THE ROLE OF GLYPH BLOCKS IN PUNCTUATING MAYAN TEXTS

Andrew Tate Knudsen

A thesis submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Master's of Arts in the Linguistics Department in the Graduate School.

Chapel Hill
2023

Approved by:

David Mora-Marín

Joseph Lam

Matthew Looper

Katya Pertsova

Jennifer Smith

ABSTRACT

Andrew Tate Knudsen: The Role of Glyph Blocks in Punctuating Mayan Texts
(Under the direction of David Mora-Marín)

This thesis investigates how glyph blocks organize Mayan texts. Understanding the factors that influence a glyph block's composition and its syntactic correspondence helps reveal the degree of metalinguistic awareness of Mayan scribes. To address these points, I sampled 60 texts from the Mayan Hieroglyphic Database, compiled by Matthew Looper and Martha Macri (2023). Each glyph block from each text was tagged for an array of variables, including syntactic value and graphic composition. Linear and logistic regressions revealed that scribes possessed a sound understanding of phrasal structure, evident in their utilization of graphic compacting to fit syntactic units within a single block. Additionally, the phrasal value of glyph blocks varied depending on region, class, and presence of purely phonological spellings, showing that scribal practices differ in their preference for phrasal blocks. In sum, Mayan scribes were keenly aware of syntactic structure, using glyph blocks as a means of punctuating their texts.

To my loving partner and family, thank you for keeping me sane and always having my back, and for all of your support as I worked through my thesis. I love you all.

ACKNOWLEDGEMENTS

First, I would like to thank my thesis advisor, Dr. David Mora-Marín. Prior to taking your historical linguistics class three years ago, I knew very little about Mayan hieroglyphs, or about epigraphy altogether. Through your courses, I developed a passion for the subject, and I am so glad I got to expand my knowledge under your advising. I could not have completed this thesis without your seemingly infinite knowledge of Mayan writing, which proved invaluable in advancing my research.

I would also like to express my gratitude towards Dr. Joseph Lam, Dr. Matthew Looper, Dr. Katya Pertsova, and Dr. Jennifer Smith. Each of you have elevated my thesis to a higher level via your helpful comments as members of my thesis committee. I appreciate your willingness to meet with me as I was grappling with different ideas in preparing for my prospectus defense, as well as your continued guidance and advice through the completion of my thesis. I deeply value your insights and am lucky to have each of you on my committee.

Next, I would like to thank Chris Wiesen at the Odum Institute for his assistance in all things statistics related. Your help with sampling and reviewing my regressions saved me much time and provided me with a deeper understanding of these methods.

TABLE OF CONTENTS

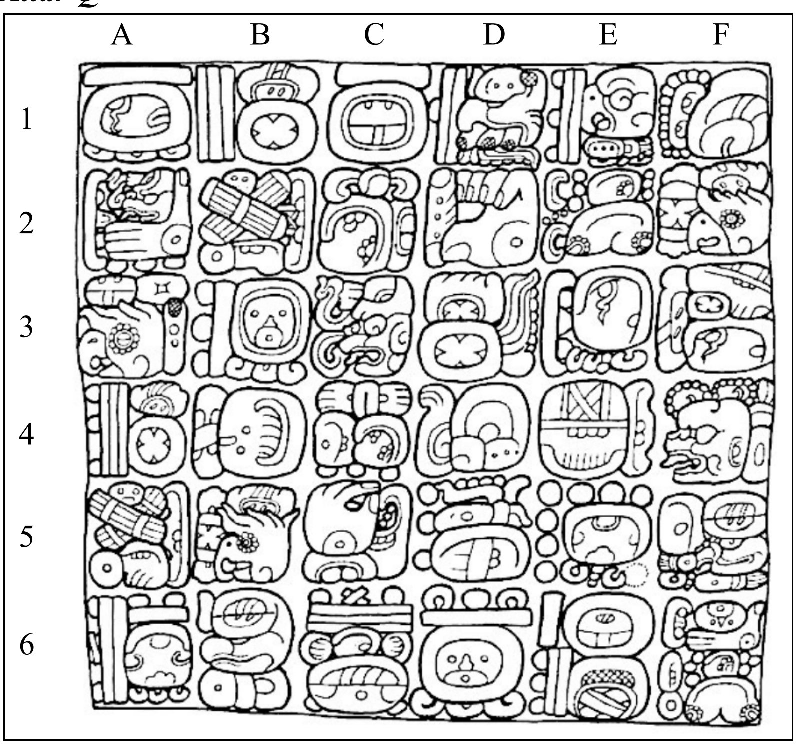
LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

Writing systems vary vastly in how they are organized. Beyond the reading direction of the script, writing systems differ in how their signs are graphically spaced, and what types of signs may be juxtaposed to each other. Languages written in the Roman alphabet use spacing to indicate word boundaries, as well as commas and periods to parse speech into phrasal and prosodic units. In contrast, Hangul, used for writing Korean, conglomerates its featural signs into syllabic units (Rogers 2005:7). Mayan writing typically organizes its script by displaying the language in a grid of square-shaped units called glyph blocks. Figure 1 below depicts Copan Altar Q, a text organized in a six-by-six glyph block grid.

Figure 1. Copan Altar Q



While direction of writing is not always consistent in Mayan inscriptions, Mayan is generally written in a double-columnar fashion. To read Copan Altar Q, this would entail starting with glyph block A1, then reading B1, then A2, then B2, and continuing until reaching the end of columns A and B. Then reading would continue in the same way with columns C and D. Glyph blocks vary in the number and type of signs they possess. Within a glyph block, reading direction goes from left to right, and from top to bottom.

Although every Mayanist is familiar with glyph blocks, very few have attempted to stringently define and characterize this unit of graphical organization. In his dissertation, Justeson describes glyph blocks as a unit of meaning, explaining that it is rare for a single word to be stretched across multiple glyph blocks (Justeson 1978:262). He even goes on to say that "the immediate graphic constituent structure of a glyph block reflects the immediate linguistic constituent structure of the corresponding linguistic sequence" (1978:276). He explains that in addition to parallels of the spoken language, artistic constraints also influence a glyph block's layout. Not only must the glyph block maintain its squarish shape, but the signs within it also must be represented in an aesthetically pleasing way. Law and Stuart provide a similar analysis, explaining that glyph blocks are a method of visual arrangement that serves to delineate words or larger syntactic structures (2017:130). Like Justeson, they find that morphemes are rarely split across glyph blocks, and instead that it is common for multiple words reflecting a syntactic constituent to be grouped within a single block. While these authors confirm common intuitions about the role of glyph blocks, no study has yet described the relationship between glyph blocks and grammatical units via inferential statistics.

This thesis seeks to expand our understanding of what constitutes "punctuation". Traditionally, punctuation is considered in a very strict sense. Specifically, it is often conceptualized as simple, small marks, such as the comma < , > or period < . > used in the Roman script to denote different phrase types. However, the main function of punctuation is to organize written language into prosodic and grammatical units. As such, I posit that punctuation need not have a specific appearance, so long as it performs this function. In Mayan writing, glyph blocks do play a vital role in organizing the written language, and potentially correlate with syntactic units. I argue that glyph blocks are a form of punctuation, which requires a broader definition of the term. Rather than restricting punctuation to types of pointed marks, we must consider the fact that visual arrangements may serve the same function in segmenting written language, thus including glyph blocks within our definition.

By uncovering how glyph blocks are composed, we will gain a better understanding of how Mayan scribes conceptualize language. If glyph blocks do systematically correspond to syntactic units, this will provide evidence that scribes had a developed metalinguistic knowledge of their language, and that they chose to write in a fashion that reflects their language's syntactic structure. Metalinguistic awareness has been discussed by Houston (2011) on the phonological level. He states that the **bu** syllabogram is derived from the **mu** syllabogram by adding two circular elements within the sign, such that the similarity in graphic representation reflects the phonological similarity between [b] and [m] (Houston 2011:32). While there is no direct evidence of metalinguistic awareness from the Classic Period in terms of syntactic structure, such awareness can be interpreted by analyzing representational choices in the script. For example, if scribes consistently write full prepositional phrases within a single glyph block, we can infer that they conceptualize prepositional phrases as a unit that cannot be broken in the writing. I expect that glyph blocks will correspond to syntactic phrases (with some expected variability), implying that scribes were cognizant of this constituency.

Discoveries about the precise function of glyph blocks in Mayan writing would open up avenues for analyzing other writing systems in a similar light. Egyptian hieroglyphs, Cuneiform, Japanese, and Chinese are all logographic systems like Mayan, and it is possible that they display interesting parallels regarding visual arrangement.

Since the defense of this research, this thesis has been revised to consider comments from my committee members. The thesis is now organized into chapters, and a final conclusions chapter has been added at the end. Additionally, the hypotheses have been reorganized so that they are thematically related, and such reorganization has also been accounted for in results and discussion chapters to reflect this change.

CHAPTER 2: BACKGROUND

Regardless of how we define the term “punctuation”, and whether visual arrangements are included within this definition, it is clear that other writing systems use visual cues to segment written language. As speakers (and writers) of English, we know that English words are separated from one another when written by means of a gap at the word boundary. Writing segmentation at the word level is common in many written languages, ranging in representation from spacing to pointed marks or vertical lines. Such written languages include Ancient Egyptian Hieroglyphs (Woodard 2015:7), Ancient South Arabian (Nebes & Stein 2015:455), Etruscan (Rix 2015:946), Luvian (Melchert 2015:577), Old Persian (Schmitt 2015:719), and Ugaritic (Rogers 2005:93). Note that Ancient Egyptian and Cuneiform (used for writing Luvian, Old Persian, and Ugaritic) are logographic scripts like Mayan, suggesting that any writing system, regardless of representational unit, utilize spacing as a device for visual segmentation. Some languages use spacing to separate smaller linguistic units as well. In Chinese writing, each logogram generally corresponds to a single syllable. As such, Chinese writing is separated on the syllabic level (Rogers 2005:26). While Hangul also spaces writing by syllable units, each syllabic block can be further divided into as many as three segmental signs (including a mandatory nucleus and an optional onset and coda).

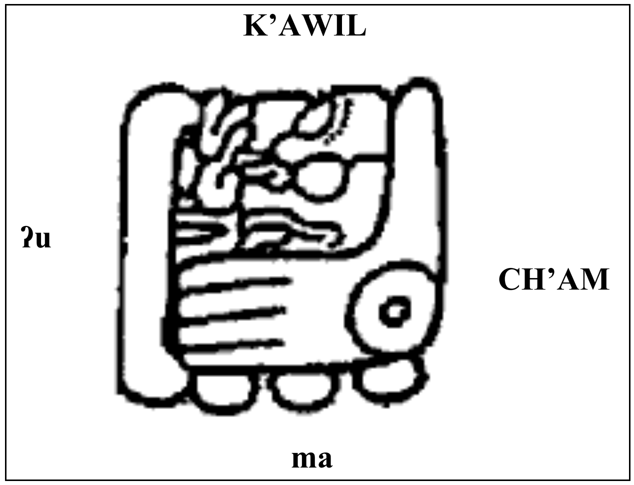
Punctuation differs from text segmentation in that the former is based on linguistic units, whereas the latter is restricted by aesthetic and spatial restrictions. While a text could be segmented in a number of different ways, the use of punctuation is reserved for delineating syntactic and prosodic units. Glyph blocks are inarguably a unit of visual segmentation: rather than writing each sign of a text into a large conglomerative mass, signs are grouped together into multiple discrete units. What remains to be seen is whether glyph blocks also punctuate texts. Visual segmentation and punctuation may overlap in Mayan writing, where a glyph block does correspond to a grammatical phrase or a full word, though it may also be possible that this correlation does not exist, with glyph blocks only containing a fragment of a word or phrase. Regardless, we can state that both punctuation and text segmentation serve to separate texts into visual components, though the units that they create may vary.

As mentioned above, Mayan writing is logosyllabic, comprised of two types of signs. The first are logograms, which are signs associated with a specific semantic value. Mayan also has a phonographic system that works in tandem with the logograms. These signs are syllabograms, representing a CV combination of segments. One aspect of Mayan writing that is absent in many other writing systems is the immense variability in representation. While logograms are available for the majority of roots in Mayan writing, it is possible to write such roots phonographically as well. For example, the Mayan word for 'snake' is *chan*, which could be represented by a single logogram **CHAN**, or it could be represented by the two syllabograms **cha-na**. Mayan roots are generally CVC in shape, and thus it was understood by readers that the final vowel of syllabic spellings should be dropped. Via phonetic complementation, scribes could write the same word using a combination of logograms and syllabograms. For the *chan* example, a scribe could use the logogram **CHAN**, followed by the syllabogram **na**. Phonetic complementation helps readers to discern the pronunciation of a logogram. This is especially important when a sign is polyvalent, meaning that it has multiple possible semantic values. Providing clues to the pronunciation allows for selection of the correct meaning by the reader. Even more variation can result when considering allographic signs. Many syllabic signs can be represented with a variety of different forms, any of which would be appropriate for use. The number of syllbograms, logograms, and phonetic complements within a glyph block tend to vary depending on a number of factors, as seen in the results section below.

One rather obvious factor that dictates how a text will be arranged graphically is the size and shape of the medium used for writing. As a result, scribes would often plan out their text layout in advance to ensure that the piece of writing would fit within the given physical margins. There are plenty of surviving artifacts that show that scribes did run out of room. Grube (2021:3) describes how some pottery vessels containing a Primary Standard Sequence (PSS) sometimes were incomplete. The PSS is a standardized sequence, containing information on the dedication of the vessel, when it was painted or inscribed, what the vessel was meant to hold, and the owner of the vessel (Coe 1973; Boot 2005:1). The PSS would be written around the vessel, so that a complete text would place the final glyph block of the text adjacent to the first glyph block. However, some vessels do not complete the full PSS, since the scribe was not able to fit the entire piece of writing in one circle around the object.

As previously mentioned, texts were generally read in a double columnar fashion (Justeson 1986:453), but this was not always the case. As with pottery vessels bearing the PSS, text could simply be read left-to-right. In rarer instances, right-to-left reading is also found, with any face-shaped signs facing the opposite direction than usual. The reading direction always opposes the direction that the face-shaped signs are pointing towards. In the majority of cases, texts are written so that the signs run parallel to the spoken language, but there are some instances of non-linguistic sign order. For example, to say that someone acceded to rulership, the phrase *ch'amaw k'awil* is used, meaning '(s)he held the scepter'. However, due to the graphic shape of the logogram **CH’AM**, the logogram for **K’AWIL** was often written first, as seen in Figure 2.

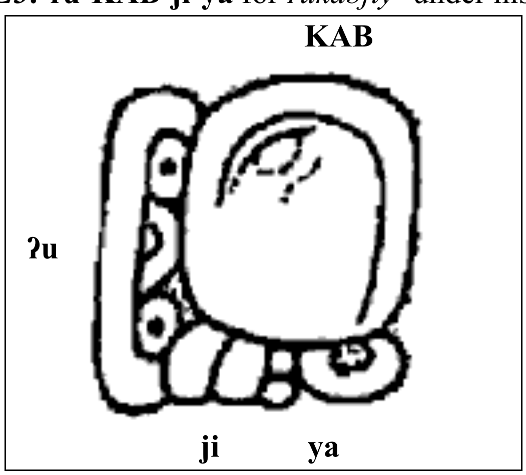
Figure 2. Copan Altar Q:A2: **ʔu-K’AWIL-CH’AM-ma** for *ʔu ch’am k’awil* ‘his taking of k’awil’



This non-linguistic arrangement serves to visually represent the action which it describes, as well as aesthetically improve the layout of the text. Aesthetics were important to Mayan scribes, and sometimes extended to having portions of the text interact with imagery on the object on which it was written. For example, Altar 1 from Mountain Cow depicts the image of a captive who is surrounded by text (Morley 1937:222). The hair of this figure is extended to contact a glyph block meaning ‘he of twenty captives’. This intentional association indicates to the reader that the figure is in fact a captive, and places emphasis on the power of the holder of this title.

