
33	M354	SPECIFIC ANIMAL/PRODUCT
34	M288	HEADER/TOTAL
	N39B, N24, N34	grand total (capacity system)

Reading: A complex Susa ledger. Opens with a numerical header (M157) and institutional framing (M387, M009). Records multiple entries of specific animal products (M354 appears 5 times) across different institutional designations (M263, M266, M036). Labor accounts (M057, M096) track workers. Closes with M288 (grand total) and capacity-system numerals. This is the kind of tablet that produces $tw = 116$: every sign depends on multiple contextual positions, creating dense co-occurrence structure.

8.6 P008013 (MDP 06, 213) — Textile Production Record

A 34-sign Susa tablet dominated by textile signs:

Pos.	Sign	Reading
1	M157	NUMERICAL MARKER
2	M391	SUBTOTAL
3	M038	DISTRIBUTION
4	M388	ACCOUNT
5	M387	INSTITUTION/OFFICIAL
	N01	1
6	M057	LABOR ACCOUNT
7–8	M376, M149	COMMODITY ACCOUNT, TEXTILE
	N14, N01	11
9	M246	LABOR ALLOCATION
10	M009	PERSON
11	M371	PERSON/CONTAINER
12–13	M376, M207	COMMODITY ACCOUNT, YIELD MARKER
	N01	1
14	M388	ACCOUNT
15–16	M110, M242	QUALITY MARKER, STANDARD MEASURE
17	M096	LABOR/WORKER
18–19	M376, M323	COMMODITY ACCOUNT, CURRENT PERIOD
20–21	M149, M391	TEXTILE, SUBTOTAL
	N34, N14	70
22–24	M128, M377, M347	PRODUCTION UNIT, ALLOCATION, TRANSACTION RECORD
25	M371	PERSON/CONTAINER
	N14	10
26–28	M207, M376, M149	YIELD MARKER, COMMODITY, TEXTILE
	N24, N14, N01	35
29–30	M296, M296	LABOR CLASSIFIER, LABOR CLASSIFIER
31	M066	MODIFIER
	N14	10
32–33	M376, M149	COMMODITY, TEXTILE
34	M153	PERSONNEL
	N34, N01	61

Reading: A textile production account. The sign pair M376+M149 (commodity+textile) appears four times, each followed by quantities. Workers (M096, M371) and institutional officials (M387) are named. Subtotals (M391) punctuate the account. The closing entry records 61 units of textile against a personnel marker. This tablet tracks textile output by worker and institution—a production report.

9 Results VI: Cross-Script Parallels

9.1 Indus Valley Script

The contextual decoding pipeline measured positional parallels between Proto-Elamite signs and the Indus Valley corpus. The pipeline assigned each Proto-Elamite sign an Indus positional profile—INITIAL, MEDIAL, TERMINAL, or FLEXIBLE—based on where structurally equivalent signs appear in Indus seal sequences.

Of 61 signs decoded by contextual analysis:

- 52 signs (85.2%) have MEDIAL Indus parallels—they appear in the middle of sequences in both scripts. Content words occupy this position: commodity names, modifiers, classifiers.

- 4 signs (6.6%) have INITIAL parallels—they open sequences. These correspond to header and account signs.
- 3 signs (4.9%) have TERMINAL parallels—they close sequences. These correspond to total/summary signs.
- 2 signs (3.3%) are FLEXIBLE—they appear in multiple positions.

The positional parallel is structural, not phonetic. Both scripts place classifiers and headers at sequence boundaries and content words in medial positions. This shared organizational principle is expected if both scripts serve administrative bookkeeping—the genre constrains the grammar.

10 Validation

Seven independent lines of evidence converge:

