

# Compositionality of Complex Graphemes in the Undeciphered Proto-Elamite Script using Image and Text Embedding Models

Logan Born<sup>1</sup>

loborn@sfu.ca

Kathryn Kelley<sup>2</sup>

kathryn.kelley@utoronto.ca

M. Willis Monroe<sup>3</sup>

willis.monroe@ubc.ca

Anoop Sarkar<sup>1</sup>

anoop@cs.sfu.ca

<sup>1</sup>Simon Fraser University  
School of Computing Science

<sup>2</sup>University of Toronto  
Department of Near and  
Middle Eastern Civilizations

<sup>3</sup>University of British Columbia  
Department of Asian Studies

## Abstract

We introduce a language modeling architecture which operates over sequences of images, or over multimodal sequences of images with associated labels. We use this architecture alongside other embedding models to investigate a category of signs called complex graphemes (CGs) in the undeciphered proto-Elamite script. We argue that CGs have meanings which are at least partly compositional, and we discover novel rules governing the construction of CGs. We find that a language model over sign images produces more interpretable results than a model over text or over sign images and text, which suggests that the names given to signs may be obscuring signals in the corpus. Our results reveal previously unknown regularities in proto-Elamite sign use that can inform future decipherment efforts, and our image-aware language model provides a novel way to abstract away from biases introduced by human annotators.

## 1 Introduction

This work sets out to understand a category of signs called complex graphemes (CGs) in the undeciphered proto-Elamite (PE) script, a writing system from ancient Iran dating to approximately 3300-2900 BC (Dahl et al., 2013).<sup>1</sup> PE is partly contemporaneous with the world’s other two earliest writing systems, Egyptian hieroglyphs and proto-cuneiform, and is the least deciphered of the three, with the underlying language(s) remaining unknown. PE was used exclusively as an accounting technology, employing several numerical systems whose bundling principles are known. Although written in continuous lines, PE, like proto-cuneiform, is most comparable to an accountant’s spreadsheet; some structures and rules governing

<sup>1</sup>Our code, data, and trained models are available at <https://github.com/sfu-natlang/pe-compositionality>

sign use have been identified (Hawkins, 2015; Dahl et al., 2018; Englund, 2004).

The corpus consists of approximately 1500 published clay tablets from excavations in Iran, almost all of which exist in electronic transliteration following the conventions of a work-in-progress sign list (Dahl, 2006). As with other decipherments, understanding the nature of signs and the nuances of sign use is as important as identifying the underlying language(s). Meaningful information can be recovered and the texts partly “read” even if the language remains unknown.

To better understand sign usage in PE, this work proposes an architecture for image-aware language modeling, which permits sharing information between visually similar signs much as sub-word units share information between words. We use sign embeddings to demonstrate patterns which are not readily apparent due to the complexity of the accounting system and the large number of sign shapes found in the script. Our analysis offers insights on complex graphemes that can aid in future hypothesis generation. We confirm that some transliteration choices by PE specialists capture meaningful semantic divisions in the script; this is not a trivial fact, due to the large number of similar looking signs. By using image-aware models, we also observe that some signs with distinct names receive very similar embeddings, implying a functional equivalence that could be exploited by merging signs to create a less sparse corpus that is more amenable to analysis by NLP methods.

## 2 Methodology

As described by Dahl (2005), CGs in proto-Elamite are signs that consist of one sign inscribed within another (transliterated |S1+S2|), or of one sign framed by two instances of another (|S1+S2+S1|). Rarely, S1 and S2 occur connected at the side, as

in  $\text{IM296}+\text{M296}$   $\{\cup\}$ . We refer to S1 as the *outer* sign and S2 as the *inner* sign, though we acknowledge this terminology is not quite appropriate in cases like  $\text{IM296}+\text{M296}$ . Most signs which occur as part of a CG can also occur as standalone signs. Exceptions to this are rare, such as M600 which only ever occurs in the hapax  $\text{IM362}+\text{M600}$ .

Although these signs are *orthographically* compositional, it is not known whether they are also semantically compositional. Similar constructions exist in proto-cuneiform (PC), including “containers” with signs inscribed to indicate specific products (Wagensonner, 2015). Some PC compounds survive into later cuneiform, and sometimes have idiomatic meanings, e.g. cuneiform  $\text{GU}_7$  “eat”, a combination of “head” and “bowl”. Chinese characters likewise exhibit varying degrees of visual and semantic compositionality (Sprout, 2006).

Past work (Mikolov et al., 2013b; Salehi et al., 2015; Cordeiro et al., 2016) suggests that embedding models capture semantic compositionality in noun compounds and multiword expressions. Often, these models assign a compound a representation which is similar to the sum of the representations of the words in the compound. Thus we predict that if CGs are semantically compositional, their embeddings will be additively compositional at a higher rate than expected by chance. Their embeddings may also exhibit other signs of internal structure, such as the ability to model proportional analogy between CGs with shared components:

$\text{IM136}+\text{M365}$  :  $\text{M136}$  ::  $\text{IM327}+\text{M365}$  :  $\text{M327}$



If this analogy holds in the embedding space (which is to say that the 3CosAdd formula  $|\text{IM136}+\text{M365}| - \text{M136} + \text{M327} \approx |\text{IM327}+\text{M365}|$  holds between the signs’ embeddings) this would give further evidence that the CGs involved have some degree of semantic compositionality.