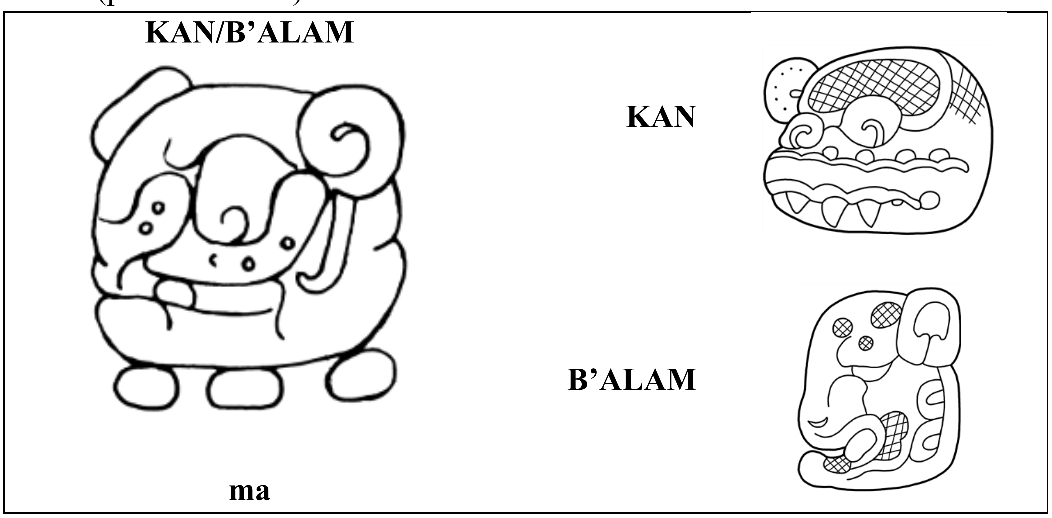
Within a glyph block, signs can be one of two types of visual shapes: a main sign or an elongated sign. Main signs are squarish in shape, and generally occupy a large portion of the glyph block (Justeson 1986:447). Elongated signs, also called graphic affixes, are more rectangular in shape, and are generally derived from more elaborate signs that can be reinterpreted as having multiple components. Depending on context, a sign may appear as a main sign in one glyph block, but as an elongated sign in another. In Figure 3 below, the sign **KAB** is a main sign, whereas the other three are all graphic affixes. While there is only one main sign per glyph block, it is possible to have any number of graphic affixes, given that they fit within the available space. Both logograms and syllabograms may appear as either graphic type. As mentioned previously, there is a wide range of ways in which words can be represented based on whether a spelling is syllabic or logographic, and whether the scribe chooses to phonetically complement certain logograms. This causes variability in how many elongated signs reside within a glyph block, and thus how tightly compacted it is. However, there are certain semantic categories that tend to have more signs per glyph block than others. One such category is emblem glyphs. These glyphs representing titles of a polity were systematically constructed with three parts: the logograms **K'UHUL** and **ʔAJAW**, meaning 'divine' and 'lord', respectively, as well as a third sign generally corresponding to the site-specific name, which was often tied to a location's toponym (Martin 2020:71-72). Thus, all complete emblem glyphs contained at least three signs.

Figure 3. Copan Altar Q:E3: **ʔu-KAB-ji-ya** for *ʔukabjiy* 'under his authority'



While signs can usually be categorized as either a main sign or a graphical affix, the practice of blending multiple signs into a single sign can be found in many Mayan texts. When one sign is completely encompassed by another (and thus only occupies the space of the larger, outer sign), this is called infixation, where the smaller sign would be the infix. Conflation occurs whenever two signs are mixed, where the resulting graph contains elements of both of its component parts. As with graphical infixing, conflation results in a single sign that contains the information of multiple pieces. Figure 4 is taken from a glyph block at Palenque, and displays the process of conflation. The logogram **KAN** 'snake' has been conflated with the logogram **B'ALAM** 'jaguar'. This conflated main sign takes aspects of both **KAN** and **B'ALAM**: The circles around the mouth and the curved element in the eye can be found in **KAN**, while the ear is taken from **B'ALAM**.

Figure 4. Conflation Example. Palenque Temple of Inscriptions:E4: **KAN/B'ALAM-ma,** 'Kaan B'ahlam (personal name)'



While it is vital to understand the different types of components that may construct a glyph block, it is also important to consider the context in which glyph blocks are used. Mayan texts could be written in a number of different discourse genres, some of which adhered to a standardized template so that every text of that genre had the same framework. Name-tagging is one of the shortest discourse genres, and involves only the name of the object on which it is inscribed, or possibly the name of the owner, though no statements about possession are made.

Proprietary texts, on the other hand, require at minimum a possessed and a possessor noun, indicating what the object is and who owns it. Dedicatory texts follow the PSS, mentioned above. These texts were extremely formulaic, such that any text bearing the PSS were nearly identical, differing only in information specific to that vessel (including what the vessel was, who possessed it, and the date of dedication). Political texts are another common discourse genre, often written on stone monuments. These narratives dictate important information about rulers, including their date of birth and date of accession to rulership. Additionally, they may describe military successes or important relationships and alliances with other nobles. The majority of political narratives begin with an Initial Series Introductory Glyph (ISIG), which generally occupies the space of four glyph blocks. This is then followed by the date, as well as any additional calendrical data, such as the current position within the lunar sequence. Through the text, new events will be introduced using an event indicator glyph and the number of days since the previous mentioned event. Ritual and cosmological texts were also popular, often including specific dates regarding the occurrences of certain celestial events, though such subject matter was somewhat sparse during the Classic Period. It should be noted that political narratives and ritual texts often overlapped, and were not necessarily distinct entities (Martin 2020:54). The fact that many of these discourse genres are very formulaic in their composition proves that standardization of writing was important, to some extent, to Mayan scribes. It would be fruitful to question whether such standardization is purely graphic or discourse-based, or if it reflects a deeper, metalinguistic knowledge of the spoken language.

Understanding the grammar of Mayan languages is essential to this study. Note that while there are approximately 31 Mayan languages today (two of which have become extinct), Mayan hieroglyphic texts were only used for three branches of the Mayan family tree at the time of writing: Yucatec, Ch'olan, and Tzeltalan. However, it is worth noting that evidence for texts written in Tzeltalan is not strong (Mora Marín, personal communication, 2023). Scribes who were native speakers of another Mayan language would have been required to learn to write in a second language, requiring them to be diglossic. Yucatec, Ch'olan, and Tzeltalan are all VOS in structure, and thus the written language follows this pattern (Bricker 2004:1065). In terms of morphosyntactic alignment, Mayan utilizes an ergative-absolutive system, which marks intransitive subjects and transitive objects in the same fashion, to the exclusion of the transitive subject (Maxwell 2006:707).

Mayan languages are also highly agglutinative, allowing for affixation of many possible regular morphemes onto verb and noun roots (Polian 2017:201). Transitive roots are prefixed with an ergative clitic, known as Set A pronouns, and may also be preceded by an aspect marker. The root is immediately suffixed by a status marker, which indicates the mood, the completive status of the verb, and the transitivity. An absolutive suffix is than added at the end (belonging to the Set B pronouns). Intransitive verbs will follow the same pattern, except that they will lack the Set A pronoun. Nouns, when possessed, will also utilize Set A prefixes along with a possessive suffix, whereas their unpossessed counterparts will not. Note that the noun will often be marked directly with the Set A proclitic, though this marking may be attached to the preceding adjective if there is one (Polian 2017:620; Coon 2017:664; Hofling 2017:719). Nouns may also have adjectival modifiers and determiners preceding the noun root. Additionally, nouns may have deictic suffixes that correspond to the determiner of the noun. Templates for each of these syntactic categories are given in Table 1 below.

Table 1. Verb and Noun Templates

Syntactic Category	**Template**
V (transitive)	(Aspect) - Set A - Transitive Root/Stem - Status - Set B
V (intransitive)	(Aspect) - Transitive Root/Stem - Status - Set B
N	(DET) - (Set A) - (Modifier) - Noun Root/Stem - (POSS) - (Deictic Enclitic)

Provided that a large portion of this thesis investigates the correlation between syntactic phrases and glyph blocks, an examination of the phrasal structure in Mayan languages is warranted. Table 2 provides outlines for maximal extensions of both verb phrases and noun phrases. These templates were derived from those provided by Polian (2017), Coon (2017), and Hofling (2017) for Tzeltal and Tzotzil, Ch'ol, and the Yucatecan languages, respectively. While there is slight variability, verb and noun phrases in each of the aforementioned languages are generally formed via these templates.

Table 2. Verb Phrase and Noun Phrase Templates

Phrase Type	**Template**
VP	(Negative Particle) - (Adverb) - (Aspect) - Verb
NP	(Determiner) - (Numeral with Classifier) - (Attributive Adjective(s)) - Noun - (Possessor NP) - (Relative Clause)

In Tzeltal and the Yucatecan family, aspect may be marked by a free morpheme preceding the verb, though in other scenarios is will appear as a verbal prefix (Polian 2017:622; Hofling 2017:707). In contrast, aspect is always free in Ch'ol, and always affixed to the verb in Tzotzil (Coon 2017:667; Polian 2017:622). It is important to note that Polian, Coon, and Hofling all avoid providing a clear, definitive template for a verb phrase. In fact, most Mayanists will remain neutral in this regard. Given the standard VOS or VSO word order of Mayan languages, in conjunction with the possibility of topicalization, it is unclear how to define a verb phrase. By and large, Mayanists will not comment on whether a transitive verb and its object form a verb phrase or not. A notable exception is in the case of antipassive verbs with incorporated objects. In such constructions, the object must immediately follow the antipassive verb: it may not be moved, and no other elements may be inserted between it and its verb (Mora-Marín 2004:340). Thus, it is generally accepted that object-incorporated antipassives constitute a verb phrase.

In terms of noun phrases, Polian, Coon, and Hofling agree upon the template provided in Table 2. Note that only the noun is required; all other elements are optional. Determiners, numerals, and adjectives will precede nouns, while possessor nouns and relative clauses will follow noun heads. Also, in Ch'ol, it is possible for the relative clause to precede the noun it describes, though this is a much rarer construction.

For the purposes of this study, a "word" consists of a root and any bound morphology, including both affixes and clitics. Compounds are also considered to constitute a single word, even though they possess more than one root. For example, a "phrase" is composed minimally of the phrasal head, and may also include additional words that provide information about the head (such as those seen in Table 2). A word alone constitutes a full syntactic phrase only when there are no other words in the phrase. Referencing the NP template from Table 2, a noun alone would only form a complete noun phrase if it did not have a determiner, numeral, adjective, possessor noun phrase, or relative clause related to it. If a noun does appear with one of these other elements, the noun alone would not be considered a syntactic phrase. Examples of noun and verb phrases in both contemporary and epigraphic Mayan can be found in Table 3.

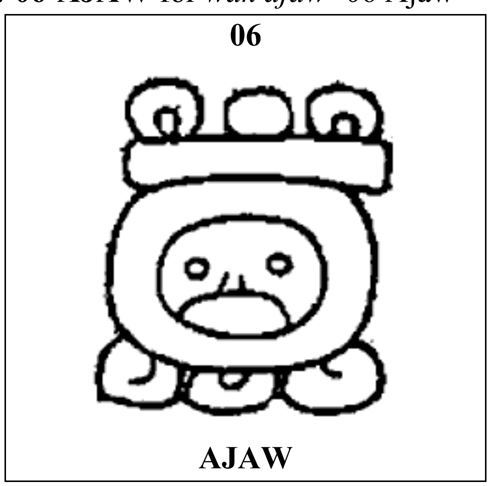
Table 3. NP and VP Examples. Phrase heads are bolded.

Phrase Type	Language	Example
NP	Tseltal	te s-bats'il **s-ton** k-alak'-tik=e. ART B3-authentic A3-egg A1-chicken-PL1=DET 'the authentic chicken eggs' {Polian 2017: 620}
	Ch'ol	ili cha'-p'ej kolem **alaxax** DET two-CLF.round big orange 'these two big oranges' {Coon 2017:664}
	Yucatec	le tumb'en **máaskab'-o'** DET new machete-DET 'that new machete' {Hanks 1984 2:1}
	Epigraphic	lachan **aajaaw** twelve Ajaw '12 Ajaw (Calendar Round date)'
VP	Tsotsil	Muk' **ch-i-bat** NEG IPFV-B1-go 'I am not going.' {Polian 2017:623}
	Ch'ol	Tyi **wäy-i-yoñ** PFV sleep-ITV-B1 'I slept.' {Coon 2017:667}
	Mopan	tan **aw-il-ik-en** DUR A2-see-ITS-B1SG 'you are seeing me' {Hofling 2017:723}
	Epigraphic	**u-chuk-uw-Ø** A3-seize-RTV:STATUS-B3 's/he seizes/d it'

Understanding the syntax of Mayan is essential for determining whether glyph blocks are in fact a form of punctuation. However, some researchers argue that there are examples of punctuation in Mayan writing aside from glyph blocks which fit within the more narrow Western definition of the term. Nikolai Grube (2021:3) sites examples of what he claims are "space fillers" or markers of the end of a text, taking the form of either one or two vertical lines or two vertically arranged dots. These phenomena are found on pottery vessels in which the text loops around the rim of the vessel, so that the end of the text comes right before the beginning of the text. Grube proposes that such "punctuation" at the end of the text indicates finality of what is being said, and thus functions as a Roman orthographic period. However, the question of whether these markings are in fact punctuation or not has been contested. Mora Marín (2022) suggests that such pointed marks may be portions of glyphs that were supposed to follow in the PSS. For example, two dots at the end of a ceramic text may represent the numeral 2, which could feasibly comprise part of someone's name or title. However, these scribes would have run out of space, so only a portion of the name or title in following glyph block would remain.

Mayan writing also utilizes certain signs that act purely as space fillers. Aesthetics were a priority to scribes, and ensuring that the glyph blocks were neatly organized and fit a squarish shape was important. One common use of space fillers was in numerals. To represent the number 6 in Mayan writing, a single circle (representing a value of 1) is placed on top of a single bar (with a value of 5). However, in practice, two place fillers are also inserted. Consider the following glyph block from Copan Altar Q:

Figure 5. Copan Altar Q:D6: **06-AJAW** for *wak ajaw* '06 Ajaw'



Here, we can see that there are two upside-down "U" elements, one on either side of the circle element of the numeral 6. These serve no semantic function, but instead create a visual arrangement that better fits the square shape of a typical glyph block.

Utilization of another type of visual arrangement as a textual tool can be found in the Palenque Tablet of 96 Glyphs. This text summarizes a few accession events at Palenque, with a focus on the ruler Kuk II (Josserand 1986:27). As is typical in political narratives, a Posterior Event Indicator (PEI) is used to introduce a new event. What is unique about this text is how the PEI glyphs are arranged visually: for the first four instances of the sign, each is lower physically than the next. This creates an ascending visual, reaching its climax at the accession of Kuk II, which is the most important event in the text. This text displays a form of discourse segmentation via visual arrangement, showing that scribes do use spatial control to demarcate linguistic units.

CHAPTER 3: HYPOTHESES

Based off of these prior works, I developed the following hypotheses that I investigate in this thesis.

3.1. Hypothesis 1: Glyph Blocks as Syntactic Phrases

First, I suspect that glyph blocks in monumental texts will correspond to syntactic units more systematically than in portable texts. Monumental inscriptions were often constructed by order of a ruler, and thus required an experienced scribe. Such scribes were themselves considered high class citizens due to their craft. Further, monumental inscriptions were often on public display. Since these writings were easily accessible to a wide population, there was a pressure that such inscriptions were very high in quality. In contrast, portable objects tended not to be displayed to a large group of people. While these objects still required a great deal of artistry, there was less influence from rulers in their construction. Furthermore, pottery inscriptions were often left incomplete, with only a part of a text written upon its vessel, supporting the claim of writings on ceramic vessels being less regulated (Mora Marín, personal communication, 2023). If monumental inscription are more systematic in their syntactic representation, then this would suggest that pressures from ruling commissioners and public display require a more structured representation of writing. It would also suggest that these elite scribes had a metalinguistic knowledge of their language, and found it important to display this through their writing.

Second, I hypothesize that glyph blocks will become more systematically represented in later years, whereas earlier texts will be more variable. With time, I suspect that writing would become more standardized, especially during more prolific times between approximately 600-800 CE. Thus, I expect less variability in phrasal representation within glyph block representations in the Late Classic (after 600 CE) than in the Early Classic Period (before 600 CE).

Third, I also suspect that more prolific sites (including Palenque and Copan) will represent their glyph blocks more systematically than less prolific sites (like Uxmal). If the same scribes/scribal schools are producing a large number of texts, then we would expect them to vary less than sites where only a few surviving inscriptions still exist. As a result, texts at these more prolific sites would be more likely to have glyph blocks that correspond to full syntactic phrases.