1. **Treewidth gradient matches known hierarchy.** Susa (tw = 116) versus periphery (tw = 2–7) reproduces the administrative structure identified by archaeological excavation.
2. **Entropy places the script correctly.** $H = 6.83$ bits/sign falls in the logosyllabic range, between rongorongo (7.08) and Linear B (6.35). This is consistent with a logographic writing system, not a syllabary or alphabet.
3. **Topic model finds only economic topics.** All 15 interpretable topics map to accounting functions: commodities, labor, textiles, animals, institutions, containers, distributions. No religious or narrative topics appear.
4. **Tablet structure matches prior work.** The HEADER → COMMODITY → ADMIN → TOTAL structure confirmed by the dictionary matches the layout identified through numerical analysis alone by Nissen, Damerow, and Englund (1993).
5. **Known signs validate.** The seven signs with prior scholarly assignments (M388 = account type per Dahl 2005; M157 = numerical per Damerow & Englund 1989; M218 = grain context per Dahl 2005; M370 = agricultural product per distributional analysis; M066 = modifier per distributional analysis; M387 = institutional per Dahl 2005; M263 = administrative per co-occurrence with M387) all receive consistent assignments from the LLM pipeline.
6. **Per-site sign distributions cohere.** Universal signs are structural; Susa-only signs are specialized. Administrative hierarchy theory predicts this pattern, and every comparable ancient bureaucracy exhibits it (Ur III Sumerian, Old Kingdom Egyptian, Mycenaean Linear B).
7. **Cross-linguistic metaphor converges.** Cross-linguistic mapping assigns M269 the reading LORD (“flowing radiance”). The same light-as-authority metaphor appears in Sumerian EN, where “lord” derives from “bright/luminous one.” Proto-Elamite and early Sumerian occupied the same trade network and chronological window (3100–2900 BCE). The convergence is semantic, not phonetic—two independent administrations mapped the same conceptual metaphor onto their highest-ranking title.

11 Content Classification

The vocabulary is economic. Sorting the 15 interpretable topics by total token coverage:

Domain	Percentage of corpus
Administrative/institutional	28.4%
Commodity accounts	18.7%
Labor/textile production	14.2%
Animal husbandry	11.3%
Distribution/rations	8.1%
Containers/vessels	6.4%
Header/totals (structural)	12.9%

No religious vocabulary appears in the high-frequency signs. No royal inscriptions. No narrative. Proto-Elamite is bookkeeping—livestock records, textile production accounts, grain distributions, personnel accounts, institutional ledgers.

12 Comparison with Other Unread Scripts

Script	tw	Entropy	Corpus	Status
Proto-Elamite	119	6.83	15,534	Ross 2026
Linear B	77	6.35	5,000+	Deciphered (Ventris 1952)
Linear A	55	6.10	4,249	Partially read
Rongorongo	41	7.08	14,906	Decoded (Ross 2026)
Indus Valley	26	5.80	1,003	Structurally decoded

Proto-Elamite has the highest treewidth of any ancient writing system measured. This reflects the combinatorial complexity of centralized administration at Susa—not linguistic complexity per se, but bookkeeping complexity. The number of variables tracked simultaneously (commodities, agents, institutions, quantities, transaction types) exceeds what any other known ancient administration recorded in a single document.

13 Conclusion

One graph-theoretic measurement decodes the administrative structure of the oldest unread script on Earth. The $58\times$ treewidth differential between Susa and peripheral sites maps the hierarchy of an empire that has been silent for 5,100 years. A dictionary of 349 candidate readings—constructed by three language models running on a laptop, seven supported by prior scholarship—reads the tablets as livestock records, textile production accounts, grain distributions, and personnel ledgers. The HEADER–COMMODITY–ADMIN–TOTAL structure, identified independently by numerical analysis three decades ago, now has semantic content.

Proto-Elamite is not mysterious. It is an accounting ledger. The civilization that invented writing independently of Mesopotamia used it for the same purpose: to keep track of what belonged to whom.

A Complete Dictionary

Table 5 in Section 5 presents the top 30 signs. The remaining 319 entries appear below.

Table 6: Proto-Elamite dictionary: signs ranked 31–349.