Unfortunately, most PE signs are rare, which impedes a model’s ability to learn meaningful information about their distributions. Yet many signs with distinct names have striking visual resemblances, and it is usually not known whether they have different meanings. Visual information may therefore help an embedding model by allowing it to share distributional information across graphically similar signs. To this end, we propose an architecture for multimodal language modeling in Figure 1. This architecture uses two separate embedding components. On the left of Figure 1, in

red, is a standard embedding layer which replaces a one-hot input with a small, learnable representation. On the right, in blue, a lookup function retrieves an image of the sign represented by the input. A CNN extracts a feature vector from the image, which is max-pooled, flattened, and passed through a dense layer to produce a low-dimensional embedding. Both embeddings are concatenated and fed to a BiLSTM<sup>2</sup> (Hochreiter and Schmidhuber, 1997; Schuster and Paliwal, 1997) which attempts to predict the name of the next sign in the text. All timesteps share the same weights for the CNN and embedding layers. By omitting the blue image-embedding component we can obtain a normal BiLSTM language model. By omitting the red text-based component, we can obtain an image-only model which never directly sees the labels assigned to the signs in the corpus.

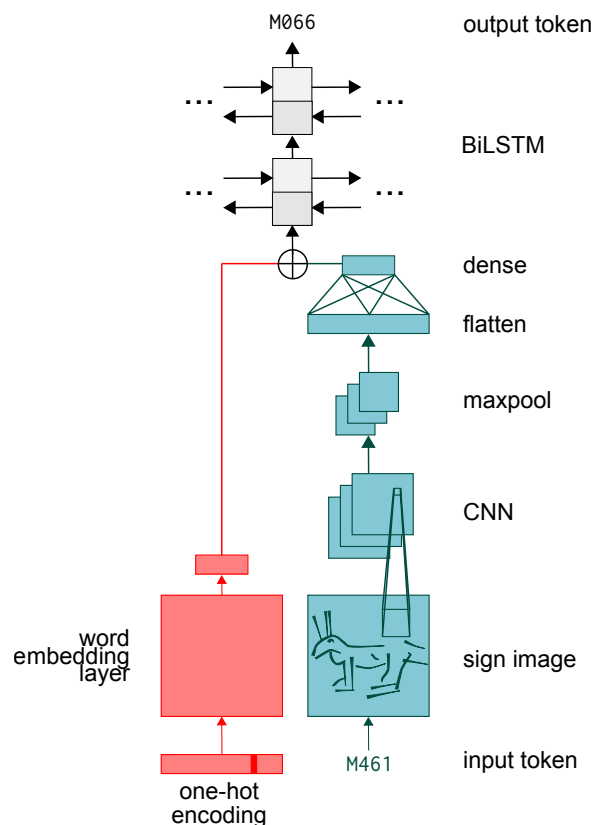


Figure 1: Architecture for image-aware, multimodal language modeling.

To verify that this architecture captures distributional properties of signs, and not just visual properties, we train a separate image recognition model to predict a sign’s name given only its image.

<sup>2</sup>We also attempted to train a Transformer model (Vaswani et al., 2017), but the corpus size proved insufficient and it always underperformed compared to the BiLSTM.

Model	Input Type	Embedding Sizes <sup>3</sup>	Other Parameters	Description
glove	seq. of sign names	16, 32, 64, 128, 256	window size: 15	Pennington et al. 2014
fasttext.cbow	seq. of sign names	16, 32, 64, 128, 256	window size: 15	Bojanowski et al. 2017
fasttext.skip	seq. of sign names	16, 32, 64, 128, 256	window size: 15	Bojanowski et al. 2017
word2vec.cbow	seq. of sign names	16, 32, 64, 128, 256	window size: 15	Mikolov et al. 2013a
word2vec.skip	seq. of sign names	16, 32, 64, 128, 256	window size: 15	Mikolov et al. 2013a
lm.text	seq. of sign names	64	hidden dimension: 64	Figure 1, blue (image embedding) omitted.
lm.image+text	seq. of sign names and images	64	hidden dimension: 64 image size: 64×64	Figure 1.
lm.image	seq. of sign images	64	hidden dimension: 64 image size: 64×64	Figure 1, red (text embedding) omitted.
image recognition	individual sign image	64	image size: 64×64	Figure 1, blue (image embedding) only.

Table 1: List of models considered in this work.

This model uses the blue image embedding component from Figure 1 to produce a representation of an input image; a dense layer predicts the name of the sign from this embedding. This model only sees signs in isolation, meaning it will not learn from distributional information. If a result holds for the multimodal LM but not for this image recognition model, this implies that the result arises from contextual information in the text, and not simply from visual resemblances between signs.

We also train CBoW and skipgram models with FastText<sup>4</sup> (Bojanowski et al., 2017) and word2vec (Mikolov et al., 2013a), as well as GloVe embeddings (Pennington et al., 2014). Table 1 summarizes all of the models used in this work and important hyperparameters. We train these models on the PE corpus from Born et al. (2019), which is a cleaned version of texts originally published by the Cuneiform Digital Library Initiative (CDLI). This contains digitized transliterations from 1399 tablets comprising 11013 lines in total, or 33778 tokens. 7508 tokens represent broken or unreadable signs, and another 11364 represent numerals, leaving only 14906 non-numerical tokens. 1107 tokens (comprising nearly half the sign types in our cleaned data) are labeled as CGs. We treat each entry of a tablet as a single input sentence for training LMs, and set aside 500 lines as a validation set.

Prior to training, we replace all signs occurring 3 or fewer times<sup>5</sup> with UNK. We replace rare signs wherever they occur, including inside of CGs. The tokens X and . . . represent broken or unreadable signs, so we also replace these with UNK. When

<sup>3</sup>To give a fairer comparison, we train the simpler models with several embedding sizes and report results from whichever dimensionality performs best on each task. Additional model information is in the supplemental material.