Fourth, I hypothesize that when words are spelled purely with syllabograms, glyph blocks will be less likely to correspond to syntactic units than when logograms are used. When syllabic spellings are used, especially in heavily grammatically affixed words, many more signs are required. Thus, scribes must face the choice of either tightly compacting blocks, or spreading these words out over multiple blocks. In the latter case, a single block would only correspond to part of the pronunciation of a single word, and thus would not correlate with a syntactic unit.

Further, I believe that increases in the instances of conflation and infixation, as well as instances of multiple glyph blocks in the space of one, will produce more phrasal glyph blocks. Assuming that Mayan scribes were well aware of the syntactic structure of their language, I suspect that they would make efforts to fit phrases within a single glyph block. Conflation, infixation, and compressing blocks are all methods by which a scribe could save space, allowing for longer phrases to fit within a single glyph block.

I further hypothesize that longer texts will produce more glyph blocks that represent syntactic phrases. Longer texts are generally correlated with specific discourse genres, namely political narratives. Political narratives, due to their public presentation on monuments, were widely viewed and provided important histories about a site's rulers. As such, I would suspect systematicity in phrasal representation, as these were prestigious works. In extension, I believe that both conflation and infixation would be more likely to occur in longer texts: by reducing space restrictions, scribes can ensure that these inherently longer texts are able to fit on the writing media.

3.2. Hypothesis 2: Glyph Block Compression

I postulate that glyph blocks will be more compacted near the end of a text. As previously mentioned, scribes would often run out of physical space on which to fit their inscriptions. Thus, in order to complete the entire text, they must write a greater number of signs per glyph block. In extension, I believe glyphs near the end of the text will be more likely to contain inner glyph blocks, graphical affix signs, more instances of conflation and infixation, and fewer phonetic complements. Juxtaposing multiple glyph blocks within the space of one glyph block, as well as avoiding larger signs, will save space. For an example of this glyph block juxtaposition, see Figure 6 below in the methods section. Similarly, conflating or infixing signs will merge two different signs into the space of one. Further, since phonetic complements are not necessary to include, they will likely be omitted in order to preserve space. In conjunction with this hypothesis, I suspect that the reverse is also true: that the beginning of texts will be less dense. Many monumental texts begin with an ISIG, which often occupies the space of two or four glyph blocks, if not more. Thus, these texts will have far fewer signs per glyph block at the start.

Next, I hypothesize that glyph blocks contributing to titles, names, and emblem glyphs will have more phonetic complements. Such semantic categories are often site-specific, and thus may be less familiar to read. Phonetic complementation will assist readers in knowing how to pronounce such words, and by having these phonetic complements, the glyph blocks will be more compacted.

I also postulate that the number of total signs will increase with time. As scribal practices are transmitted and developed, writers will be more familiar with the sign inventory, and different ways to arrange them in a text. Late Classic scribes could review Early Classic writings and note that there were numerous unfinished texts due to space limitations. Such scribes would likely compact their glyph blocks to remedy this issue.

3.3. Hypothesis 3: Preference for Syllabograms

Through background research, it has become apparent that many Mayanists share an intuition that texts written in the northern Mayan region contain more syllabograms than in other regions. I suspect that this is true, given that the Yucatecan speakers of the north may choose to reflect their native pronunciations in a Ch'olan writing system (Wichmann & Davletshin 2006:103).

Lastly, I hypothesize that portable object will be more likely to possess purely syllabic spellings than monuments. The use of syllabic spellings is an innovation in Mayan writing, with a rise in popularity in the Late Classic Period. Provided that portable texts are less formal than monumental texts, I believe this innovation would be more readily accepted in portable objects.

CHAPTER 4: METHODS

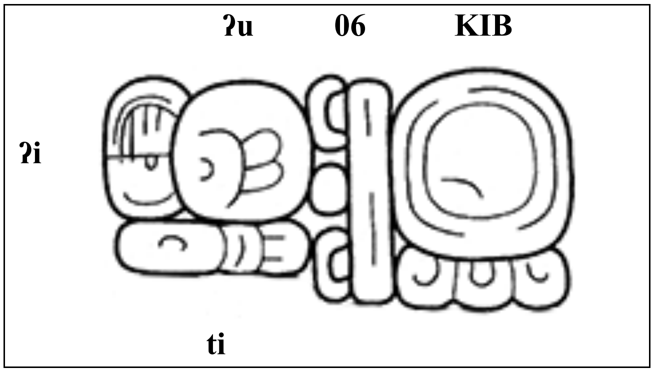
For this analysis, I utilized the Mayan Hieroglyphic Database (MHD), compiled by Matthew Looper and Martha Macri (1991-2023). The MHD is a database that stores information about Mayan texts and the glyph blocks that compose them. The MHD is organized by glyph block, with each entry corresponding to a different block. Each entry provides information about the syllabic/logographic value of the sign, where the inscription was located, the media it was written on, when it was written, a translation, and more. Data from the MHD was uploaded into R Studio for analysis. In total, the MHD online has a total of 70, 217 glyph block records.

While some of the variables present in the MHD were of interest in the present analysis, there were many more variables that needed to be added. The values for these variables were assigned after obtaining a representative sample and analyzing images of the glyph blocks in the MHD. *phrasal_unit* indicates whether a glyph block is comprised of part of a phrase only or a full phrase, following the phrasal requirements dictated by Polian, Coon, and Hofling. Similarly, *word_unit* indicates whether a full word or only a part of a word is found in a glyph block. *grammatical_class* provides the syntactic category of the glyph block, with options including a noun, noun phrase, verb, verb phrase, adjective, and prepositional phrase. To break these units down further, the variable *semantic_category* was added to show possible trends among words with the similar meanings. While the MHD does provide the semantic value for many glyph blocks, the present variable is broader, labeling blocks more generally to indicate names, titles, emblem glyphs, and dates. Also of interest was *text_length*, which is valued with how many glyph blocks make up a specific text. *text_part* indicates how far through a text the glyph block occurs, separated into *beginning*, *middle*, and *end*. *dispersion* indicates the position of a glyph block on a numeric scale. This is given as a ratio, dividing the position of the glyph block by the length of the text. Thus, any value for *dispersion* will be on a scale from 0 to 1, with 0 representing the beginning of the text and 1 representing the end. This variable is colinear with *text_part*, since both indicate position within the text.

#_syllabogram and *#_logogram* are numeric variables which simply count how many syllable signs and logographic signs, respectively, are present within a glyph block. In cases where the value of a sign was undecipherable or not clearly evident, this value was left blank. If a polyvalent sign was used, it was recorded as having the value that was used in its particular context. Similarly, *#_other* counts the number of signs that are neither logographic nor syllabic. While these are rarer, day sign cartouches and the reduplication diacritic would both be counted here. *#_total* adds the number of syllabograms, logograms, and other signs within the glyph block. The variable *only_syllabic* will take the value *syllabic* if the glyph block contains no logograms, and the value *logosyllabic* if it has at least one logogram. *#_main_sign* and *#_affix* are also numeric variables, with *#_main_sign* counting the number of signs which tend to be more squarish in shape and occupy a greater portion of space within the block, and *#_affix* counting the number of elongated signs that graphically affix the main sign(s). Here, I follow Justeson's definititions of 'main signs' and 'graphical affixes'. *conflation/infixation* is another numeric variable that counts the instances of both conflation and infixation within the glyph block. *phonetic_comp* counts the number of syllabic signs used to clarify pronunciation of logograms within a glyph block. *blocks_in_blocks* takes a numeric value corresponding to the number of glyph blocks presiding within a single glyph block, if any. If a single glyph block

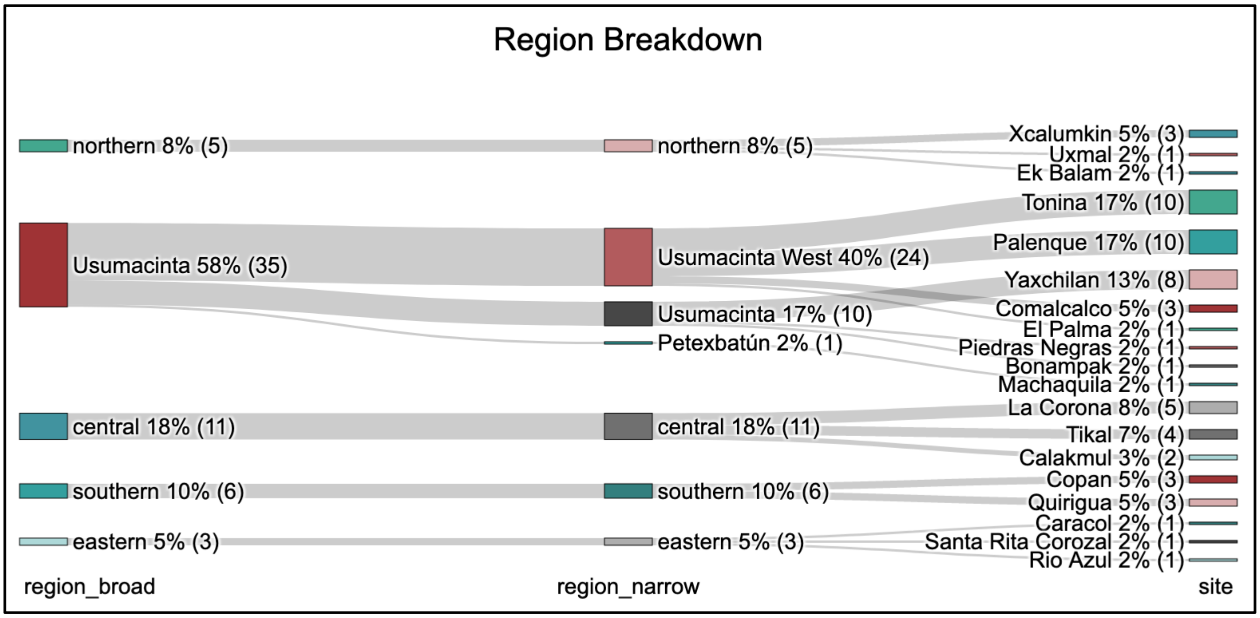
could be bifurcated into two smaller square-shaped glyph block units, then this variable would be valued with “2”. An example of glyph blocks residing within a single block can be seen in Figure 6, from Aguateca Stela 19, which shows the juxtaposition of two blocks within one.

Figure 6. Aguateca Stela 19:E2a-F2a: **ʔi-ʔu-ti-06-KIB** for *ʔi ʔuhti wak kib* ‘and then it happens on 06 Kib’



In order to analyze trends related to the texts’ place of origin, *region_broad*, *region_narrow*, and *site* were added. These colinear variables provide different degrees of specificity in locating where a text was written. *region_broad* is taken directly from the MHD, and may be valued with *northern*, *southern*, *eastern*, *central*, and *Usumacinta*. Since the Usumacinta region is rather large, *region_narrow* was created, providing the same values as *region_broad*, except that the value *Usumacinta* is divided into *Usumacinta*, *Usumacinta West*, and *Petexbatún*. *site* is simply valued with the polity name at which the inscription was written. Figure 7 depicts a Sankey diagram that breaks down the relationship between *region_broad*, *region_narrow*, and *site*, as well as showing the number of texts in each category. See also Appendix B for a tabulated version of this regional breakdown.

Figure 7. Number of texts in region_broad, region_narrow, and site



class is also provided by the MHD, and is valued with either *monumental* or *portable_object*. *period* and *date* are colinear variables that both indicate when a text was written, with *period* either being *early* or *late* (referring to the Early and Late Classic periods). *date* is valued with the Gregorian Calendar date, if a date was written on the text. Table 4 summarizes these variables.

Table 4. Variable Summary

	Variable Name	Description	Possible Values	Variable Type
1	*phrasal_unit*	Does the glyph block corresponds to a complete phrase or part of a phrase?	*phrase, phrase_part*	Nominal
2	*word_unit*	Does the glyph block correspond to a complete word or part of a word?	*word, word_part*	Nominal
3	*grammatical_class*	What syntactic unit corresponds to the glyph block?	*N, NP, V, VP, A, PP, Num, NumClassifier, Pred_Part*	Nominal
4	*semantic_category*	Does the glyph block contribute to a title, name, or emblem glyph?	*title, name, emblem_glyph, date*	Nominal

5	*text_length*	How long is the text that the glyph block belongs to?	Numerical	Metric
6	*text_part*	Is the glyph block found in the beginning, middle, or end of a text?	*beginning, middle, end*	Ordinal
7	*dispersion*	Where is the glyph block located in the text?	Numerical	Metric
8	*#_syllabogram*	How many syllabic signs are in the glyph block?	Numerical	Metric
9	*#_logogram*	How many logographic signs are in the glyph block?	Numerical	Metric
10	*only_syllabic*	Does the glyph block only contain syllabograms?	*syllabic, logosyllabic*	Nominal
11	*#_other*	How many signs that are not syllabograms or logograms are present?	Numerical	Metric
12	*#_total*	How many total signs are in the glyph block?	Numerical	Metric
13	*#_main_sign*	How many main signs are in the glyph block?	Numerical	Metric
14	*#_affix*	How many graphical affix signs are in the glyph block?	Numerical	Metric
15	*conflation/infixation*	How many instances of conflation/infixation occur in the glyph block?	Numerical	Metric
16	*blocks_in_blocks*	How many smaller glyph blocks are compressed within the glyph block?	Numerical	Metric
17	*phonetic_comp*	How many phonetic complements are in the glyph block?	Numerical	Metric
18	*region_broad*	What region was the text written in?	*northern, eastern, southern, central, Usumacinta*	Nominal
19	*region_narrow*	Same as *region_broad*, but further divides *Usumacinta*	*Usumacinta, Usumacinta West, Petexbatún,*	Nominal

20	*site*	What city was the glyph block written in?	Various	Nominal
21	*class*	Was the text monumental or portable?	*monumental, portable_object*	Nominal
22	*period*	Was the block written in the Early or Late Classic Period?	*early, late*	Ordinal
23	*dated*	When was the text written?	Gregorian calendar value	Metric

In addition, there are some variables provided in the MHD that will be useful in valuing the variables found in Table 4. These variables are summarized in Table 5.

Table 5. Supplemental Variables

	Variable Name	**Description**	**Variable Type**
1	*objabbr*	Provides a unique code that identifies the object on which the text is written.	Nominal
2	*blcoord*	Indicates where a glyph block is in a text, providing an alphanumeric code.	Nominal
3	*bllogosyll*	Transcribes a glyph block by displaying the graphemic values of each sign.	Nominal
4	*blmaya1*	Transcribes a glyph block as it would be said in Mayan.	Nominal
5	*blengl*	Translates the glyph block into English	Nominal
6	*blsem*	When applicable, indicates the semantic value of a glyph block. (This is narrower than *semantic_category* above).	Nominal

Due to the time required to analyze each glyph block and fill out the variables of interest which were not provided by the MHD, working with the entire dataset was not feasible. Thus, with the assistance of Chris Wiesen at the Odum Institute, a representative sample was created. To accomplish this, the data was separated into six different ranges corresponding to the length of the text. The length ranges are depicted in the left column of Table 6. After creating these categories, the order of the texts within them was randomized. I then went through this data text by text, referencing the MHD to ensure that the images were usable. If a text contained any glyph blocks that were eroded or unclear, it was culled. This process continued until I was able to acquire the appropriate number of texts, as seen in Table 6 below. The reason for creating the sample in this way was to ensure that the measures of central tendency of the sample reflected those of the entire dataset. The sample dataset contained 60 different objects. A link to the sample can be found in Appendix A.

Table 6. Composition of Sample by Text Length

Text Length	Number of Texts
1 - 4	11
5 - 7	9
8 - 12	11
13 - 20	10
21 - 37	9
38 +	10

Coding in R was used to separate the data and analyze trends (see Appendix C). Note that the code provided here is applicable on the entire MHD, not solely the sample. One such analysis picked out single words that spanned multiple glyph blocks. After culling the data of illegible and unclear signs (cf. lines 3-4 of Appendix C.1), 33,098 observations remained. A subset of the data was taken to examine the prevalence of words that spanned multiple glyph blocks (cf. lines 6-12 of Appendix C.1). To accomplish this, glyphs that were found on the same object (objabbr) and had the same transcription (blmaya1) but different values with regards to syllabographic and logographic make-up (bllogosyll) were collected. The fact that objabbr and blmaya1 matched for two glyph blocks indicated that they belonged to the same text and the same word, while the difference in bllogosyll indicated that the glyph block was different. 263 words were found to span multiple glyph blocks, containing a total of 740 signs.