Rank	Sign	Count	Meaning
31	M365	109	LIGHT
32	M106	109	PROMINENT/PLOW/LABOR
33	M175	106	BACK
34	M038	104	PROMINENT
35	M352	102	LABOR
36	M362	101	DIVINE
37	M059	98	UNDETERMINED
38	M260	98	LIGHT/BRIGHT
39	M320	97	UNDETERMINED
40	M131	96	SEVEN
41	M296	96	UNDETERMINED
42	M342	96	DEC
43	M206	93	BRIGHT/DIVINE
44	M269	92	LORD (<i>lán liú guāng máng</i> = flowing radiance)
45	M219	90	DIVINE
46	M347	88	EXPLAIN/INTRODUCE
47	M056	87	DIVINE/LABOR
48	M195	87	REAL
49	M048	83	LIGHT
50	M010	77	UNDETERMINED
51	M386	75	MANDATE
52	M373	75	SEVEN/PROMINENT
53	M203	74	UNDETERMINED
54	M262	74	UNDETERMINED
55	M102	73	SEVEN
56	M259	72	LABOR
57	M317	71	DIVINE
58	M240	69	LIGHT
59	M005	69	UNDETERMINED
60	M101	67	UNDETERMINED
61	M149	65	PROMINENT/LABOR
62	M024	64	TRUST/UNUSUAL
63	M111	64	UNDETERMINED
64	M006	62	BACK
65	M295	62	UNDETERMINED
66	M304	62	UNDETERMINED
67	M243	59	DIVINE
68	M264	59	SUN
69	M217	59	BACK
70	M146	54	UNDETERMINED
71	M380	54	LABOR
72	M309	54	LABOR
73	M145	53	LIGHT
74	M242	53	CONSTANT
75	M348	53	UNDETERMINED
76	M292	53	UNDETERMINED

Rank	Sign	Count	Meaning
77	M351	52	UNDETERMINED
78	M318	52	DIVINE
79	M097	51	DIVINE/LABOR
80	M325	50	LIGHT
81	M002	50	LABOR
82	M338	50	DIVINE/LABOR
83	M223	48	LIGHT/BRIGHT/DIVINE
84	M099	48	PROMINENT/POTTERY/INTRODUCE/LABOR
85	M046	47	DIVINE
86	M254	47	UNDETERMINED
87	M075	46	RADIANCE
88	M029	45	MANDATE
89	M210	44	UNDETERMINED
90	M314	43	LIGHT
91	M341	43	SUN
92	M033	42	LIGHT
93	M081	42	UNDETERMINED
94	M329	41	MANDATE
95	M139	41	UNDETERMINED
96	M041	40	SUN
97	M128	40	PRESS/BRIGHT/LABOR
98	M384	40	SEVEN
99	M228	39	LIGHT
100	M332	39	BACK
101	M372	38	PROMISE
102	M301	38	CONSTANT
103	M266	37	MANDATE/DIVINE
104	M265	36	UNDETERMINED
105	M051	35	SEVEN
106	M391	34	BRIGHT/DIVINE
107	M281	34	PROMINENT
108	M390	34	ABUNDANT
109	M383	33	UNDETERMINED
110	M246	33	PROMINENT/LABOR
111	M207	33	BACK
112	M340	32	LABOR/INTRODUCE
113	M105	32	POTTERY
114	M247	31	COVENANT
115	M147	31	LABOR
116	M250	30	DEC
117	M374	29	PROMINENT/MANDATE
118	M321	29	PROMINENT
119	M379	29	PROMINENT/RADIANCE
120	M039	28	PROMINENT/DIVINE/LABOR
121	M110	27	BRIGHT
122	M251	27	BACK
123	M122	27	UNDETERMINED
124	M103	27	MANDATE
125	M230	26	UNDETERMINED
126	M112	26	PROMINENT
127	M417	25	REAL
128	M248	24	SUN/MANDATE

Rank	Sign	Count	Meaning
129	M049	24	INTRODUCE
130	M125	23	UNDETERMINED
131	M302	23	MANDATE
132	M278	23	UNDETERMINED
133	M323	22	PRESENT
134	M205	22	UNDETERMINED
135	M220	22	ABUNDANT
136	M286	22	SEVEN/INTRODUCE
137	M069	22	DIVINE
138	M050	22	PROMINENT
139	M319	21	LIGHT
140	M073	21	LIGHT
141	M080	21	INTRODUCE
142	M261	21	BRIGHT/MANDATE
143	M418	21	LIGHT
144	M381	20	BRIGHT/SEVEN
145	M158	20	LABOR
146	M291	20	PROMINENT/FLOW/POTTERY
147	M044	19	SEVEN
148	M209	19	DIVINE
149	M316	18	UNDETERMINED
150	M109	18	UNDETERMINED
151	M328	17	UNDETERMINED
152	M193	17	UNDETERMINED
153	M249	17	ABUNDANT
154	M285	17	PRESS/DIVINE
155	M369	17	UNDETERMINED
156	M221	16	BACK
157	M324	16	PROMINENT/LABOR
158	M315	16	UNDETERMINED
159	M252	16	LABOR
160	M343	15	INTRODUCE
161	M293	15	ABUNDANT/DIVINE/COMMONER
162	M180	14	LABOR
163	M312	14	SEVEN/UNUSUAL
164	M277	14	UNDETERMINED
165	M123	13	MANDATE
166	M134	13	SUN/INTRODUCE/MANDATE
167	M222	13	MANDATE
168	M311	13	LABOR
169	M035	13	LIGHT
170	M361	13	UNDETERMINED
171	M144	13	LABOR/MANDATE/PLOW
172	M026	13	PROMINENT
173	M151	12	UNDETERMINED
174	M053	12	PROMISE
175	M310	12	INTRODUCE/LABOR
176	M382	12	MANDATE
177	M152	11	INTRODUCE
178	M058	11	LABOR
179	M061	11	MANDATE/INTRODUCE
180	M226	10	DIVINE/SUN