<sup>4</sup>Sign names are largely arbitrary, so we disable sub-words in FastText by setting the maximum sub-word length to 0.

training language models, we do not backpropagate losses from samples where the target word is UNK, since it so often represents broken material. To make the data less sparse, we remove annotations marking sign variants, so that for example M157 and M157~a are considered the same sign.

### 3 Experimental Results

#### 3.1 Additive Composition

We predict that if a CG is semantically compositional, its embedding will approximately equal the sum of the embeddings of the signs it comprises.

Given a sign  $s$ , let  $e_s$  denote the embedding of  $s$ . If  $s$  is a CG let  $\sigma(s)$  denote the list of signs which make up  $s$ . For every CG  $s$  in the signary, we check whether  $\sum_{t \in \sigma(s)} e_t \approx e_s$ . If  $\sum_{t \in \sigma(s)} e_t$  is within the  $k$  nearest neighbors of  $e_s$  for some threshold  $k$ , we say that  $s$  appears to have a compositional representation.

For different thresholds  $k$ , we measure how many CGs have compositional representations. Since many PE signs have low frequency, we predict that noise may drown out any signal when  $k$  is small. However, when  $k$  is large enough to overcome this noise, we predict that the number of CGs with compositional representations will be greater than expected by chance, as we expect that some CGs have meanings which are semantically compositional rather than idiomatic. Table 2 shows the results from this evaluation.

In text-only models, when  $k$  is small the number of CGs with compositional representations is no higher than expected by chance. However, for image-aware models, and for text-only models with large  $k$ , the number of CGs which are close to the

<sup>5</sup>To determine frequency, we count how often a sign occurs both independently and as part of a CG.

model	$k$				
	1	3	5	10	15
glove.256	0	0	1	3	<b>13</b>
word2vec.cbowl6	0	0	1	9	<b>12</b>
word2vec.skip.32	0	0	5	<b>13</b>	<b>16</b>
fasttext.cbowl28	0	2	3	5	9
fasttext.skip.128	0	3	<b>10</b>	<b>15</b>	<b>20</b>
lm.text.64	0	0	0	0	1
lm.image+text.64	1	<b>14</b>	<b>21</b>	<b>40</b>	<b>51</b>
lm.image.64	<b>11</b>	<b>16</b>	<b>27</b>	<b>48</b>	<b>61</b>
image recognition.64	<b>3</b>	<b>7</b>	<b>15</b>	<b>28</b>	<b>38</b>

Table 2: Number of compositional CGs for different similarity cutoffs  $k$ . Bold numbers represent cases where the number of compositional graphemes is significantly larger than expected by chance.

sum of their components is significant. Even for  $k = 15$ , the signs identified as compositional by `lm.image.64` average  $>0.97$  cosine similarity to the sum of their parts, suggesting this is not too generous a threshold.

Notably, the number of compositional CGs in `lm.image.64` is always larger than the number in any of the other models, including the image recognition model.<sup>6</sup> This has the important implication that compositionality in the embeddings is not solely a consequence of visual compositionality. If that were the case, the contextual information available to the LM would not be useful for this task, and the image LM would not be expected to find more compositional CGs than the image recognition model. Moreover we would not expect to find a significant amount of compositionality in any of the text-only models for any  $k$ . Table 3 shows examples of signs which appear to be compositional in the image LM but not the image recognition model. These are signs for which contextual information plays a deciding role in making them appear semantically compositional, and which may therefore be of interest to analyze in future work.

M153	+ M106	$\approx$  M153+M106
M175	+ M286	$\approx$  M175+M286
M327	+ M348	$\approx$  M327+M348
M362	+ M244	$\approx$  M362+M244
M157	+ M288	$\approx$  M157+M288
M175	+ M153	$\approx$  M175+M153
M218	+ M388	$\approx$  M218+M388

Table 3: Sample of signs which appear to be compositional in the image LM but not the image recognition model.

We emphasize that the text-only models have no information about sub-words (such as CG com-

<sup>6</sup>The image recognition model has fewer parameters than the LMs, but it attains  $>99\%$  accuracy on its original task, suggesting that it does not suffer from being a smaller model.

ponents), so any compositionality in these models must exclusively reflect distributional properties. From these results we conclude that there is legitimate evidence for some CGs having semantically compositional meanings in PE.

### 3.2 Pairing Consistency

To assess the contribution of a sign to the CGs it occurs in, we consider the pairing consistency score (PCS) from Fournier et al. (2020). This metric measures whether the offsets between pairs of words are more parallel than expected by chance. If a sign  $s$  always contributes the same meaning to the CGs in which it occurs, then the offset between the pair of signs  $(t, |t + s|)$  is expected to be roughly parallel to the offset between the pair  $(u, |u + s|)$  for most choices of  $t$  and  $u$ . If CGs containing  $s$  have idiomatic meanings (so the contribution of  $s$  is not consistent), the offsets between such pairs are not likely to be parallel. Thus PCS serves as a proxy for compositionality, and allows us to investigate the impact of individual signs on the representations of CGs in which they occur. This is distinct from a measure like mutual information which depends on raw sign counts and does not account for the internal structure of sign embeddings.