Additionally, a concentration dispersion analysis was conducted to examine where the glyph block(s) with the greatest number for words was located within a text (see Appendix C.2). First, the number of words per glyph block was calculated by counting the number of words listed in the variable *blmaya1* (cf. lines 6-8 of Appendix C.2). Then, the number of glyph blocks per text was calculated, and texts containing illegible glyph blocks were culled (cf. lines 12-34 of Appendix C.2). To avoid misleading conclusions, all texts with fewer than eight glyph blocks were culled as well (cf. line 35 of Appendix .C2). Finally, for each object, the glyph block(s) with the largest number of words within the text was located, and this index was divided by the entire length of the text, indicating the location of the glyph block(s) proportionally on a scale from zero to one, with zero at the beginning of the text and one at the end of the text (cf. lines 37-46 of Appendix C.2).

For ease of obtaining texts, three functions were created to extract the Mayan transcription and the English translation for a given object code (see Appendix C.3). The first function outputs a dataframe, with each row providing the object code, transcription, and translation for a single glyph block (cf. lines 3-6 of Appendix C.3). The second function outputs the same three pieces of information, but each as its own cell (cf. lines 8-46 of Appendix C.3). Thus, the transcriptions for each block are concatenated together, and all the translations are treated the same. The third function does the same as the second, except that it places square brackets around each glyph block of both the transcription and the translation (cf. lines 48-90 of Appendix C.3).

Appendix C.4 was created to determine the textual productivity of each city. The dataframe *text_per_city* counts up and displays the number of texts that were written at each location (cf. line 9 of Appendix C.4). Similarly, *gb_per_city* adds together the number of glyph blocks among all of the texts at each locations (cf. line 10 of Appendix C.4). These dataframes help provide an idea of the number and lengths of texts written at different sites.

To test my hypotheses, I conducted various linear and logistic regressions. Regression analyses were chosen because they allow for stringent statistic testing of the relationship between a dependent variable and one or more independent variables. Linear regressions are used when the dependent variable is metric, and logistic regressions are used when the dependent variable is nominal or ordinal (DATATab 2023). A linear regression analysis will produce a model equation that predicts the relationship between independent and dependent variables, such that each independent variable is multiplied by some coefficient indicating the effect size it has, and each scaled independent variable is then added together to determine the value of the dependent variable. With logistic regressions, the sum of the scaled independent variables are negated and used in an exponential function, which is then added to 1 and inverted. The reason for this setup is that the logistic model seeks to produce a binary result, where the dependent variable either is or is not present. Regardless of the values of the coefficients on the independent variables, logistic models will always produce a value between 0 and 1. For example, Regression 1.1a, found in Table 9, depicts a logistic regression analyzing the effect of *class*, *period*, *region_broad*, and *only_syllabic* on *phrasal unit*. The dependent variable is binary, with glyph blocks either corresponding to a full phrase or not. If an independent variable has a positive coefficient, than increasing it will increase the likelihood that the outcome will be 1 (corresponding to a full phrase). Similarly, negative coefficients indicates that the independent variable and the dependent variable are inversely related.

To determine if the linear regressions were significant, ANOVA tests were conducted. The ANOVA test is used to conclude whether the coefficients of the independent variables are significantly different from 0 (DATATab 2023). If the model does produce noticeably different results then when the independent variables are excluded, then the model proves to be significant. For logistic regressions, Chi-squared tests are used to check for significance. Similarly to ANOVAs, a Chi-squared test compares the model produced by the logistic regression to a model where each independent variable coefficient is 0. If these two models produce significantly different results, then the Chi-squared test indicates that the logistic model is a good predictor of the dependent variable. While ANOVAs and Chi-squared tests are essentially achieving the same goal, ANOVAs are used for numerical dependent variables, whereas Chi-squared tests are used for categorical dependent variables.

All data tagging was recorded in Microsoft Excel, referencing images and transcriptions provided in the MHD. All regressions, except for the Set 1.3 Regressions, were generated using DataTab. When using this online tool for the Set 1.3 Regressions, DataTab stated that relevant coefficients in the regression models were exactly 0, since DataTab includes fewer significant digits in its output than was needed here. Thus, these regressions were also run in R Studio in order to obtain more precise values.

CHAPTER 5: RESULTS

Table 7 below connects the earlier stated hypotheses with their corresponding linear and logistic regressions. The following tables for each logistic regression only provides the regression coefficient, p-value, and odds-ratio for variables that either were significant or approached significance. For the linear regressions, only the unstandardized coeffient, standardized coefficient, and p-vlaue were included. More detailed tables for each are provided in Appendix D.

Table 7. Hypotheses and their Regressions

	Hypothesis Summary	**Regression**
1	Object class, time period, region, and purely syllabic spellings will influence the phrasal value of a glyph block.	Set 1 Regressions; Tables 8-13
	An increase in the instances of conflation, infixation, and blocks within a block will correlate with glyph blocks that constitute full phrases.	Regression 2; Tables 14-15
	Longer texts will increase the likelihood that glyph blocks will correspond to full phrases, and will increase the number of conflations and infixations.	Set 3 Regressions; Tables 16-19
2	Glyph blocks nearer the end of a text will have more graphical affixes, conflations and infixations, phonetic complements, total signs, and instances of multiple blocks within a single glyph block .	Set 4 Regressions; Appendix D.4
	Glyph blocks denoting names, titles, and emblem glyphs will have more phonetic complements.	Regression 5; Tables 20-21
	Glyph blocks will have a higher number of signs, number of syllabograms, and blocks within blocks when they are written in the Late Classic Period.	Set 6 Regressions; Tables 22-27
3	The northern region will use more syllabograms than the other regions.	Set 7 Regressions; Tables 28-31
	Portable objects will have more purely syllabic spellings than monumental texts.	Regression 8; Tables 32-33

5.1 Glyph Blocks as Syntactic Phrases

5.1.1 Phrasal Unit Regressions (Set 1.1 Regressions)

This first set of regressions tests the impact of text class, date of writing, place of origin, and syllabic spellings on whether or not the resulting glyph block constitutes a syntactic phrase, using the variables *class*, *period*, *region_broad*, and *only_syllabic*. A Chi-squared test indicated that the model produced by the logistic regression is significant.

Table 8. Chi-Squared Test for Regression 1.1a

Chi^2	df	p
65.87	7	<.001

Table 9 depicts the model produced by the logistic regression. In terms *region_broad*, the values *Usumancinta*, *southern*, and *eastern* are omitted, since their p-values were further from significance than that of *central*. Note that *northern* was taken as the reference value.

***Table 9. Logistic Regression 1.1a Model: Effect of* region_broad, class, period, *and* only_syllabic *on* phrasal_unit**

	Coefficient B	p-value	Odds Ratio
Central	0.71	.136	2.03
Monument	0.69	<.001	2
Early	-0.7	.278	0.5
Only_syllabic	-1.06	<.001	0.35

From this data, we can conclude that both *class* and *only_syllabic* have a significant impact on whether a glyph black will correspond to a grammatic phrase, both yielding a p-value of <0.001. The positive value of 0.69 for the value of *monument* under the variable *class* indicates that monuments are more likely to have glyph blocks equating to phrasal units, while portable objects will show less systematicity in grammatical representation. The negative coefficient of -1.06 associated with *only_syllabic* indicates that glyph blocks comprised only of syllabograms are less likely to correspond to a syntactic phrase than glyph blocks that have one or more logograms within them.

Neither *region_broad* nor *period* proved influential in determining phrasal representation when considered with these other variables. The *central* value for *region_broad* was the closest to yielding significance, with a p-value of 0.136. For *period*, the p-value associated with the value *early* is 0.278. *region_broad* was substituted with both *region_narrow* and *site*, though both of these variables only produced p-values further from significance. Similarly, *dated* was tested in place of *period*, again not portraying any significance.

It is worth noting that when taken as the only independent variable, *region_broad* does reveal some significance. Again, a Chi-squared test reveals that the model produced by the logistic regression is significant, where the regression takes only *region_broad* as an independent variable and *phrasal_unit* as a dependent variable.

Table 10. Chi-Squared test for Regression 1.1b

Chi2	df	p
10.36	4	.035

***Table 11. Logistic Regression 1.1b Model: Effect of* region_broad *alone on* phrasal_unit**

	Coefficient B	p	Odds Ratio
Usumacinta	0.57	.209	1.76
Central	0.93	.042	2.54
Southern	0.78	.093	2.19
Eastern	0.8	.181	2.23

With a p-value of 0.042 and a coefficient of 0.93, the value *central* does significantly affect phrasal representation within glyph blocks, indicating that glyph blocks from texts written in the central region are more likely to depict full phrases than in the northern region, since *northern* is the reference. *region_broad* fails to produce any significance when considered with *class* and *only_syllabic* simply because both of these other variables have a stronger correlation with the dependent variable. The large impact of text class and purely syllabic blocks overshadows any influence of region on syntactic representation, though when considered on its own, different regions do prove to have significant differences in their writing.

Another logistic regression was run to determine if any individual sites had an influence on the syntactic correspondence in glyph blocks. Similarly to *region_broad*, *site* only yielded significant results when viewed as the only independent variable. The Chi-squared test yields a p-value of <0.001, and thus we can consider this model.

Table 12. Chi-Squared test for Regression 1.1c

Chi2	df	p
74.78	18	<.001

Table 13 only includes the sites which have a significant effect, or approach significance. A table containing all sites can be found in Appendix D.1. The reference value for this variable is *Santa Rita Corozal.*

***Table 13. Logistic Regression 1.1c Model: Effect of* site *on* phrasal_unit**

	Coefficient B	p	Odds Ratio
La Corona	2.29	.038	9.88
El Palma	2.3	.077	10
Piedras Negras	2.22	.055	9.17
Quirigua	1.83	.099	6.2
Rio Azul	2.56	.035	13
Tonina	2.65	.019	14.17

Table 13 reveals that La Corona, Rio Azul, and Tonina all have positive coefficients and significant p-values (0.38, 0.35, and 0.19, respectively). This indicates that these sites all tend to tie their glyph blocks to syntactic units more often than other sites do.

5.1.2 Use of Graphic Compression Strategies (Regression 1.2)

Again examining the dependent variable of *phrasal_unit*, another regression was run to test the influence of glyph blocks within larger glyph blocks and instances of conflation and infixation on the phrasal value of a glyph block. The regression model was significant, as seen in Table 14.

Table 14. Chi-Squared Test for Regression 2

Chi2	df	p
35.97	2	<.001

A more detailed table for the model given by Table 15 is provided in Appendix D.2.

Table 15. Logistic Regression 2 Model: Effect of conflation/infixation and blocks_in_blocks on phrasal_unit

	Coefficient B	p	Odds Ratio
Conflation/infixation	0.49	.012	1.63
Blocks_in_blocks	1	<.001	2.72

As indicated by Table 15, both *conflation/infixation* and *blocks_in_blocks* have a significant effect on the value of *phrasal_unit*. The positive coefficient on each signifies that glyph blocks are more likely to correspond to syntactic phrases when there are more instances of conflation and infixation, and when multiple glyph blocks are juxtaposed within the space of one block.

5.1.3 Text Length Regressions (Set 1.3 Regressions)

Next, a couple of regressions were generated to examine the potential impact of the length of a text on various dependent variables. First, Regression 1.3a models the relationship between length and the phrasal status of glyph blocks. After running a Chi-squared Test, this model proved to be significant with a p-value of 0.016.

Table 16. Chi-Squared Test for Regression 1.3a

Chi2	df	p
5.8	1	0.016

Table 17. Logistic Regression 1.3a Model: Effect of text_length on phrasal_unit

	Coefficient B	p	Odds Ratio
Text_length	-0.003	0.016	0.994

Also with a p-value of 0.016, the variable *text_length* does show a significant effect on the value of *phrasal_unit*. While the coefficient is quite small in magnitude, the negative value suggests that the shorter a text is, the more likely its glyph blocks will correspond to phrases. Next, a second regression concerning text length was created, taking the presence of conflation and infixation as a dependent variable.

Table 18. ANOVA for Regression 1.3b

Model	df	F	p
Regression	1	4.19	0.041

Table 19. Linear Regression 1.3b Model: Effect of text_length on conflation/infixation

	Unstandardized Coefficients	Standardized Coefficients	
Model	**B**	**Beta**	**p**
Text_length	-0.0004	-0.0572	0.041

Tables 18 and 19 show that both the model and the relationship between *text_length* and *conflation/infixation* are significant (p-values of 0.044 and 0.041, respectively). The negative coefficient indicates that the longer a text is, the fewer graphical blends it uses (though the effect size reflected by the coefficient is quite small).

5.2 Glyph Block Compression

5.2.1 Dispersion Regressions (Set 2.1 Regressions)

To test the influence of location within a text on the number of graphical affixes, the total number of signs, number of phonetic complements, instances of blocks within a larger block, and instances of conflation and infixation, various linear regressions were created. Such regressions used either *text_part* or *dispersion* as the independent variable, and *#_affix*, *#_total*, *conflation/infixation*, *blocks_in_blocks*, or *phonetic_comp* as the dependent variable. This yielded a total of 10 linear regressions. For each, an ANOVA was conducted to determine whether the model was significant. In each case, the model failed to reach significance, suggesting that location within a text has no effect on the composition of the glyph block. Results of each ANOVA are given Appendix D.4.

Since all of the independent variables are numeric, a correlation analysis was conducted for each pair of *dispersion* and one independent variable to see if this yielded more interesting results. However, each correlation analysis similarly failed to show any significant relationship between the variables. These can also be found in Appendix D.4

5.2.2 Phonetic Complement Regression (Regression 2.2)

Another regression was conducted to determine the effect of a glyph block's semantic category on the number of phonetic complements used. Specifically, the values of *name*, *title*, and *EG* were investigated. An ANOVA was run, proving that the model generated by the regression is significant, with a p-value of <0.001.

Table 20. ANOVA for Regression 2.2

Model	df	F	p
Regression	36	3.55	<.001

Table 21. Linear Regression 2.2 Model: Effect of semantic_category on phonetic_comp

	Unstandardized Coefficients	Standardized Coefficients	
Model	**B**	**Beta**	**p**
Name	0.14	0.12	<.001
Title	0.04	0.03	.386
EG	0.03	0.01	.712

Table 21 reveals that glyph blocks corresponding to names of individuals do significantly impact the use of phonetic complements. With a positive coefficient, glyph blocks containing a name are more likely to have a phonetic complement than blocks containing words of other semantic categories. However, neither titles nor emblem glyphs depict the same trend, with p-values of 0.386 and 0.712, respectively. A full version of this table can be found in Appendix D.5.

5.2.3 Time Period Regressions (Set 2.3 Regressions)

Next, three linear regressions were generated to determine the effect of the time period on the composition of glyph blocks. The first regression tests the relationship between time period and the number of total signs within a block. Table 22 indicates that the model is significant via an ANOVA test.

Table 22. ANOVA for Regression 2.3a

Model	df	F	p
Regression	1	9.43	.002

More detailed tables for the model given by Table 23, as well as the models for the following regressions found in Tables 25 and 27, are provided in Appendix D.6.

Table 23. Linear Regression 2.3a Model: Effect of period on #_total

	Unstandardized Coefficients	Standardized Coefficients	
Model	**B**	**Beta**	**p**
Early	-0.74	-0.09	.002

With a p-value of 0.002, this model proves the significance of *period* in determining how many signs are used in a block. The negative coefficient associated with the value *early* indicates that texts written in the Late Classic Period tend to have more signs per glyph block, while Early Classic texts tend to have a smaller ratio.

The relationship between time period and the number of syllabograms in a text also proved significant:

Table 24. ANOVA for Regression 2.3b

Model	df	F	p
Regression	1	9.69	.002

Table 25. Linear Regression 2.3b Model: Effect of period on #_syllabograms

	Unstandardized Coefficients	Standardized Coefficients	
Model	**B**	**Beta**	**p**
Early	-0.68	-0.09	.002

Similarly to Table 23, Table 25 reveals a negative coefficient associated with the Early Classic Period. Thus, in addition to the total number of signs per block, the number of syllabograms specifically increased from the Early Classic to the Late Classic Periods. Another regression was created to see if the number of logograms increased as well; however, this model was not significant.

A third linear regression was run to analyze the link between period and the occurrence of multiple glyph blocks within the space of a single glyph block. Tables 26 and 27 show the significance of the model and the relationship between the variables.

Table 26. ANOVA for Regression 2.3c

Model	df	F	p
Regression	1	4.59	.032

Table 27. Linear Regression 2.3c Model: Effect of period on blocks_in_blocks

	Unstandardized Coefficients	Standardized Coefficients	
Model	**B**	**Beta**	**p**
Early	-0.11	-0.06	.032

Again, the regression reveals a negative coefficient linked with the *early* value of the *period* variable. Thus, placing multiple glyph blocks in the space of one was more common in the Late Classic Period than in the Early Classic Period.