Rank	Sign	Count	Meaning
181	M214	10	INTRODUCE
182	M115	10	LABOR/SUN/POTTERY
183	M298	10	UNDETERMINED
184	M461	10	BACK
185	M064	10	LABOR
186	M375	10	LABOR
187	M047	9	UNDETERMINED
188	M143	9	BRIGHT
189	M244	9	PROMINENT/ABUNDANT
190	M208	9	LABOR
191	M393	9	LABOR/PROMINENT
192	M290	8	UNDETERMINED
193	M270	8	BACK
194	M037	8	LIGHT
195	M176	8	LIGHT
196	M268	8	DISTRICT
197	M356	8	LIGHT
198	M224	7	UNDETERMINED
199	M294	7	DISTRICT
200	M188	7	UNDETERMINED
201	M052	7	UNDETERMINED
202	M201	6	BRIGHT/SEVEN/DIVINE
203	M150	6	LABOR
204	M007	6	PROMINENT/PRESS/INTRODUCE
205	M016	6	UNDETERMINED
206	M014	6	MANDATE
207	M141	6	LIGHT
208	M359	6	CONSTANT
209	M271	6	LABOR
210	M074	6	UNDETERMINED
211	M126	6	LIGHT/BRIGHT/MANDATE
212	M086	5	BACK
213	M063	5	LABOR/DIVINE
214	M034	5	UNDETERMINED
215	M358	5	DIVINE
216	M368	5	LIGHT
217	M076	5	LIGHT
218	M447	5	UNDETERMINED
219	M449	5	MANDATE
220	M289	5	LIGHT
221	M177	4	UNDETERMINED
222	M357	4	LIGHT/BRIGHT/LABOR
223	M308	4	CONSTANT
224	M025	4	PROMINENT
225	M068	4	INTRODUCE
226	M129	4	UNDETERMINED
227	M283	4	PROMINENT/LABOR
228	M385	4	UNDETERMINED
229	M098	4	UNDETERMINED
230	M284	4	LIGHT
231	M104	4	SUN
232	M011	4	UNDETERMINED

Rank	Sign	Count	Meaning
233	M062	4	LIGHT
234	M030	4	LIGHT
235	M012	4	SEVEN
236	M107	4	PROMINENT/LIGHT/LABOR
237	M276	4	PROMINENT
238	M482	4	DIVINE/LABOR
239	M089	3	POTTERY
240	M462	3	BACK
241	M135	3	UNDETERMINED
242	M255	3	UNDETERMINED
243	M213	3	UNDETERMINED
244	M140	3	DIVINE
245	M094	3	SEVEN
246	M154	2	BACK
247	M282	2	UNDETERMINED
248	M256	2	WEIGHT/LIGHT
249	M458	2	LIGHT/SEVEN
250	M015	2	UNDETERMINED
251	M017	2	DIVINE/INTRODUCE/PROMINENT/NEXT DAY
252	M087	2	LABOR
253	M184	2	REAL
254	M088	2	INTRODUCE/MANDATE
255	M504	2	UNDETERMINED
256	M167	2	LIGHT/BRIGHT/DIVINE
257	M267	2	LIGHT
258	M093	2	LIGHT
259	M229	2	UNDETERMINED
260	M090	2	UNDETERMINED
261	M322	2	SEVEN/ABUNDANT
262	M438	2	UNDETERMINED
263	M420	2	UNDETERMINED
264	M503	2	ABUNDANT/DIVINE
265	M155	2	BRIGHT
266	M510	2	BACK
267	M349	2	MANDATE
268	M339	2	UNDETERMINED
269	M159	2	DIVINE
270	M330	2	SEVEN/UNUSUAL
271	M470	2	INTRODUCE
272	M471	2	DIVINE
273	M142	2	DIVINE/INTRODUCE
274	M344	2	INTRODUCE
275	M335	2	DIVINE/SEVEN
276	M127	2	UNDETERMINED
277	M489	2	UNDETERMINED
278	M492	2	LIGHT/BRIGHT/DIVINE
279	M299	1	PROMINENT
280	M287	1	DIVINE/UNUSUAL/SEVEN
281	M486	1	SUN/PROMINENT
282	M495	1	DIVINE/MANDATE/UNUSUAL/PROMINENT
283	M496	1	UNDETERMINED
284	M045	1	UNDETERMINED