For each sign  $s$  we construct two relations.  $R_{s,in}$  contains all CGs with  $s$  as the inner sign, paired with whichever sign forms the outer part of the CG.  $R_{s,out}$  contains all CGs with  $s$  as the outer sign, paired with whichever sign forms the inner part of the CG. Formally, given a CG  $c$  containing a sign  $s$ , let  $\delta(c, s)$  denote the element of  $c$  which is not  $s$ . Further, let  $I(s)$  be the set of all CGs with  $s$  as the inner element and  $O(s)$  be the set of all CGs with  $s$  as the outer element. Then

$$R_{s,in} = \{(\delta(c, s), c) \mid c \in I(s)\}$$

$$R_{s,out} = \{(\delta(c, s), c) \mid c \in O(s)\}$$

Table 4 reports the average PCS<sup>7</sup> of  $R_{s,in}$  and  $R_{s,out}$  for each model, averaged across all signs  $s$ . On average, we find that inner signs have higher PCS than outer signs. This difference is statistically significant in the image-aware LMs, the image recognition model, and FastText. This implies that inner signs have a more consistent and predictable impact on the representation of compounds in which they occur. The fact that this holds for

<sup>7</sup>We compute PCS using the code published by Fournier et al. (2020), but we adjust the permutation-finding function to avoid infinite loops when a relation contains few items.



some text-only models as well as for the image-aware LMs implies that it is due to distributional properties of signs and not simply their appearance.

model	Mean PCS	
	$R_{s,in}$	$R_{s,out}$
glove.64	0.542	0.544
word2vec.cbow.64	0.525	0.492
word2vec.skip.64	0.521	0.495
fasttext.cbow.64	<b>0.562</b>	<b>0.484</b>
fasttext.skip.64	<b>0.539</b>	<b>0.500</b>
lm.text.64	0.465	0.529
lm.image+text.64	<b>0.719</b>	<b>0.482</b>
lm.image.64	<b>0.760</b>	<b>0.536</b>
image recognition.64	<b>0.929</b>	<b>0.493</b>

Table 4: Comparison of pairing consistency for the inner and outer parts of compound signs in 64-dimensional models. Bolded rows represent pairs where the difference between columns is significant.

Fournier et al. (2020) note that different categories of relations in English have different average PCS. They find that relations involving inflectional morphology (for example, between a verb and its gerund) have high PCS, relations involving derivational morphology (as between *heat* and *reheat*) have lower PCS, and other semantic relations (as between *hot* and *cold*) have the lowest PCS of the relations they examine.

We expect that absolute PCS values will not be comparable between PE and English, owing to the very different nature of the two writing systems. However, it may be possible to draw broad comparisons between different categories. As the category with the highest PCS, inner signs appear to pattern with inflectional morphology, while outer signs pattern more closely with regular lexical items. This does not imply that inner signs actually encode inflectional morphology: most PE signs likely correspond to objects or ideograms, and most types of morphological marking were absent in the earliest phases of Near Eastern writing (Nissen et al., 1993). Rather, we suggest that inner signs may offer minor refinements to the meaning of an outer sign without fundamentally changing its value, parallel to the way that inflecting a verb refines its role in a sentence but does not change its basic meaning.

### 3.3 Analogy

Our PCS results measure sign behaviour in aggregate, but do not provide specific examples of relations between signs. We augment these results by searching for concrete analogies which hold in the embedding models.

Given two CGs  $s$  and  $t$ , let  $s - t$  denote the signs that are in  $s$  but not  $t$ , and let  $s \cap t$  denote the signs both CGs have in common. Consider the vector

$$A(s, t) = e_s - \sum_{u \in s-t} e_u + \sum_{v \in t-s} e_v$$

This vector represents the analogical formula  $s : (s - t) :: t : (t - s)$ . If  $A(s, t) \approx e_t$  in a particular embedding model, then this analogy appears to hold true according to that model.

We compute how often  $A(s, t)$  is within the  $k$  nearest neighbors of  $e_t$  for different thresholds  $k$  when  $s \cap t \neq \emptyset$ . We also compute how often  $A(s, t)$  is close to  $e_t$  when  $s$  and  $t$  are randomly chosen CGs. We predict that CGs which have signs in common also have some meaning in common, and consequently that the former value will be significantly larger than the latter value.

Table 5 shows the results of this evaluation. As in the compositionality task, more analogies hold between CGs with shared components in image-aware models than in text-only models, and the largest number by far occur in the image LM. Once again, in `lm.image.64` the target vector averages  $>0.97$  similarity to the computed vector even when  $k = 15$ . Bold numbers in the table represent cases where analogies are significantly more likely to hold between CGs with shared components than between random pairs of CGs. We see that the number of analogies is larger than expected by chance even in some text-only models, suggesting that there is a meaningful relationship between some CGs which have elements in common. The fact that the image LM outperforms the image recognition model further implies that these analogies reflect legitimate distributional properties and are not purely due to visual resemblance.

model	$k$				
	1	3	5	10	15
glove.256	0	8	11	25	48
word2vec.cbow.256	0	17	36	<b>65</b>	<b>90</b>
word2vec.skip.128	0	8	29	<b>97</b>	<b>140</b>
fasttext.cbow.128	0	9	22	<b>64</b>	<b>98</b>
fasttext.skip.256	0	11	30	91	<b>145</b>
lm.text.64	0	2	7	16	21
lm.image+text.64	<b>27</b>	<b>82</b>	<b>134</b>	<b>233</b>	<b>320</b>
lm.image.64	<b>69</b>	<b>172</b>	<b>258</b>	<b>393</b>	<b>521</b>
image recognition.64	<b>29</b>	<b>67</b>	<b>92</b>	<b>133</b>	<b>174</b>

Table 5: Number of analogies which hold between CGs with signs in common, for different similarity cutoffs  $k$ . Bold numbers represent values which are significantly larger than expected by chance.