5.3 Preference for Syllabograms

5.3.1 Effect of Region on Block Composition (Set 3.1 Regressions)

The next two regressions were generated to determine the effect that region had on the number of different types of signs within a glyph block. First, we analyze the effect of region on the number of syllable signs.

Table 28. ANOVA for Regression 3.1a

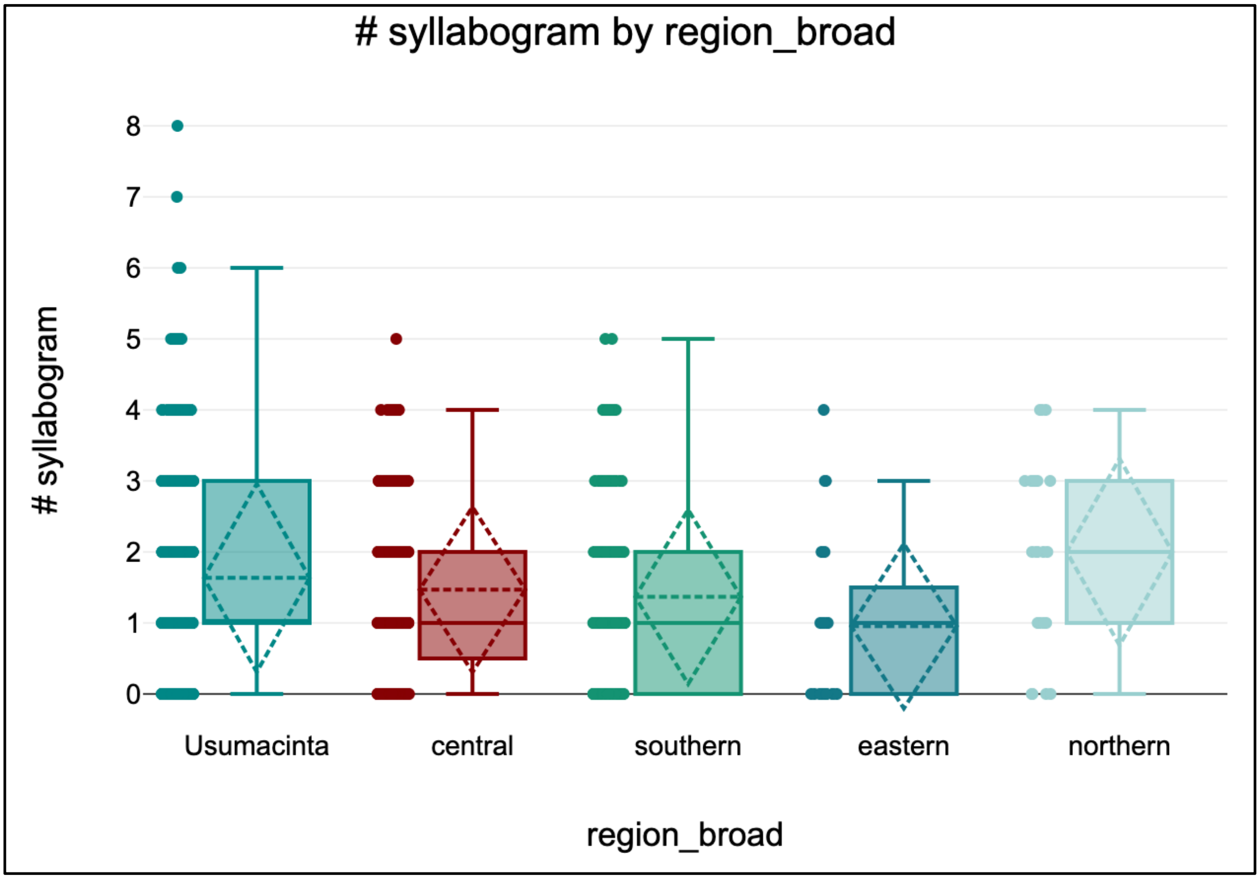
Model	df	F	p
Regression	4	3.98	.003

***Table 29. Linear Regression 3.1a Model: Effect of* region_broad *on* #_syllabograms**

	Unstandardized Coefficients	Standardized Coefficients	
Model	**B**	**Beta**	**p**
Central	-0.53	-0.18	.059
Southern	-0.63	-0.18	.027
Eastern	-1.04	-0.11	.006

Table 28 reveals that the model is significant, with a p value of 0.003. Table 29 indicates that there is a significant effect of region on the number of syllabograms for both the southern and eastern regions, when compared to the northern reference. The central region also approaches significance. The negative coefficients on each indicate that they use fewer syllabograms per glyph block than the northern region. The Usumacinta region was left out here, as it did not approach significance. A full table can be found in Appendix D.7. Figure 8 below depicts this relationship between region and the number of syllabograms per glyph block graphically.

Figure 8. Number of Syllabograms per Glyph Block by Region



The next regression seeks to determine whether logogram use is also dependent on region.

Table 30. ANOVA for Regression 3.1b

Model	df	F	p
Regression	4	4.56	.001

***Table 31. Linear Regression 3.1b Model: Effect of* region_broad *on* #_logograms**

	Unstandardized Coefficients	Standardized Coefficients	
Model	**B**	**Beta**	**p**
Usumacinta	0.84	0.36	.001
Central	0.6	0.22	.019
Southern	0.75	0.23	.004
Eastern	0.74	0.09	.034

Again, the model is significant, as indicated by the p value of 0.001. Here, we can see that each region differs significantly from the northern reference. With positive coefficients, this signifies that the Usumacinta, central, southern, and eastern regions all use more logograms on average than texts in the northern region.

5.3.2 Effect of Object Class on the Presence of Syllabic Spellings (Regression 3.2)

The class of a text (whether it is a monumental or portable object inscription) reveals an interesting pattern in terms of how often purely syllabic spellings are used, as opposed to logosyllabic or logographic representations. The logistic regression that shows this correlation is significant, as seen in Table 32.

Table 32. Chi-Squared Test for Regression 3.2

Chi2	df	p
11.49	1	0.001

***Table 33. Logistic Regression 3.2 Model: Effect of* class *on* only_syllabic**

	Coefficient B	p	Odds Ratio
Monumental	-0.76	<0.001	0.47

With a negative coefficient of -0.76 and a p-value of less than 0.001, the model reveals that monumental inscriptions are less likely to have glyph blocks with purely syllabic spellings. In contrast, portable objects are more likely to use syllabic spellings in lieu of representations involving logograms.

CHAPTER 6: DISCUSSION

6.1 Glyph Blocks as Syntactic Units

6.1.1 Set 1.1 Regressions

As speculated in the hypotheses, monumental texts tend to show a higher degree of linkage between glyph blocks and phrasal units than portable objects do. Monumental texts generally had much more oversight and social pressure in their production, forcing organized, regulated visual representation of the subject material. In contrast, scribes of texts on portable objects had freer range, and may not have required the same degree of scribal schooling as monumental writers did, thus resulting in less of a need to have glyph blocks equating with grammatical units. Regression 1.1a also showed that glyph blocks spelled purely with syllabograms are less likely to be full phrases. This makes sense, since generally speaking, two syllabograms are required to convey the same information as a single logogram. Logograms usually have a CVC phonological structure, so its corresponding syllabic spelling would require one syllabogram for the first consonant and the vowel of the root, and a second syllabogram for the second consonant. Choosing to use a syllabic spelling thus requires more signs, and as such the scribe may require another glyph block to fit the longer spelling. The same is true for logosyllabic spellings: while utilizing syllabograms to either provide grammatical information or to phonetically complement a logogram may ease reading, more signs are used, and thus more area is required. This process may resulting in splitting a phrase or even an individual word that would have otherwise resided within a single glyph block.

Time period ultimately had no effect on the phrasal value of a glyph block. It was hypothesized that phrasal correspondence would increase in later years, under the assumption that writing would become more standardized over time. However, this notion is not necessarily true. Another possibility is that scribes associated glyph blocks as a unit of completeness due to the high proportion of shorter texts through Classic Mayan literature. Since such a large number of texts were short, many of which being only one glyph block in length, scribes would be familiar with such texts and may infer that this visual unit also possessed a grammatical delineation component. This high proportion of shorter texts did not increase with time, so there is no reason to believe that such a reanalysis of glyph blocks by Mayan scribes would only happen in later years.

Region of origin similarly did not show a significant effect when considered with other variables. However, when considered as the sole indicator of the dependent variable *phrasal_unit*, we do see that the central Mayan region produced phrasal glyph blocks significantly more often than the other regions. In the sample, eleven of the texts fell in this region: two from Calakmul, five from La Corona, and four from Tikal. Each of these sites are rather prolific, and we can observe a strong correlation between syntactic phrases and glyph blocks. This seems to support my hypothesis, showing that more prolific sites are more systematic in their representation of glyph blocks as grammatical units. Further testing would be required to determine if this generalization holds statistically for all prolific sites.

6.1.2. Regression 1.2

Conflation, infixation, and juxtaposing multiple glyph blocks in the space of a single block are all methods by which a scribe could compress information into a smaller space. With both conflation and infixation, signs are either blended into a single unit, or one sign is completely encompassed by the other. Either way, two distinct signs wind up occupying the space of one. Writing two glyph block units as a single, more compact glyph block clearly saves space, not within a single block as is seen with conflation and infixation, but spatially in a text. Logistic regression 1.2 reveals that both graphical merging and compression of multiple blocks had a significant effect on the phrasal value of a glyph block, with higher degrees of compression correlating with complete phrases. The major limiting factor that prevents glyph blocks from representing full syntactic phrases is restrictions on space, so employing these tools will help to mitigate this obstacle. This finding suggests that scribes possessed a metalinguistic awareness of phrase structure: if a scribe were writing a syntactic unit that required more signs than is typical, they could choose to conflate, infix, or place multiple blocks in the space of one to guarantee phrasal alignment with the graphic arrangement.

6.1.3. Set 1.3 Regressions

The fact that shorter texts are more likely to have glyph blocks that correspond to grammatical phrases than longer texts was initially surprising. Intuitively, it seemed that longer texts were generally more formal and well regulated. For example, there are plenty of political narratives that span more than 100 glyph blocks, and these were commissioned by high ranking nobles, if not the ruler themselves. In contrast, shorter inscriptions with only a few glyph blocks in total are less likely to be as well planned out and overseen. This coincides with the distinction between portable objects and monuments, where portable texts feasibly cannot reach the same lengths as monumental texts can. As discussed earlier, portable objects were not as well regulated. Despite this, there are a few factors that logically support shorter texts having a higher correspondence between glyph blocks and phrasal units. One reason is the fact that some of the texts in my sample, especially the shortest ones, are only fragmental remains. While at one time these texts were longer, erosion and other damage has caused only a portion to remain. Generally, one physical end of a text will be more damaged than the other. While the end of the text may vary in terms of what semantic content is found there, the beginning of many texts include dates. Dates are very systematically represented, and glyph blocks containing date information almost always correspond to syntactic phrases. Generally, a date glyph block will contain a numeral and either a day sign or a period sign of the long count, either scenario forming a full noun phrase. Thus, when the beginning of the text is the part that survives, it is probable that these glyph blocks will contain date information, and will thus correspond to full phrases. In the smallest text length range of my sample (ranging from one to four blocks), only two of the 11 texts did not contain glyph blocks possessing a phrase with either a calendar round or long count date. Another reason why a shorter text may be more likely to show a phrasal correspondence is the existence of texts intended to have only one glyph block. Name-tagging texts were sometimes one block in length, indicating either the name of the object on which it was inscribed, or the name of the possessor of the object. Thus in name-tagging, these one-block texts would correspond to a full phrase. It would be odd for a scribe to plan a text with only one glyph block that did not correspond to a phrase, since this would imply that the entire text was only one piece of a syntactic phrase. Thus, we would expect that all texts containing only one glyph block before any potential erosion or damage should correspond to a full phrase, and this may skew the effect of text length on phrasal correspondence.

<u>6.2 Glyph Block Compression</u>

6.2.1 Set 2.1 Regressions

Ultimately, placement of a glyph block in relation to the beginning of a text had no impact on its composition. The number of logograms, syllabograms, affixes, conflations and infixations, and the total number of signs were not impacted by position. Thus, it seems that it is not the case that scribes needed to compact signs within glyph blocks at the end of pieces of writing in order to fit the entire text on the writing media. One possible explanation as to why compacting does not occur is that texts were often planned in advance, and sometimes were sketched or traced before carving or painting began. This is especially true for monuments, as portable objects with a linear writing order would not require grid-like planning. By outlining the text prior to writing, scribes could ensure that they utilized their media appropriately. Perhaps scribes did run out of space when sketching the text, but considering that carving or painting had not yet commenced, the text could be erased and rearranged to allow for a more even spread of the number of signs within a glyph block across the text.

6.2.2. Regression 2.2

Regression 2.2 revealed that names do contain more phonetic complements when compared to other semantic categories. Names were, of course, tied to specific individuals. As a result, the logograms used to denote these individuals may have been less recognizable to any given reader. In order to ensure readers would be able to accurately read these logograms, phonetic complements could be used to provide clues on their phonological values. The names of rulers should have been relatively easy to recognize, since they were often the subject of a polity's monumental texts and therefore appeared frequently. As such, phonetic complements may not be as vital in these cases. In contrast, the names of individuals from foreign cities would be harder to discern, and thus would have a larger need for this aid. Names of foreigners do appear regularly in political narratives, either indicating a successful diplomatic mission or the capture of prisoners from a rival city. These names would not be familiar, and thus phonetic complementation would be very beneficial.

Interestingly, neither titles nor emblem glyphs displayed a significant relationship with the number of phonetic complements within the glyph block. Following the logic of complementing names, it seems logical that emblem glyphs and titles, especially those that are specific to a site, would require a reading aid. However, this is not the case. A possible reason for this result is that there are many titles that are used across Mayan texts, regardless of the site of origin. Furthermore, such titles were often spelled purely syllabically, and rarely ever appear represented by a logogram. Trivially, syllabic spellings cannot have phonetic complements: the pronunciation is already overtly represented, so there is no question as to how a syllabic spelling should be read. For example, the title *sajal*, 'feared one', is one of the most common titles, and is generally spelled with three syllabograms: **sa-ja-la**. Similarly, the common title *b'ahkab'*, 'head of land', is consistently spelled as **b'a-ka-b'a**. Given the widespread use of these titles that avoid using phonetic complements, it is understandable that there is not a significant relationship between titles and the use of phonetic clues. Emblem glyphs, on the other hand, are always site specific. Unlike titles, however, emblem glyphs have a very systematic representation, and thus are easily recognizable as emblem glyphs. As discussed in the background section, they always contain three elements: the logograms **K'UHUL** and **AJAW**, and a sign or signs that indicate the specific polity it references. Thus, even if the pronunciation of an emblem glyph was unclear, a reader would still understand its semantic value. Additionally, it was common for sites to have monuments bearing the emblem glyph of a nearby polity, usually in the context of regaling the defeat of a rival leader. Since this sign would be somewhat commonly found in texts at the site, phonetic complementation would be unnecessary.

6.2.3. Set 2.3 Regressions

While the time that a text was written has no effect on the phrasal correspondence with glyph blocks, it does impact the number of signs that are used per block. When comparing the Early Classic Period and the Late Classic Period, the latter utilizes more signs within glyph blocks. Further, the number of syllabograms per block also increased in the Late Classic Period. This supports the findings of other Mayanists, who have found that usage of syllabic signs does increase with time, perhaps as a result of a higher proportion of texts from the northern region being present in later years (Looper & Macri 2022:9). The reason for this compacting may be an effort to utilize space more conservatively: placing more signs per block will result in a need for fewer blocks. Another possibility is that in the Late Classic Period, scribes were more concerned with representing grammatical affixes graphically, rather than leaving them up to interpretation. Grammatical affixes are represented with syllabograms, though I did not code for the number of grammatical affixes per glyph block, so this suggestion in merely speculatory. It is also possible that the number of available syllbograms increased between the Early and Late Classic Periods. Many Mayan syllabograms are derived acrophonically from existing logograms, so it is possible that some syllabograms present in later Mayan texts had not yet evolved from their logographic origins. Similarly, many logographic signs will maintain their form, but adopt new, additional values over time. These new allographs, like the newly derived syllabic signs, would provide scribes in the Late Classic Period with a larger inventory of signs to choose from than their Early Classic counterparts.