Rank	Sign	Count	Meaning
285	M513	1	PROMINENT/MANDATE
286	M257	1	PROMISE
287	M197	1	PROMINENT/DIVINE
288	M500	1	LIGHT/BRIGHT/DIVINE
289	M120	1	MANDATE
290	M501	1	DIVINE
291	M114	1	SEVEN
292	M412	1	SUN/DIVINE
293	M421	1	UNDETERMINED
294	M272	1	PROMINENT
295	M378	1	DEC
296	M117	1	SEVEN/BRIGHT/WAVE/LABOR
297	M366	1	BRIGHT/SEVEN/LABOR/DIVINE
298	M227	1	LABOR
299	M084	1	SEVEN
300	M505	1	LIGHT
301	M253	1	UNDETERMINED
302	M424	1	SEVEN
303	M431	1	UNDETERMINED
304	M433	1	LIGHT/SEVEN
305	M509	1	UNDETERMINED
306	M077	1	UNDETERMINED
307	M441	1	SEVEN
308	M444	1	PROMINENT/SEVEN
309	M445	1	UNDETERMINED
310	M446	1	BACK
311	M215	1	CONSTANT/BACK
312	M448	1	LIGHT
313	M130	1	INTRODUCE
314	M450	1	SEVEN
315	M452	1	PROMINENT
316	M506	1	PRESS/SEVEN
317	M453	1	DIVINE/MANDATE
318	M173	1	BACK
319	M455	1	UNDETERMINED
320	M021	1	PROMINENT
321	M465	1	INTRODUCE
322	M466	1	SUN/DIVINE
323	M467	1	DIVINE/PROMINENT
324	M468	1	EXPLAIN
325	M469	1	SUN
326	M454	1	EXPLAIN/PROMINENT
327	M363	1	SUN
328	M018	1	LIGHT
329	M360	1	UNDETERMINED
330	M172	1	LABOR
331	M165	1	LIGHT/BRIGHT/DIVINE
332	M043	1	UNDETERMINED
333	M241	1	LIGHT/PROMINENT/UNUSUAL
334	M479	1	BACK
335	M402	1	PROMINENT
336	M401	1	SUN

Rank	Sign	Count	Meaning
337	M511	1	LIGHT
338	M326	1	PROMINENT/LABOR
339	M040	1	UNDETERMINED
340	M353	1	UNDETERMINED
341	M410	1	REAL
342	M204	1	DIVINE
343	M085	1	SUN
344	M411	1	DIVINE/SEVEN
345	M600	1	LIGHT
346	M512	1	BACK/PROMISE
347	M071	1	LIGHT
348	M490	1	RADIANCE/SEVEN
349	M491	1	UNDETERMINED

References

- [1] J. L. Dahl, “Complex Graphemes in Proto-Elamite,” *CDLI Journal*, 2005:3, 2005.
- [2] P. Damerow and R. K. Englund, *The Proto-Elamite Texts from Tepe Yahya*, Cambridge, MA: Peabody Museum, 1989.
- [3] H. J. Nissen, P. Damerow, and R. K. Englund, *Archaic Bookkeeping: Early Writing and Techniques of Economic Administration in the Ancient Near East*, University of Chicago Press, 1993.
- [4] L. Ross, “Reading rongorongo,” Zenodo, March 2026. DOI: 10.5281/zenodo.19362491.
- [5] L. Ross, *Shadow & Mirror: Complementarity of Computation and Consciousness*, Zenodo, March 2026. DOI: 10.5281/zenodo.19263974.
- [6] Cuneiform Digital Library Initiative, <https://cdli.mpiwg-berlin.mpg.de/>.