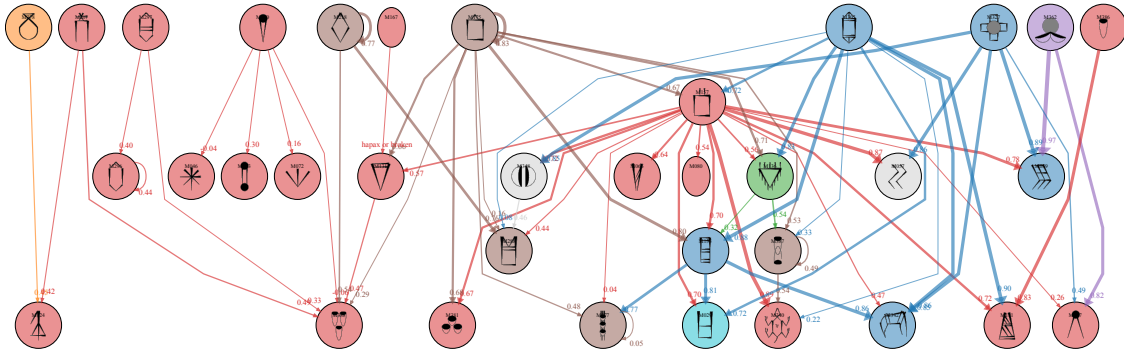


Figure 2: Containment hierarchy for a subset of the signs which can occur in CGs. Directed edges point from outer signs to the inner signs they can contain. Note that (excluding self-loops) the graph is acyclic and all edges point from higher nodes to lower ones. Thicker edges represent CGs which are more strongly compositional. Nodes are colored according to modularity class (Blondel et al., 2008) such that nodes are most strongly connected to like-colored nodes. Full hierarchy, showing all signs which occur in CGs, is available in supplemental material.

As was the case for additive compositionality, the image+text LM underperforms the image-only LM, and on this task the difference is much more pronounced. This suggests that sign names act as distractors: if sign names conveyed information which was helpful to the analogy task, their inclusion would be expected to improve performance. This fact has implications about the labeling of the data which we return to in Section 4.

Taken altogether, the results suggest that many CGs have compositional meanings which can be understood by comparison to the meanings of their component parts and the other CGs with which they share components. We next consider which pairs of signs are able to combine into CGs and which pairings are never observed.

### 3.4 CG Containment Rules

Some signs which occur as the inner part of one CG may also occur as the outer part of another, as with M348 (Ⓜ) in ⓂM327+M348 (Ⓜ) and ⓂM348+M004 (Ⓜ). We may therefore expect to find pairs of signs where either one can contain the other, and yet, no such pairs actually exist. In fact, we find that CGs appear to be constructed according to a strict hierarchy whereby a sign may only contain itself or another sign which is lower on this hierarchy. We can visualize this as a lattice with directed edges from outer signs to the inner signs they are observed to contain, as in Figure 2 (excerpted from the full hierarchy available in the supplemental material). The thickness of an edge in this figure is proportional to the compositionality of the corresponding CG in `lm.image+text.64`.

There appears to be some relation between a sign’s compositionality and its position in this lattice. The signs on the left half of Figure 2 have low compositionality (seen as thinner edges in the figure) while the nodes to the right have higher compositionality (seen here as thicker edges). This suggests that there may be different kinds of CG, of which some are idiomatic and some are not, and that these categories have sufficiently little overlap to appear as separate modules in the lattice.

This “grammar” governing CG construction has not been noted in previous PE scholarship. The ordering of signs within this hierarchy deserves attention in future work, as it may reflect different levels of administrative units in PE society, degrees of specificity in qualifying commodities, or other information which can be exploited to understand the content of these texts.

## 4 Analysis

Little is known about the role of CGs in PE, although these signs make up a significant portion of the corpus. Some occur in “headers” appearing at the beginning of a text. In headers, outer signs (such as M157) are hypothesized to indicate the type of household or institution to which the entire account relates. The outer sign may be further specified by an inner sign, but many (including M157) can also appear without another sign inscribed within. Inner signs are hypothesized to specify a particular kind of item being recorded, a person, profession, or administrative department related to an account, and more.

Our results are consistent with these hypothe-

ses. The PCS results point to inner signs playing a specializing role; this is corroborated by visual inspection of the embedding space, which reveals that CGs cluster according to their outer sign rather than their inner sign (cf. Figure 3 below).

According to Table 2, our text-only models detect additive composition in at most one of every 10 CGs; the image LM detects it in one of every 4 CGs. Likewise, the image LM suggests that a meaningful analogical relation obtains between slightly less than one-third of all pairs of CGs with signs in common. These values depend on the threshold  $k$ , but they suggest the presence of a least a small core of compositional CGs in PE. In several places, compositional and non-compositional CGs appear separated from one-another in the CG containment hierarchy (cf. Figure 2), which may point to this being a legitimate distinction in the writing system and not a failure of our models to detect compositionality in some cases where it is really present.

We can make some inferences about the CGs which are compositional. They are not likely to represent either combinations of ideograms with an emergent lexical value (like the Sumerian cuneiform sign for *naη* “drink” combining the signs for human head and water) or ideograms with phonetic complements (signs indicating the proper reading of the CG), as both cases should be expected to produce non-compositional meanings. Our results may also counter-indicate “coat-of-arms”-like symbols (Farmer et al., 2004), since we show that the components of CGs can often be understood in relation to their use elsewhere in texts, and since CG elements on their own often seem to reference products (including foodstuffs and livestock) and their distribution. Future work may train embedding models on proto-cuneiform, a structurally-similar writing system containing compound signs with occasionally known meanings that could act as useful points of comparison.