The time period also influenced the frequency of putting multiple blocks in the space of a single glyph block, showing an increase of this phenomenon in later times. Similar to the increased number of signs, we can see that this scribal tool resulted in denser glyph blocks in the

Late Classic Period. Again, it is possible that the juxtaposing of glyph blocks is a method of saving space. Another potential reason is that this method of visual arrangement simply had not yet been innovated, or at least not widely spread. All of the texts from the sample that exhibit this phenomenon come from the Late Classic Period, so it seems that Early Classic scribes either rejected this option or had not yet considered it as a possibility.

6.3 Preference for Syllabograms

6.3.1. Set 3.1 Regressions

These regressions revealed that the region in which a text was written does significantly impact its composition, in terms of types of signs. The northern region deviated the most from the other four regions, showing a preference for syllabogram usage higher than that of the other regions, as well as a tendency to avoid logograms more than the others. This finding agrees with the observations of other Mayanists as well. Justeson and Fox claim that phonetic spellings used for recording Yucatecan were common, and can be used to track sound changes from ancestor languages (1989:27). Similarly, Wichmann and Davletshin concluded that the northern Yucatan displays the highest degree of phoneticism in the Mayan region (2006:103). They attribute this trend to the fact that Yucatec Mayan speakers in the Classic Period lived in a diglossic society. While the northern lowlands were occupied by Yucatec speakers, the writing system was adopted from Ch’olan speakers. In order to reflect the Yucatecan language with the Ch’olan script, scribes would add syllabograms to indicate the regional pronunciation of glyphs, as lexemes would have differed between the two language groups. Wichmann and Davletshin also suggest that the increase in phoneticism is a result of linguistic pride: scribes wanted to display their regional variety, and in doing so established a sense of independence from other Mayan writing traditions. By using inferential statistics, the present study shows that the region where a text was written and the number of syllabograms are significantly correlated, which supports the findings of other Mayanists.

6.3.2. Regression 3.2

Regression 3.2 proves that texts inscribed upon monuments and portable objects behaved differently in terms of their use of purely syllabic spellings. Scribes of portable objects show a preference for writing glyph blocks that only contain syllabograms, in comparison to their monumental counterparts. A possible motivation here is a difference in style associated with monumental and portable texts. As previously mentioned, portable objects were not under as much scrutiny as monumental texts, the latter of which would have been on display for a large viewing population. As a result, monuments needed to retain a formal tone, while portable objects were allowed more liberty, permitting more informality in representation.

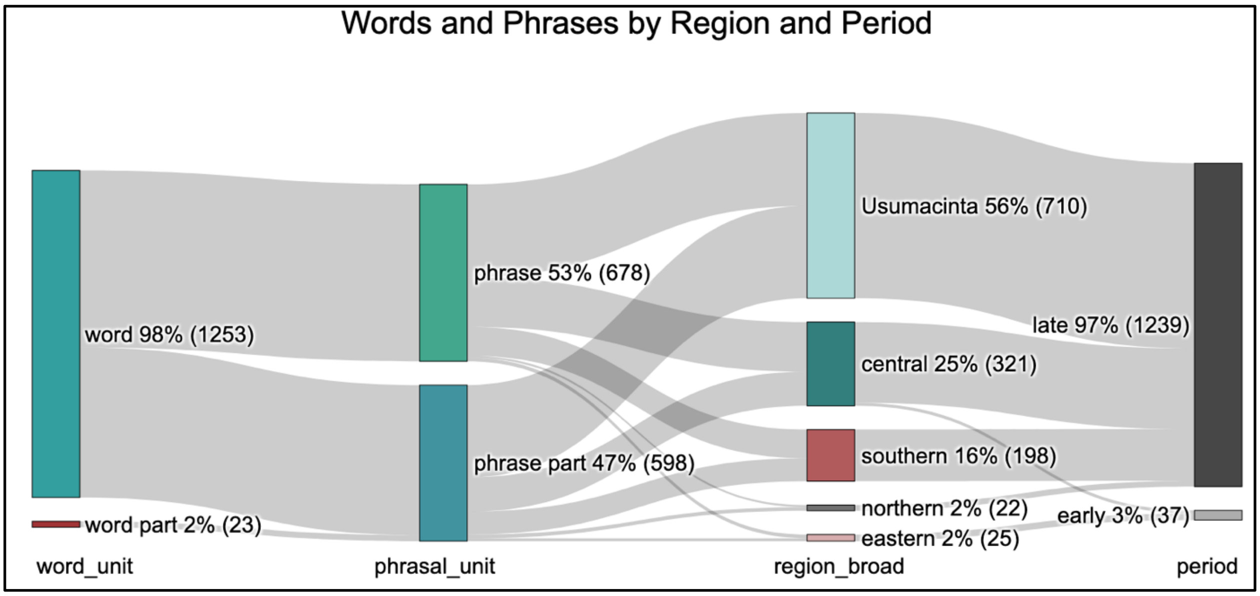
Differences in formality based on object class can be seen in other areas of Mayan writing as well. In Proto-Ch'olan, the third person singular pronoun was given by **ha7in*-, which later derived to **hin*- in its descendent Proto-Western Ch'olan. What is interesting here is that the **hin*- form only appears in portable objects, either as the syllabogram sequence **hi-ni** or as **hi-na** (Mora-Marín 2009:20). Thus, this later innovated form only appears in portable texts, and never appears in monumental inscriptions. Mora-Marín suggests that such a pattern does reflect a difference in formality in terms of media, and that the Proto-Western Ch'olan **hin*- may not have been appropriate for use in monumental texts. Furthermore, instances of quoted speech on portable vessels sometimes used the **hin*- representation, which perhaps indicates that spoken language had become more colloquial than the more conservative textual tradition, at least on monuments. Returning to the idea of syllabic spellings in portable texts, innovative forms tend to be less formal. Indeed, syllabic spellings do increase with time, with the Late Classic Period presenting more syllabic spellings than the Early Classic Period. Thus, it seems that scribes would have viewed such representations as innovative, and made the conscious choice to allow for more informality via this new representational tool in portable objects than in the more conservative monumental texts. Conversely, this may also indicate that logographic and logosyllabic spellings were more prestigious.

CHAPTER 7: CONCLUSION

7.1. General Conclusions

Scribes of Mayan texts display a sufficient degree of metalinguistic awareness through their inscriptions. While glyph blocks may not correspond to full syntactic phrases, they do correspond to full words, and sometimes multiple words, the vast majority of the time. Only 2% of the glyph blocks in my sample depicted part of a single word, suggesting that scribes actively attempted to ensure that words were not bifurcated over multiple glyph blocks (see Figure 9 below). Out of these 11 instances of words spanning multiple glyph blocks, two depict a split of a compound word, such that each block of the compound would normally equate to a single root. For example, on the Hummingbird Vase from Tikal (TIKMT176), we see the compound *ixi'mte'el* 'maize tree?' spread over glyph blocks C and D. Glyph block C contains the prepound *ixi'm* 'maize?', while block D contains the postpound root *te'* 'tree', with two additional syllabograms to form the grammatical suffix *-el*. Thus, each block does consist of a full stem, but in this context, each only comprises a piece of the compound word. Five of the remaining split words all involve the third person singular ergative pronoun *u-* existing in isolation, followed by a noun that is possessed. Further, four of these five instances all appear at La Corona, indicating the this site may have had a preference for a full-sized allogram, which was not widespread throughout the Mayan region. Regardless, it is clear that scribes preferred to keep a word within a single glyph block, and instances where this does not occur are likely attributed to site specific representations or simply a lack of scribal knowledge. The fact that this ergative proclitic is not included in the same glyph block as the root to which it attaches could provide insight as to how scribes thought about clitics. It is even more rare to see grammatical affixes separated from their roots (only twice in the present sample), indicating that graphic cohesion of roots and suffixes was a higher priority than cohesion of roots and proclitics. Thus, it is clear that scribes were metalinguistically aware of the fact that clitics were distinct entities from affixes, and as a result it seems that it was somewhat more permissible to place clitics in a separate glyph block.

Figure 9. Words and Phrases by Region and Period.



While not present in this sample, there are instances of vowel-insertion ligatures in various texts that demonstrate the separation of the ergative clitic *u-* from its root. A ligature describes a syllabogram that is written in one glyph block, but part of its phonological value is tied to a root in an adjacent block. An example is found at Chichen Itza, which includes the sequence of signs **k'a-k'u-pa-ka-la** for *k'ak' upakal* 'Fire-his-shield' (Mora-Marín nd:3). Here, we can see that the syllabogram **k'u** acts as a ligature: the consonantal part serves to provide the final sound in the word *k'ak'*, whereas the vowel represents the ergative proclitic which is morphologically attached to the noun *pakal*. Thus, even though the proclitic *u-* is separated graphically from the noun *pakal*, the scribe was able to choose this representation, perhaps in an effort to save space. It should be noted that this is only possible due to the fact that this proclitic is vowel initial, rather than beginning with an expected glottal stop. Words in proto-Mayan were always consonant initial, having a glottal stop where there other would have been no other consonant (Kaufman 2015:2). However, the third person singular ergative proclitic, as well as the second person proclitics, do not follow this requirement, which allows them to be used in ligatures as in the previous example.

Considering that a large portion of existing Mayan texts are one glyph block in length, scribes would be very familiar with such pieces of writing. Out of the total 1,847 texts present in the MHD, 120 are only one glyph block long (including both full texts and fragments), comprising 6.5% of all texts. As a result, scribes would be used to associating a glyph block as a complete unit of visual arrangement. Such exposure is likely to influence their later writings: if the input that is received is that a large number of texts pack all of their content within a single glyph block, the textual output would similarly place cohesive information within a single unit of space. Such an association with completeness could be reanalyzed with time, so that glyph blocks are interpreted as visual arrangement of syntactic units.

Scribes also show a keen awareness of the metalinguistics of their language through their utilization of spatial compression tools, namely conflation, infixation, and placing multiple glyph blocks within the space of one. As indicated by Regression 1.2, the use of these tools is significantly linked to the phrasal value of a glyph blocks, where more instances of conflation, infixation, and block compression will lead to full phrases, rather than phrase parts. This reflects a conscious knowledge of phrasal constituents, as scribes would compress more information within a glyph block in order to make a glyph block contain a full syntactic phrase.

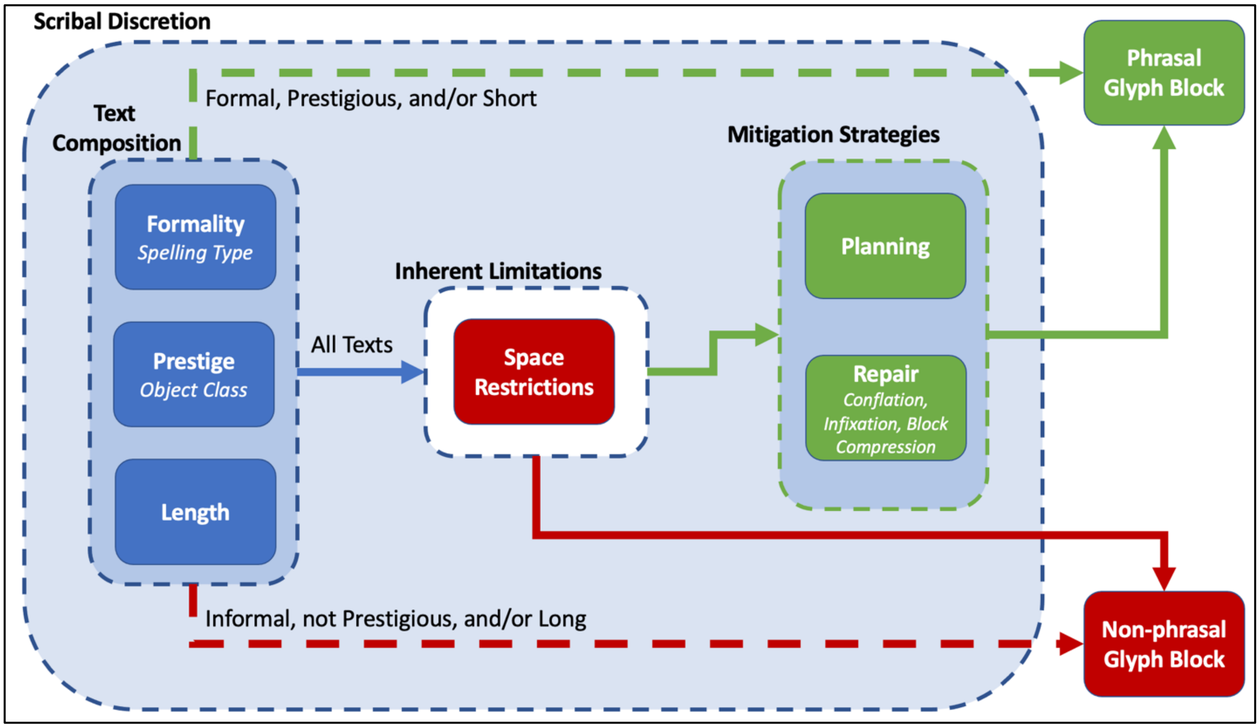
Based on the present study, I argue that glyph blocks act as a form of punctuation via visual arrangement. While I had originally been operating under the assumption that glyph blocks would systematically correspond to full syntactic phrases, this does not seem to be the case. From the sample, just over half (53%) of the glyph blocks contained full grammatical phrases, whereas the others either contained only part of a phrase, or a full phrase along with an additional phrase part (see Figure 9). However, it is clear that having blocks that represented full words was of high importance to scribes. Thus, scribes are utilizing visual arrangement to segment a text into constituent parts; the constituents are simply on a smaller scale than I was expecting. Of course, there were a small number of instances where a glyph block contained only a piece of a word, but such variability is expected, since different scribes would have had different degrees of practice and education related to their craft, and for the other aforementioned potential factors.

Having said this, there are times where it is clear that scribes were cognizant of full syntactic phrases, and not simply words. In the case of transitive verb constructions, scribes would often include the direct object within the same glyph block. In these instances where the verb is antipassive and has an incorporated object, it is evident that scribes were making an effort to keep the verb and the noun together in order to contain the verb phrase within a glyph block. In extension, keeping the verb and its incorporated object together helps to satisfy the preferred argument structure (PAS), which could be another motivating factor for scribes (Mora-Marín 2004:358). PAS indicates that there is a preferred number of clauses associated with verbal arguments. While antipassive verbs may incorporate objects into a single clause, they may also express such entities obliquely by introducing it with a preposition, creating two clauses. However, PAS indicates that having one argument is preferred, encouraging the binding between the antipassive verb and its object. Beyond verbs, we can also see this preference for keeping entire phrases within a glyph block in prepositional phrases. Generally, prepositions are included within the same glyph block as the following noun phrase, ensuring that the entire phrase stays within a single glyph block. Thus, while a phrasal requirement for glyph blocks was not necessary, there are many instances where knowledge of phrasal structure is clear.

While grammaticality does influence glyph block composition, there are other factors that affect this segmentation. One major influence is the size and shape of the media used for writing. Such dimensions will place restrictions on how a text can be segmented, and by extension may change how a glyph block is constructed. For example, a scribe writing a text on a limited amount of space may utilize glyph blocks containing a larger number of signs than would be typical when writing without spacing limitations. Thus, while ensuring grammaticality at the word level in glyph blocks was a priority, non-linguistic variables could also influence how glyph blocks were constructed.

Figure 10 below depicts a flowchart detailing the factors that influence whether a glyph block will be phrasal or not. To start, a text is made to relay certain information to a certain audience. Thus, the text will need to reflect the appropriate amount of formality and prestige, and must also be long enough to convey the message. Formality may surface as the use of syllabic spellings, where such spellings are more informal, than those including logograms. Prestige is tied with the object class, such that monumental inscriptions hold a higher degree of overt prestige. While formal, prestigious, and shorter texts show a preference for phrasal glyph blocks, and informal, non-prestigious, and longer texts show more instances of non-phrasal glyph blocks, all texts are restricted by the size and the shape of the writing medium. Such external forces could sacrifice the grammaticality of a glyph block in order to compensate for the writing surface, though scribes could also use different strategies to mitigate such issues. By sketching texts in advance, and by using repair strategies such as conflation, infixation, and block juxtaposition, scribes are able to create phrasal glyph blocks despite media limitations.

Figure 10. Flowchart of Glyph Block Grammaticality.