The two components of a CG can occur independently, within the same text or even side-by-side. A dramatic example comes from  $\text{IM218+M288}$   $\diamond$ , the components of which appear 37 times as the bigram  $\text{M218 M288}$ .  $\text{M288}$   $\text{𒀭}$  (“grain container”) is the most frequent sign in PE, appearing in diverse contexts but often before numerical measures of capacity.  $\text{M218}$   $\diamond$  is among the signs speculated to function “syllabically” to write personal names (Dahl, 2019), though it may also have other uses. It is not clear yet whether  $\text{IM218+M288}$

and  $\text{M218 M288}$  operate identically, particularly since  $\text{IM218+M288}$  is not strongly additively compositional in any of our embedding models. The possible polyvalence of  $\text{M218}$  and broad distribution of  $\text{M288}$  may impact models’ ability to detect compositionality in  $\text{IM218+M288}$ . Despite this difficulty, the image LM identifies analogies between  $\text{IM218+M288}$ ,  $\text{IM175+M288}$ , and  $\text{IM305+M288}$  (the analogy vector has  $>0.99$  cosine similarity to the target in both cases) implying that we should at least consider  $\text{M218}$ ,  $\text{M175}$ , and  $\text{M305}$  as parallel categories each with relation to grain capacities.

Some signs rarely occur outside of CGs, such as the productive inner sign  $\text{M342}$   $\text{𒀭}$ , about which practically nothing is known. Our data show that it has moderately high PCS (0.69 in `lm.image.64`) and that analogies hold between all but one of the CGs which contain  $\text{M342}$  ( $\text{IM157+M342}$ ,  $\text{IM304+M342}$ ,  $\text{IM305+M342}$ ,  $\text{IM325+M342}$ ,  $\text{IM327+M342}$ , and  $\text{IM351+M342}$ , excluding  $\text{IM153+M342}$ ). These analogies hold strongly for the image LM but not the image recognition model, meaning they reflect primarily distributional properties. Many of these signs are also additively compositional. We believe that these signs may be suitable starting points for future analysis, as our results imply that they are probably not idiomatic and are likely to have related meanings.

$\text{IM157+M342}$	: M157	::	$\text{IM304+M342}$	: M304
$\text{IM157+M377+M377}$	: M157	::	$\text{IM175+M377+M377}$	: M175
$\text{IM370+M046+M370}$	: M046	::	$\text{IM370+M072+M370}$	: M072
$\text{IM175+M377+M377}$	: M175	::	$\text{IM201+M377+M377}$	: M201
$\text{IM351+1(N14)}$	: 1(N14)	::	$\text{IM351+M380}$	: M380
$\text{IM036+1(N39C)}$	: 1(N39C)	::	$\text{IM036+M035}$	: M035
$\text{IM136+M365}$	: M136	::	$\text{IM327+M365}$	: M327
$\text{IM157+M057}$	: M157	::	$\text{IM327+M057}$	: M327

Table 6: Sample of analogies which hold in the `lm.image+text.64` model.

Table 6 gives additional examples of analogies which hold in `lm.image+text.64`. We see that inner and outer signs both participate in analogical relations, as do both  $\text{IS1+S2}$ -type CGs and  $\text{IS1+S2+S1}$ -type CGs. Some analogies hold between a CG with a numeric inner sign and one with a non-numeric inner sign, as between  $\text{IM036+1(N39C)}$  and  $\text{IM036+M035}$ . Such cases may have implications to the meaning of the signs involved; if  $\text{1(N39C)}$  and  $\text{M035}$  truly have parallel functions in these two CGs, this may imply a kind of quantifying role for  $\text{M035}$ , or alternatively that  $\text{1(N39C)}$  is used for its pronunciation or possible syllabic value rather than as a true numeral. The

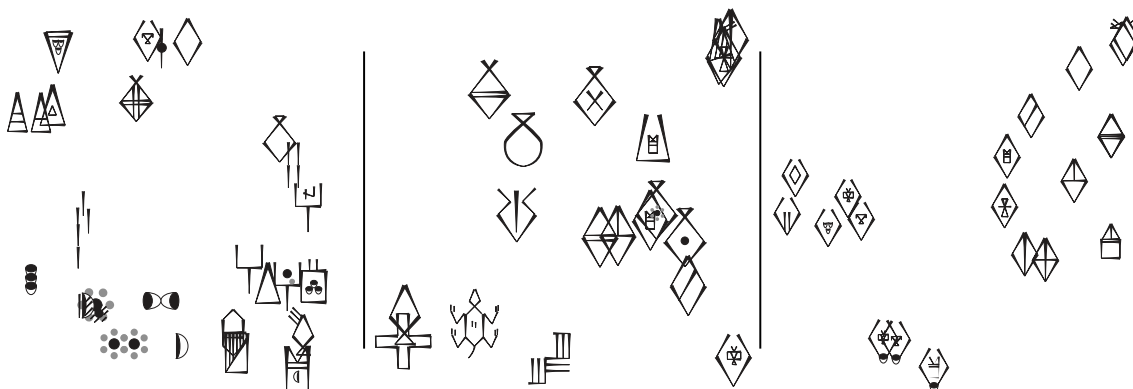


Figure 3: Detail from t-SNE decompositions of the GloVe embeddings (left), the image LM (centre) and the image recognition model (right).

existence of other M036 compounds containing numerals (e.g. |M036+1(N30D)| and |M036+1(N14)|) would seem to favor the former interpretation.

The image-only LM found stronger signals for compositionality and analogical relations than the image+text LM, suggesting that sign names acted as distractors for those tasks. This has significant implications for the ongoing process of revising the PE sign list. Our work relies on the sign labels assigned through an exhaustive manual transliteration process; since it is easy to automate merging signs, this process assumed that most signs are unique until proven otherwise. However, we now believe this choice weakens signals in the text data by making most signs very rare. Moreover, some signs which appear graphically compositional are not currently labeled as CGs, usually when the inner part is never attested as a standalone sign. For these reasons, future work may benefit from relabeling signs based on a combination of context and sign shape.








At the same time, the current transliteration system may record meaningful (if fine-grained) information reflected in minor graphical details (consider M263  and M262 ) such as (hypothetically) “jug of red beer” versus “jug of dark beer”. Such similarly functioning signs might obtain similar embeddings, but retaining their distinction in the published transliterations still improves our understanding of the texts. However for both manual and machine-learning analysis, significant reductions in the sign list may open new avenues for decipherment: for instance, [Born et al. \(2019\)](#) note that frequency-based approaches to decipherment are currently difficult in PE owing to the very small number of repeated  $n$ -grams in the corpus.

Figure 3 shows details from the embedding spaces learned by GloVe, the image LM, and the

image recognition model.<sup>8</sup> GloVe produces small clusters of visually similar signs even though it does not have access to sign images: note the proximity of M353 , M354 , and 2(N30C) , as well as the variants of M036 . These clusters occur in sufficient number that we have confidence the model is detecting meaningful similarities in the usage of visually similar signs. The image recognition model produces much clearer groupings of visually related signs, as would be expected. The image LM replicates some clusters from the image recognition model: a cluster of lozenge-shaped signs is visible in both the image LM figure and the image recognition figure. However, contextual information causes the image LM to relocate other lozenge-shaped signs like M218  to a different part of the embedding space, implying a functional difference between it and the signs in the figure. Overall, these observations confirm that our multimodal architecture is finding a balance between contextual and visual information as intended.

## 5 Related Work

[Sun et al. \(2019\)](#) introduce “character-enhanced” embeddings of Chinese words. Their architecture roughly parallels our own, but requires a deeper CNN due to the visual complexity of Chinese characters. We train with a full context language modeling objective whereas they use a sampling scheme similar to word2vec. They use character-level information to improve word embeddings, where we exclusively learn character embeddings. Our application of this architecture to decipherment is novel.

[Liu et al. \(2017\)](#) explicitly learn compositional embeddings for Chinese characters. They use su-

<sup>8</sup>Full figures are available in our supplemental data.



pervised data to help identify when two visually-distinct signs use the same radical (as in 水 and 池). In our data, it is not known which signs are truly related to one another, thus we refrain from giving the model explicit information about compositionality.

Yin et al. (2019) segment and transcribe undeciphered scripts based on visual similarities between glyphs. Although their transcription error rate is high, they still achieve partial decipherments with no human intervention.

Dencker et al. (2020) perform OCR-style sign detection on images of Sumerian cuneiform tablets, recognizing signs which may be written very differently across the corpus. Their task benefits from the existence of supervised Sumerian training data.

Born et al. (2019) train topic models on PE texts and cluster PE signs in a simple mutual information-based embedding model. The present work considers more sophisticated embedding models and performs a more detailed investigation of the embedding space.

Luo et al. (2019) perform automated decipherment of Ugaritic. Their technique finds alignments between orthographic representations of phonetic information, and thus is not easily applicable to ideographic scripts. It also requires multilingual data, and cannot extract information from a script with no known surviving relatives.

Our work exploits the embedding space learned by a neural language model, but the actual task of language modeling is otherwise irrelevant to our results. By contrast, Kambhatla et al. (2018) actually sample text from a neural language model to help estimate the quality of a proposed decipherment. Future work could similarly sample from a language model as a means of counteracting the small size of the PE corpus; this should be done with caution, however, given the difficulty of evaluating whether the sampled text is fluent.

Salehi et al. (2015) and Cordeiro et al. (2016) demonstrate that English word embeddings tend to be additively compositional and can capture human intuitions about semantic compositionality. Har-tung et al. (2017) investigate other methods for decomposing word embeddings.

Sproat (2006) discusses a variety of writing systems and the degrees to which they employ phonetic versus semantic information. The discussion is largely taxonomic and addresses subtle nuances between scripts which are already well-understood. In this way it demonstrates the wide range of varia-

tion observed between scripts, and by extension the range of possibilities which should be considered when analyzing an undeciphered script such as PE.

## 6 Conclusion

Interpreting what a word embedding model has learned typically involves a comparison to native speaker intuitions. In contrast, in this work we have shown how exploiting graphical compositionality and carefully examining sequences of image embeddings can lead to new insights in proto-Elamite (PE), an undeciphered script with no living users and relatively little available data. Abstracting away from human annotations, we introduced a novel architecture for multimodal or image-based language modeling, which shares information between visually similar signs to better model contextual patterns. This provides a new toolkit for decipherment of an unknown language, distinct from translation-based approaches.

As one of the world's earliest experiments in writing, employing 774 signs and variants by current estimates, reasonable concerns have existed over PE's level of standardisation and the impact this may have on decipherment (Dahl, 2019:71, 82). The corpus is small and filled with lacunae, and prior work has done little to understand how NLP techniques function on early writing systems which may reflect linguistic content differently from modern writing systems. Despite these challenges, this work has shown that embedding models can indeed identify meaningful patterns in proto-Elamite.

We have presented evidence that a subset of complex graphemes are semantically compositional rather than idiomatic, and we have discovered the existence of a simple grammar or partial ordering which appears to govern the construction of CGs. Our results should give domain experts confidence that the proto-Elamite script contains sufficient regularities to allow for describing its mechanics and potentially understanding the underlying content.

## Acknowledgments

We thank the anonymous reviewers for their comments. The research was partially supported by the Natural Sciences and Engineering Research Council of Canada grants NSERC RGPIN-2018-06437 and RGPAS-2018-522574 and a Department of National Defence (DND) and NSERC grant DGDND-2018-00025 to the fourth author.