The fact that words rarely ever spread across multiple glyph blocks, and that full phrases were frequently contained within a single glyph block as well, suggest that scribes were using glyph blocks as a way of visually punctuating a text. The glyph blocks denote syntactic constituents of varying sizes, with the smallest possible unit generally being a single word. However, as we have seen, this same tool can also denote larger syntactic phrases comprised of multiple words. Thus, glyph blocks do play a syntactic role, and not merely a visual one.

7.2. Limitations and Recommendations

One limitation in this study was interference by partial texts. Not all text fragments are labeled as such in the MHD, and thus we must be cautious when considering results that involve the length of texts. The length of a text fragment would be calculated as being shorter than the original text, and thus these writings could introduce bias into length-related results.

Another factor that should be considered more deeply is the effect of physical shape and size of the writing medium on representational choices. The present work places a strong focus on the influence of grammatical units, and does not investigate such physical limitations in writing. Shape and size of the medium inarguably govern the glyph layout and composition of glyph blocks to some degree, so examining this variable could help to explain the effect of non-linguistic elements on glyph blocks.

In future studies, it would be helpful to tag the sample for glyph blocks that were both a single word and a full phrase. Based on how I recorded values for *phrasal_unit* and *word_unit*, all that was recorded is if a glyph block is a full phrase or not, and whether there are any partial words within the glyph block. Some of the glyph blocks that were given the value *phrase* are only a single word. For example, a single logogram representing a noun would be a full phrase as long as it was not possessed and did not have any corresponding numerals, demonstratives, or adjectives in neighboring glyph blocks. Thus, it is likely that a portion of these “phrasal” glyph blocks are only comprised of a single word. Making a distinction here in tagging the sample would help to see how often scribes would place phrases of multiple words within a single glyph block, without extra noise due to these one word phrases.

A deeper understanding of the degree of metalinguistic awareness possessed by Mayan scribes could be gained via a more stringent tagging of syntactic categories. It would be valuable to see the frequency with which scribes placed transitive verbs with their direct objects in the same glyph block. If scribes did group these elements together, this would suggest that scribes did consider transitive verbs and their objects as verb phrases. In the present study, this situation was not quantified, nor were transitive verbs and intransitive verbs differentiated in the way that they were tagged. Remedying this process could shine some light on the perception of verb phrases by Mayan scribes.

Similarly, it would be fruitful to be more detailed in tagging noun phrases. Some nouns in the sample existed in isolation, and thus formed a syntactic phrase on their own, whereas others existed in more complex constructions. Examining the effect of words other than the noun head in a noun phrase on the composition of glyph blocks could potentially reveal trends regarding the placement of certain elements. For example, there were many instances of possessor noun phrases inhabiting the glyph block after that which contained the possessed noun head. Quantifying this phenomenon would allow us to interpret whether scribes considered possessor noun phrases to exist within the same syntactic phrase as the noun it possesses.

While valuing each of the variables for the glyph blocks in the sample was generally straight forward, *blocks_in_blocks* was sometimes difficult to discern, and thus subject to bias. While most glyph blocks were clearly either separable or inseparable into smaller glyph block units, a few instances proved to be more difficult. Oftentimes glyph blocks that contain smaller, juxtaposed units are more rectangular in shape, given that each individual part is more squarish. The difficulty arises when the glyph block is squarish, but it is possible to draw either a horizontal or vertical line through the block without splitting any individual signs. Should this be considered as one glyph block or two? Consider a glyph block that is comprised of four squarish signs, all about the same size. If the left two blocks are read first in this hypothetical example, we would need to decide whether these four signs simply make up a single block, or if the left two signs and right two signs would each comprise a smaller glyph block, which itself is rectangular in shape. To tackle this bias in the tagging process, a more stringent definition of this variable would be required.

7.3. Future Steps

To further this analysis, I would like to tackle these same questions with a larger sample size. While the sample size of 60 texts for the current study was sufficient in looking at general trends, there were few generalizations that could be drawn in some more specific areas. For example, texts from 19 different sites of origin were included in the sample, but many of these sites only appeared once, and thus it is difficult to draw conclusions about the impact of the *site* variable on the composition of glyph blocks. To obtain a better understanding of the influence of *site*, a more targeted sample would be required, selecting multiple texts from some of the more prolific locations. In extension, a larger sample may reveal more interesting results regarding the impact of time of writing on the resulting glyph blocks.

I would also like to further investigate the effects of conflation and infixation on phrasal representation. While the logistic regression did reveal that such instances of graphical blending do increase the chances that a glyph block will be a grammatical unit, there are certain factors governing conflation and infixation that were not addressed in the present study. Specifically, there are certain pairs of signs that appear conflated or infixed quite often, suggesting that these tools have a preference for merging particular signs. For example, the day name in the 365 day cycle *k'anasiy* is often written where the **K'AN** logogram (with the appearance of a cross) is infixed into the eye of the bird representing the syllable **ʔa**. This pairing of signs appears repeatedly throughout different texts, and suggests that these signs have an affinity toward infixing with each other. Another example of the spelling of the name of a Palenque Lord:

K'inich Janab Pakal I. The sign for **JAN** appears repeatedly within the outline of the **PAKAL**, such that the first is infixed within the latter.

Investigating these limitations and future steps could help to elucidate more ways in which Mayan scribes applied their metalinguistic knowledge towards their craft. While it is clear that scribes used glyph blocks to punctuate their texts, there are many more intricacies and motivations behind their representational choices to be discovered.

APPENDIX A: SAMPLE

Link to sample set

	objabbr (object abbreviation in MHD)
1	CPN21aFrg
2	TNAStu20
3	XLMLnt02
4	TNAStu23
5	CLKFrag06
6	TNAStu25
7	TNAStu07
8	TNAStu33
9	CRCAlt24
10	TNAStu14
11	UXMAlt04
12	YAXBn02
13	CPNT11Crn
14	XLMCol05
15	TIKMT035
16	XLMCol06
17	TIKMT051B
18	CRNEl04
19	CPNT18NGa
20	SRCStvsl
21	PALTISF
22	TIKMT140
23	CMLU26Sp03
24	TNAMon145
25	CRNEl05
26	CLKMSK855
27	PMASt01
28	PALTIST
29	CMLU26Sp09
30	PALTCL
31	EKBCST07
32	CMLSpn02
33	RAZT12mu
34	TNAMon149

35	YAXLnt24
36	TNAMon069
37	PALTIEF
38	YAXLnt39
39	YAXLnt14
40	MQLSt07
41	PALSUBT
42	YAXLnt02
43	YAXLnt01
44	YAXSt35
45	PALHCWF
46	PNGSP
47	PALTISS
48	TIKMT176
49	CRNSQPan04
50	TNAMon159
51	CRNPan03
52	QRGStA
53	PALTISL
54	CRNPan01
55	BPKSt02
56	YAXLnt25
57	QRGStJ
58	PALSLAV
59	PALTIm
60	QRGStC

APPENDIX B: REGION BREAKDOWN

region_broad	*region_narrow*	*site*	Number of texts
Central	Central	Calakmul	2
		La Corona	5
		Tikal	4
Eastern	Eastern	Caracol	1
		Rio Azul	1
		Santa Rita Corozal	1
Northern	Northern	Ek Balam	1
		Uxmal	1
		Xcalumkin	3
Southern	Southern	Copan	3
		Quirigua	3
Usumacinta	Petexbatún	Machaquila	1
	Usumacinta	Bonampak	1
		Piedras Negras	1
		Yaxchilan	8
	Usumacinta West	Comalcalco	3
		El Palma	1
		Palenque	10
		Tonina	10

APPENDIX C: R PROGRAMMING

Appendix C.1: Finding Words Spanning Multiple Glyph Blocks

```
MHD_v5_30_copy <- read_csv("Downloads/MHD_v5_30 copy.csv")
mayan =
MHD_v5_30_copy[,c("objabbr","objclass","blcoord","bllogosyll","blmaya1","blengl")]
mayan2 = mayan[-grep("(\\?\\?)|(^_$)|(^_\\s_$)",mayan$blmaya1),]
mayan3 = mayan2[-
c(grep("(^_$)|(^_\\s_$)",mayan2$bllogosyll),grep("^_$",mayan2$blengl)),]

match = c()
for (i in 1:(nrow(mayan3)-1)){
  if (mayan3[i,"objabbr"]==mayan3[i+1,"objabbr"] &
mayan3[i,"blmaya1"]==mayan3[i+1,"blmaya1"] &
mayan3[i,"bllogosyll"]!=mayan3[i+1,"bllogosyll"]){
    match = rbind(match,mayan3[i,])
    match = rbind(match,mayan3[i+1,])
  }
}

match_new = unique(match)
```

Appendix C.2: Dispersion

```
MHD_v5_30_copy <- read_csv("Downloads/MHD_v5_30 copy.csv")
mayan = MHD_v5_30_copy[,c("objabbr","regionorigin","objclass","katundate","blcoord","bllogosyll","blmaya1","blengl")]

mayan$words_per_block = 0

for(i in 1:nrow(mayan)){
  mayan$words_per_block[i] = length(strsplit(mayan$blmaya1[i],"\\s")[[1]])
}

mayan_ordered = mayan[order(mayan$objabbr),]

gb = MHD_v5_30_copy[MHD_v5_30_copy$bltag==0,]
gb_ordered = gb[order(gb$blsort),]
gb_ordered$counter = 1
gb_num = aggregate(counter ~ objabbr, data = gb_ordered, FUN = sum)

gb = MHD_v5_30_copy[MHD_v5_30_copy$bltag==0,]
mayan2 = gb[-grep("(\\?\\?)|(^_$)|(^_\\s_$)",gb$blmaya1),]
mayan3 = mayan2[-c(grep("(^_$)|(^_\\s_$)",mayan2$bllogosyll),grep("^_$",mayan2$blengl)),]
gb_ordered_new = mayan3[order(mayan3$blsort),]
gb_ordered_new$counter = 1
gb_num_new = aggregate(counter ~ objabbr, data = gb_ordered_new, FUN = sum)

legible = c()
for(i in 1:nrow(gb_num)){
  for(j in 1:nrow(gb_num_new)){
    if((gb_num$objabbr[i]==gb_num_new$objabbr[j])&(gb_num$counter[i] == gb_num_new$counter[j])){
      legible = rbind(legible,c(gb_num$objabbr[i],gb_num$counter[i]))
    }
  }
}

legible = as.data.frame(legible)
legible$V2 = as.numeric(legible$V2)
```

```
legible = legible[legible$V2>=8,]

dispersion = c()

for(i in 1:nrow(legible)){
  obj = mayan_ordered[mayan_ordered$objabbr==legible$V1[i],]
  max = max(obj$words_per_block[1:legible$V2[i]])
  loc = which(obj$words_per_block==max)/legible$V2[i]
  dispersion = rbind(dispersion,c(legible$V1[i],paste(loc,collapse = ", ")))
}

dispersion = as.data.frame(dispersion)
```

Appendix C.3: Obtaining Text

```
MHD_v5_30_copy <- read_csv("Downloads/MHD_v5_30 copy.csv")

text <- function(object){
  obj <- MHD_v5_30_copy[MHD_v5_30_copy$objabbr==object,c("objabbr","blmaya1","blengl")]
  return(obj)
}

text2 <- function(object){
  MHD_v5_30_copy$objkerr[is.na(MHD_v5_30_copy$objkerr)] <- 0
  if(sum(grep("^[0-9]+$",object,value=TRUE)==object)==1){
    mayan <- MHD_v5_30_copy[MHD_v5_30_copy$objkerr==object,"blmaya1"][[1]]
  }
  else{
    mayan <- MHD_v5_30_copy[MHD_v5_30_copy$objabbr==object,"blmaya1"][[1]]
  }
  mayan_split <- strsplit(paste(mayan,collapse = " "),"\\s")[[1]]
  mayan_unique <- c()
  mayan_unique <- append(mayan_unique,mayan_split[1])
  for(i in 1:(length(mayan_split)-1)){
    if(mayan_split[i]!=mayan_split[i+1]){
      mayan_unique <- append(mayan_unique,mayan_split[i+1])
    }
  }
  if(sum(grep("^[0-9]+$",object,value=TRUE)==object)==1){
    english <- MHD_v5_30_copy[MHD_v5_30_copy$objkerr==object,"blengl"][[1]]
  }
  else{
    english <- MHD_v5_30_copy[MHD_v5_30_copy$objabbr==object,"blengl"][[1]]
  }
  english_unique1 <- c()
  english_unique1 <- append(english_unique1,english[1])
  for(i in 1:(length(english)-1)){
    if(english[i]!=english[i+1]){
      english_unique1 <- append(english_unique1,english[i+1])
    }
  }
```

```
  english_split <- strsplit(paste(english_unique1,collapse = " "),"\\s")[[1]]
  english_unique2 <- c()
  english_unique2 <- append(english_unique2,english_split[1])
  for(i in 1:(length(english_split)-1)){
    if(english_split[i]!=english_split[i+1]){
      english_unique2 <- append(english_unique2,english_split[i+1])
    }
  }
  return(c(object,paste(mayan_unique,collapse = " "),paste(english_unique2,collapse = " ")))
}

text3 <- function(object){
  MHD_v5_30_copy$objkerr[is.na(MHD_v5_30_copy$objkerr)] <- 0
  if(sum(grep("^[0-9]+$",object,value=TRUE)==object)==1){
    mayan <- MHD_v5_30_copy[MHD_v5_30_copy$objkerr==object,"blmaya1"][[1]]
  }
  else{
    mayan <- MHD_v5_30_copy[MHD_v5_30_copy$objabbr==object,"blmaya1"][[1]]
  }
  mayan_split <- strsplit(paste(mayan,collapse = "] ["),"\\s")[[1]]
  mayan_unique <- c()
  mayan_unique <- append(mayan_unique,mayan_split[1])
  for(i in 1:(length(mayan_split)-1)){
    if(mayan_split[i]!=mayan_split[i+1]){
      mayan_unique <- append(mayan_unique,mayan_split[i+1])
    }
  }
  mayan_unique[1] <- paste(c("[",mayan_unique[1]),collapse="")
  mayan_unique[length(mayan_unique)] <- paste(c(mayan_unique[length(mayan_unique)],"]"),collapse="")
  if(sum(grep("^[0-9]+$",object,value=TRUE)==object)==1){
    english <- MHD_v5_30_copy[MHD_v5_30_copy$objkerr==object,"blengl"][[1]]
  }
  else{
    english <- MHD_v5_30_copy[MHD_v5_30_copy$objabbr==object,"blengl"][[1]]
  }
  english_unique1 <- c()
  english_unique1 <- append(english_unique1,english[1])
  for(i in 1:(length(english)-1)){
```

```
    if(english[i]!=english[i+1]){
      english_unique1 <- append(english_unique1,english[i+1])
    }
  }
  english_split <- strsplit(paste(english_unique1,collapse = "] ["),"\\s")[[1]]
  english_unique2 <- c()
  english_unique2 <- append(english_unique2,english_split[1])
  for(i in 1:(length(english_split)-1)){
    if(english_split[i]!=english_split[i+1]){
      english_unique2 <- append(english_unique2,english_split[i+1])
    }
  }
  english_unique2[1] <- paste(c("[",english_unique2[1]),collapse="")
  english_unique2[length(english_unique2)] <- paste(c(english_unique2[length(english_unique2)],"]"),collapse="")
  return(c(object,paste(mayan_unique,collapse = " "),paste(english_unique2,collapse = " ")))
}
```

Appendix C.4: Number of Texts and Glyph Blocks per City

```
MHD_v5_30_copy <- read_csv("Downloads/MHD_v5_30 copy.csv")
gb = MHD_v5_30_copy[MHD_v5_30_copy$bltag==0,]
gb_ordered = gb[order(gb$blsort),]
gb_ordered$textlength = 1
gb_num = aggregate(textlength ~ objabbr+siteorigin+regionorigin+objclass+katundate,
data = gb_ordered, FUN = sum)

gb_num$numtext = 1

text_per_city <- aggregate(numtext ~ siteorigin, data = gb_num, FUN = sum)
gb_per_city <- aggregate(textlength ~ siteorigin, data = gb_num, FUN = sum)
```

APPENDIX D: FULL LINEAR AND LOGISTIC REGRESSIONS

Appendix D.1: Set 1.1 Regressions

Logistic Regression 1.1a Model: Effect of *region_broad*, *class*, *period*, and *only_syllabic* on *phrasal_unit*