## References

- Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. 2008. [Fast unfolding of communities in large networks](#). *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10):P10008.
- Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov. 2017. Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics*, 5:135–146.
- Logan Born, Kate Kelley, Nishant Kambhatla, Carolyn Chen, and Anoop Sarkar. 2019. [Sign clustering and topic extraction in Proto-Elamite](#). In *Proceedings of the 3rd Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*, pages 122–132, Minneapolis, USA. Association for Computational Linguistics.
- Silvio Cordeiro, Carlos Ramisch, Marco Idiart, and Aline Villavicencio. 2016. [Predicting the compositionality of nominal compounds: Giving word embeddings a hard time](#). In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1986–1997, Berlin, Germany. Association for Computational Linguistics.
- Jacob L. Dahl. 2005. Complex graphemes in proto-elamite. *Cuneiform digital library journal*, 4(3).
- Jacob L. Dahl. [Proto-elamite sign list](#) [online]. 2006.
- Jacob L. Dahl. 2019. Tablettes et fragments proto-élamites / Proto-Elamite tablets and fragments. *Textes Cuneiform Tomes XXXII Musée de Louvre*.
- Jacob L. Dahl, Laura Hawkins, and Kate Kelley. 2018. [Labor administration in proto-Elamite Iran](#). In Agnès Garcia-Ventura, editor, *What’s in a Name? Terminology related to the Work Force and Job Categories in the Ancient Near East*, pages 15–44. Alt Orient und Altes Testament 440. Ugarit Verlag: Münster.
- Jacob L. Dahl, Cameron A. Petrie, and Daniel T. Potts. 2013. Chronological parameters of the earliest writing system in iran. In Cameron A. Petrie, editor, *Ancient Iran and Its Neighbours*. Oxbow Books, Oxford, UK.
- Tobias Dencker, Pablo Klinkisch, Stefan M. Maul, and Björn Ommer. 2020. [Deep learning of cuneiform sign detection with weak supervision using transliteration alignment](#). *PLOS One*.
- Robert K. Englund. 2004. [The state of decipherment of proto-Elamite](#). *The First Writing: Script Invention as History and Process*, pages 100–149.
- Steve Farmer, Richard Sproat, and Michael Witzel. 2004. [The collapse of the indus-script thesis: The myth of a literate harappan civilization](#). *Electronic Journal of Vedic Studies*, 11(2).
- Louis Fournier, Emmanuel Dupoux, and Ewan Dunbar. 2020. [Analogies minus analogy test: measuring regularities in word embeddings](#). In *Proceedings of the 24th Conference on Computational Natural Language Learning*, pages 365–375, Online. Association for Computational Linguistics.
- Matthias Hartung, Fabian Kaupmann, Soufian Jebbara, and Philipp Cimiano. 2017. [Learning compositionality functions on word embeddings for modelling attribute meaning in adjective-noun phrases](#). In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, pages 54–64, Valencia, Spain. Association for Computational Linguistics.
- Laura F. Hawkins. 2015. [A new edition of the Proto-Elamite text MDP 17, 112](#). *Cuneiform Digital Library Journal*, 1.
- Sepp Hochreiter and Jürgen Schmidhuber. 1997. [Long short-term memory](#). *Neural Comput.*, 9(8):1735–1780.
- Nishant Kambhatla, Anahita Mansouri Bigvand, and Anoop Sarkar. 2018. [Decipherment of substitution ciphers with neural language models](#). In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 869–874, Brussels, Belgium. Association for Computational Linguistics.
- Frederick Liu, Han Lu, Chieh Lo, and Graham Neubig. 2017. [Learning character-level compositionality with visual features](#). In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2059–2068, Vancouver, Canada. Association for Computational Linguistics.
- Jiaming Luo, Yuan Cao, and Regina Barzilay. 2019. [Neural decipherment via minimum-cost flow: From Ugaritic to Linear B](#). In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 3146–3155, Florence, Italy. Association for Computational Linguistics.
- Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013a. [Efficient estimation of word representations in vector space](#). In *International Conference on Learning Representations (ICLR)*.
- Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013b. Distributed representations of words and phrases and their compositionality. In *Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2, NIPS’13*, page 3111–3119, Red Hook, NY, USA. Curran Associates Inc.
- H.J. Nissen, P. Damerow, and R.K. Englund. 1993. *Archaeic Bookkeeping: Early Writing and Techniques of Economic Administration in the Ancient Near East*. University of Chicago Press.

- Jeffrey Pennington, Richard Socher, and Christopher D. Manning. 2014. [Glove: Global vectors for word representation](#). In *Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543.
- Bahar Salehi, Paul Cook, and Timothy Baldwin. 2015. [A word embedding approach to predicting the compositionality of multiword expressions](#). In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 977–983, Denver, Colorado. Association for Computational Linguistics.
- M. Schuster and K.K. Paliwal. 1997. [Bidirectional recurrent neural networks](#). *Trans. Sig. Proc.*, 45(11):2673–2681.
- Richard Sproat. 2006. *A Computational Theory of Writing Systems (Studies in Natural Language Processing)*. Cambridge University Press, USA.
- Chi Sun, Xipeng Qiu, and Xuanjing Huang. 2019. [VCWE: visual character-enhanced word embeddings](#). In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers)*, pages 2710–2719. Association for Computational Linguistics.
- Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. [Attention is all you need](#). In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*, pages 5998–6008.
- K. Wagensooner. 2015. Vessels and other containers for the storage of food according to the early lexical record. *Origini Preistoria e protostoria delle civiltà antiche*, XXXVII(1).
- Xusen Yin, Nada Aldarrab, Beáta Megyesi, and Kevin Knight. 2019. [Decipherment of historical manuscript images](#). In *2019 International Conference on Document Analysis and Recognition, IC-DAR 2019, Sydney, Australia, September 20-25, 2019*, pages 78–85. IEEE.