	Coefficient B	Standard error	z	p	Odds Ratio	95% conf. interval
Usumacinta	0.26	0.47	0.55	.584	1.29	0.52 - 3.22
Central	0.71	0.47	1.49	.136	2.03	0.8 - 5.14
Southern	0.28	0.48	0.59	.558	1.33	0.51 - 3.44
Eastern	1.06	0.9	1.18	.237	2.88	0.5 - 16.72
Monument	0.69	0.19	3.68	<.001	2	1.38 - 2.89
Early	-0.7	0.64	1.09	.278	0.5	0.14 - 1.76
Only_syllabic	-1.06	0.19	5.72	<.001	0.35	0.24 - 0.5
Constant	-0.7	0.49	1.44	.149		

Logistic Regression 1.1b Model: Effect of *region_broad* alone on *phrasal_unit*

	Coefficient B	Standard error	z	p	Odds Ratio	95% conf. interval
Usumacinta	0.57	0.45	1.26	.209	1.76	0.73 - 4.25
Central	0.93	0.46	2.04	.042	2.54	1.04 - 6.22
Southern	0.78	0.47	1.68	.093	2.19	0.88 - 5.45
Eastern	0.8	0.6	1.34	.181	2.23	0.69 - 7.2
Constant	-0.56	0.44	1.26	.207		

Logistic Regression 1.1c Model: Effect of *site* on *phrasal_unit*

	Coefficient B	Standard error	z	p	Odds Ratio	95% conf. interval
Bonampak	1.29	1.14	1.13	.259	3.64	0.39 - 34.21
Calakmul	0.69	1.24	0.56	.578	2	0.17 - 22.95
Comalcalco	1.2	1.16	1.04	.298	3.33	0.35 - 32.2
Copan	1.95	1.24	1.57	.117	7	0.61 - 79.87
Caracol	-17.67	9318.88	0	.998	0	0 - Infinity
La Corona	2.29	1.1	2.08	.038	9.88	1.14 - 85.94
Ek Balam	1.1	1.32	0.83	.404	3	0.23 - 39.61
Machaquila	1.9	1.22	1.55	.12	6.67	0.61 - 73.04
Palenque	1.6	1.1	1.46	.145	4.97	0.57 - 43
El Palma	2.3	1.3	1.77	.077	10	0.78 - 128.78
Piedras Negras	2.22	1.15	1.92	.055	9.17	0.96 - 87.79
Quirigua	1.83	1.11	1.65	.099	6.2	0.71 - 54.15
Rio Azul	2.56	1.22	2.11	.035	13	1.2 - 140.74
Tikal	0.92	1.13	0.81	.419	2.5	0.27 - 23.03
Tonina	2.65	1.13	2.35	.019	14.17	1.55 - 129.57
Uxmal	20.99	9813.79	0	.998	1308993867.94	0 - Infinity
Xcalumkin	0.8	1.25	0.64	.523	2.22	0.19 - 25.72
Yaxchilan	1.24	1.11	1.12	.262	3.46	0.4 - 30.21
Constant	-1.61	1.1	1.47	.142		

Appendix D.2: Regression 1.2

Logistic Regression 1.2 Model: Effect of *conflation/infixation* and *blocks_in_blocks* on *phrasal_unit*

	Coefficient B	**Standard error**	**z**	**p**	**Odds Ratio**	**95% conf. interval**
Conflation/infixation	0.49	0.19	2.51	.012	1.63	1.11 - 2.38
Blocks_in_blocks	1	0.21	4.87	<.001	2.72	1.82 - 4.08
Constant	-1.02	0.23	4.44	<.001		

Appendix D.3: Set 1.3 Regressions

Logistic Regression 1.3a Model: Effect of *text_length* on *phrasal_unit*

	Coefficient B	**Standard Error**	**z**	**p**	**Odds**	**95% conf. interval**
(Intercept)	0.037	0.088	0.42	0.675	1.037	0.874-1.232
Text_length	-0.003	0.001	-2.40	0.016	0.997	0.995-0.999

Linear Regression 1.3b Model: Effect of *text_length* on *conflation/infixation*

	Coefficient B	**Standard Error**	**z**	**p**	**Odds**	**95% conf. interval**
(Intercept)	-2.054	0.146	-14.058	0.000	0.128	0.096-0.17
Text_length	-0.004	0.002	-1.952	0.051	0.996	0.992-1.00

Appendix D.4: Set 2.1 Regressions and Correlations

<u>Regression ANOVAs</u>

Effect of *text_part* on #_*affix*

Model	df	F	p
Regression	2	0.45	.638

Effect of *text_part* on #_*total*

Model	df	F	p
Regression	2	0.26	.77

Effect of *text_part* on *conflation/infixation*

Model	df	F	p
Regression	2	0.06	.945

Effect of *text_part* on *blocks_in_blocks*

Model	df	F	p
Regression	2	0.03	.971

Effect of *text_part* on *phonetic_comp*

Model	df	F	p
Regression	2	1.97	.14

Effect of *dispersion* on #_*affix*

Model	df	F	p
Regression	1	0.43	.51

Effect of *dispersion* on #_*total*

Model	df	F	p
Regression	1	0.12	.724

Effect of *dispersion* on *conflation/affixation*

Model	df	F	p
Regression	1	0.33	.567

Effect of *dispersion* on *blocks_in_blocks*

Model	df	F	p
Regression	1	0.8	.371

Effect of *dispersion* on *phonetic_comp*

Model	df	F	p
Regression	1	0.4	.527

Correlations

Correlation between *dispersion* and *#_affix*

	r	p (2-tailed)
Dispersion and # affix	0.02	.51

Correlation between *dispersion* and *#_total*

	r	p (2-tailed)
Dispersion and # value_total	0.01	.724

Correlation between *dispersion* and *conflation/affixation*

	r	p (2-tailed)
Dispersion and Conflation/infixation	-0.02	.567

Correlation between *dispersion* and *blocks_in_blocks*

	r	p (2-tailed)
Dispersion and Blocks_in_blocks	0.03	.371

Correlation between *dispersion* and *phonetic_comp*

	r	p (2-tailed)
Dispersion and Phonetic_comp	0.02	.527

Appendix D.5: Regression 2.2

Linear Regression 2.2 Model: Effect of *semantic_category* on *phonetic_comp*

	Unstandardized Coefficients	Standardized Coefficients				95% confidence interval for B	
Model	**B**	**Beta**	**Standard error**	**t**	**p**	**lower bound**	**upper bound**
(Constant)	0.24		0.03	8.81	<.001	0.19	0.29
Name	0.14	0.12	0.04	3.69	<.001	0.07	0.21
Title	0.04	0.03	0.04	0.87	.386	-0.05	0.13
EG	0.03	0.01	0.08	0.37	.712	-0.13	0.19

Appendix D.6: Set 2.3 Regressions

Linear Regression 2.3a Model: Effect of *period* on #_*total*

	Unstandardized Coefficients	Standardized Coefficients				95% confidence interval for B	
Model	**B**	**Beta**	**Standard error**	**t**	**p**	**lower bound**	**upper bound**
(Constant)	3.45		0.04	85.59	<.001	3.37	3.53
Early	-0.74	-0.09	0.24	-3.07	.002	-1.21	-0.27

Linear Regression 2.3b Model: Effect of *period* on #_*syllabograms*

	Unstandardized Coefficients	Standardized Coefficients				95% confidence interval for B	
Model	**B**	**Beta**	**Standard error**	**t**	**p**	**lower bound**	**upper bound**
(Constant)	1.56		0.04	42.69	<.001	1.49	1.64
Early	-0.68	-0.09	0.22	-3.11	.002	-1.11	-0.25

Linear Regression 2.3c Model: Effect of *period* on *blocks_in_blocks*

	Unstandardized Coefficients	Standardized Coefficients				95% confidence interval for B	
Model	**B**	**Beta**	**Standard error**	**t**	**p**	**lower bound**	**upper bound**
(Constant)	1.11		0.01	126.4	<.001	1.09	1.13
Early	-0.11	-0.06	0.05	-2.14	.032	-0.21	-0.01

Appendix D.7: Set 3.1 Regressions

Linear Regression 3.1a Model: Effect of *region_broad* on #_*syllabograms*

	Unstandardized Coefficients	**Standardized Coefficients**				**95% confidence interval for B**	
Model	**B**	**Beta**	**Standard error**	**t**	**p**	**lower bound**	**upper bound**
(Constant)	2		0.27	7.38	<.001	1.47	2.53
Usumacinta	-0.37	-0.14	0.28	-1.33	.184	-0.91	0.17
Central	-0.53	-0.18	0.28	-1.89	.059	-1.08	0.02
Southern	-0.63	-0.18	0.29	-2.21	.027	-1.19	-0.07
Eastern	-1.04	-0.11	0.38	-2.75	.006	-1.79	-0.3

Linear Regression 3.1b Model: Effect of *region_broad* on #_*logograms*

	Unstandardized Coefficients	**Standardized Coefficients**				**95% confidence interval for B**	
Model	**B**	**Beta**	**Standard error**	**t**	**p**	**lower bound**	**upper bound**
(Constant)	1.05		0.25	4.22	<.001	0.56	1.53
Usumacinta	0.84	0.36	0.25	3.33	.001	0.34	1.33
Central	0.6	0.22	0.26	2.35	.019	0.1	1.11
Southern	0.75	0.23	0.26	2.86	.004	0.24	1.26
Eastern	0.74	0.09	0.35	2.13	.034	0.06	1.42

Appendix D.8: Regression 3.2

Logistic Regression 3.2 Model: Effect of *class* on *only_syllabic*

	Coefficient B	Standard error	z	p	Odds Ratio	95% conf. interval
Monument	0.76	0.21	3.56	<.001	2.14	1.41 - 3.25
Constant	1.26	0.19	6.61	<.001		

REFERENCES

Boot, Erik. 2005. A Preliminary Overview of Common and Uncommon Classic Maya Vessel Type Collocations in the Primary Standard Sequence. *Electronic document*: 1-22.

Bricker, Victoria. 2004. Mayan. In Roger Woodard (ed.), *The Cambridge Encyclopedia of the World's Ancient Languages*, 1041-1070. Cambridge: Cambridge University Press.

Carlson, John B., & Linda C. Landis. 1985. Bands, Bicephalic Dragons, and Other Beasts: The Skyband in the Maya Art and Iconography. *The Palenque Round Table Series* (6): 115-140.

Carlson, John B. 1988. Skyband Representations in Classic Maya Vase Painting. In Elizabeth P. Benson & Gillett G. Griffin (eds.), *Mayan Iconography*, 277-369. Princeton: Princeton University Press.

Coe, Michael D. 1973. The Maya Scribe and His World. *Grolier Club*, 160. New York.

Coon, Jessica. 2017. Ch'ol. In Judith Aissen, Nora C. England, & Roberto Zavala Maldonado (eds.), *The Mayan Languages*, 648-684. London & New York: Routledge.

DATAtab Team. 2023. DATAtab: Online Statistics Calculator. DATAtab e.U. Graz, Austria. URL https://datatab.net.

Grube, Nikolai. 2021. Punctuation Marks in Ceramic Texts. Textdatenbank und Wörterbuch des Klassischen Maya 19: 1-9.

Hofling, Charles A. 2017. Comparative Maya (Yucatec, Lacandon, Itzaj, and Mopan Maya). In Judith Aissen, Nora C. England, & Roberto Zavala Maldonado (eds.), *The Mayan Languages*, 685-759. London & New York: Routledge.

Houston, Stephen D. 2011. All Things Must Change: Maya Writing over Time and Space. In Elizabeth H. Boone & Gary Urten (eds.), *Their Way of Writing: Scripts, Signs, and Pictographies in Pre-Colombian America*, 21-42. Washington, DC: Dumbarton Oaks Research Library and Collection.

Josserand, J. Kathryn. 1986. The narrative structure of hieroglyphic texts at Palenque. *Sixth Palenque Round Table*, 8: 12-31.

Justeson, John S. 1978. Mayan Scribal Practice in the Classic Period: A Test-case of an Explanatory Approach to the Study of Writing Systems. *University Microfilms International*. Stanford: Stanford University PhD Dissertation.

Justeson, John S. 1986. The Origin of Writing Systems: Preclassic Mesoamerica. *World Archaeology* 17: 437-458.

Justeson, John S. 1989. The Representational Conventions of Mayan Hieroglyphic Writing. In William F. Hanks & Don S. Rice (eds.), *Word and Image of Maya Culture: Explorations in Language, Writing, and Representation*, 25-38. Salt Lake City: University of Utah Press.

Justeson, John S., & James A. Fox. 1989. Hieroglyphic evidence for the languages of the Lowland Maya. *Unpublished*: 1-50.

Kaufman, Terrence. 2015. Initial glottal stop in Mayan languages: 1-15.

Law, Danny & David Stuart. 2017. Classic Mayan: An Overview of Language in Ancient Hieroglyphic Script. In Judith Aissen, Nora C. England, & Robert Z. Maldonado (eds.), *The Mayan Languages*, 128-172. London & New York: Routledge.

Looper, Matthew G. and Martha J. Macri. 1991-2023. Maya Hieroglyphic Database. Department of Art and Art History, California State University, Chico. www.mayadatabase.org.

Looper, Matthew G. and Martha J. Macri. 2022. MHD Reference Materials 5: The Historical Development of the Maya Script: Preliminary Results. *Glyph Dwellers* 75: 1-14.

Martin, Simon. 2020. Ancient Maya Politics: A Political Anthropology of the Classic Period 150–900 CE. *Cambridge University Press*: 1-520.

Maxwell, Judith M. 2009. Mayan Languages. In Keith Brown & Sarah Ogilvie (eds.), *Concise Encyclopedia of Languages of the World*, 705-709. Oxford: Elsevier.

Melchert, H. Craig. 2015. Luvian. In Roger D. Woodard (ed.), *The Cambridge Encyclopedia of the World's Ancient Languages,* 576-584. Cambridge: Cambridge University Press.

Mora-Marín, David. 2004. The Preferred Argument Structure of Classic Lowland Mayan Texts: 337-359.

Mora-Marín, David. 2009. Reconstruction of the Proto-Ch'olan Demonstrative Pronouns, Deictic Enclitics, and Definite Articles. *Transactions of the Philological Society* 107 (1): 98-129.

Mora-Marín, David. 2022. An Evaluation of the Recent Proposal of Punctuation Marks on Mayan Ceramic Texts. *Mesoamerican Linguists at UNC.*

Mora-Marín, David. Nd. Vowel-insertion Ligatures in Mayan Hieroglyphic Writing: Implications for Spelling Conventions: 1-45.

Morley, Sylvanus G. 1937. 38 The Inscriptions of the Peten. *Carnegie Institution of Washington Publication 437* 2: 1-628. Washington, D. C.: Carnegie Institution of Washington.

Nebes, Norbert & Peter Stein. 2015. Ancient Southern Arabian. In Roger D. Woodard (ed.), *The Cambridge Encyclopedia of the World's Ancient Languages,* 454-487. Cambridge: Cambridge University Press.

Polian, Gilles. 2017. Morphology. In Judith Aissen, Nora C. England, & Robert Z. Maldonado (eds.), *The Mayan Languages*, 201-225. London & New York: Routledge.

Polian, Gilles. 2017. Tseltal and Tsotsil. In Judith Aissen, Nora C. England, & Roberto Zavala Maldonado (eds.), *The Mayan Languages*, 610-647. London & New York: Routledge.

Proskouriokoff, Tatiana. 1960. Historical implications of a pattern of dates at Piedras Negras, Guatemala. *American Antiquity* 25 (4): 454-475.

Rix, Helmut. 2015. Etruscan. In Roger D. Woodard (ed.), *The Cambridge Encyclopedia of the World's Ancient Languages,* 943-966. Cambridge: Cambridge University Press.

Rogers, Henry. 2005. Cuneiform. In Henry Rogers (ed.), *Writing Systems: A Linguistic Approach,* 1-322. Oxford: Blackwell Publishing.

Schmitt, Rüdiger. 2015. Old Persian. In Roger D. Woodard (ed.), *The Cambridge Encyclopedia of the World's Ancient Languages,* 717-741. Cambridge: Cambridge University Press.

Wichmann, Søren, & Albert Davletshin. 2006. Writing with an Accent: Phonology as a Marker of Ethnic Identity. *Maya ethnicity: The construction of ethnic identity from the preclassic to modern times* (9): 99-106.

Woodard, Roger D. 2015. Introduction. In Roger D. Woodard (ed.), *The Cambridge Encyclopedia of the World's Ancient Languages,* 1-18. Cambridge: Cambridge University Press.

Zender, Marc U. 1999. Diacritical Marks and Underspelling in the Classic Maya Script: Implications for Decipherment. *National Library of Canada*. Calgary: University of Calgary MA Thesis.