

# Chapter 10

## Analyticity and syntheticity in Coptic: Noun incorporation, word segmentation, and clitics

So Miyagawa<sup>a</sup>

<sup>a</sup>University of Tsukuba

This chapter examines Sahidic Coptic morpho-syntax, focusing on prenominal verb states, clitics, word segmentation, and noun incorporation using the Coptic SCRIPTORIUM corpus. It analyzes noun and pseudo-noun incorporation, word segmentation complexities, and clitic categorization. The study addresses three main questions: the characteristics of Coptic noun incorporation, the impact of segmentation on morpho-syntactic boundaries, and the distinction and role of clitics in Coptic grammar. This research contributes to understanding Coptic morpho-syntax and its typological features, supporting further linguistic studies and comparisons with Afro-Asiatic languages.

本章では、Coptic SCRIPTORIUMコーパスを用いて、動詞の前名詞形（連語形）、接語、語分割、名詞抱合に焦点を当て、コプト語サイド方言の形態統語的特徴を調査する。名詞抱合と擬似名詞抱合の分析、語分割の複雑さの検討、接語の分類を行う。研究では主に3つの問いに取り組む：コプト語における名詞抱合の特徴、語分割が形態統語的境界の解釈に与える影響、およびコプト語文法における接語の区別と役割である。本研究は、コプト語の形態統語論とその類型論的特徴の理解に貢献し、アフロ=アジア諸語との言語学的研究と比較を支援する。

### 1 Introduction

This study seeks to elucidate the characteristics of the Coptic Egyptian morphosyntax, with a particular focus on the prenominal state of the verb in the



context of clitics, word segmentation, and noun incorporation through the lens of linguistic typology. Coptic Egyptian represents the final historical phase of the Egyptian language lineage, a unique branch of the Afro-Asiatic language family. With a recorded history spanning over five millennia (see Kammerzell 2000: 97), Egyptian holds the distinction of being the language with the longest traceable record of grammatical change via written documents.

This study delves into the morphological transitions of Coptic verbs, casting light on the syntactic and morphological synthesis within Coptic and exploring the concept of "wordhood" in this context. The research questions addressed in this chapter are threefold. First, what are the characteristics and extent of noun incorporation in Coptic? This question aims to investigate how noun incorporation manifests in Coptic, examining its morpho-syntactic properties, semantic constraints, and productivity across different noun classes. Second, how do word-segmentation strategies influence the interpretation of morpho-syntactic boundaries in Coptic, and which approach is optimal for typological analysis? This question explores the impact of various word-segmentation practices on the understanding of Coptic morpho-syntax and seeks to identify the most suitable segmentation method for cross-linguistic comparison. Finally, what are the morphological, syntactic, and phonological properties of the prenominal state of the verb in Coptic, and how does it function in marking grammatical relationships and interacting with other elements in the language's morpho-syntactic structure? This question delves into the nature of the prenominal state of the verb, its role in Coptic grammar, and its significance for understanding the language's typological characteristics.

The degree of synthesis in the Coptic language has been a subject of considerable debate among scholars. According to the experts, Coptic is:

- Polysynthetic (Loprieno 1995: 51, 92, 220)
- Synthetic (Haspelmath 2014: 121)
- Analytic (Reintges 2011a,b, Egedi 2007)

Synthesis in linguistics refers to the degree to which words in a language are comprised of multiple morphemes, which are the smallest units of meaning. A high index of synthesis indicates a synthetic language, where words often contain several morphemes. Conversely, a low index denotes an analytic language, characterised by a prevalence of single morphemes per word. At the extreme end of the synthetic spectrum we find polysynthetic languages, where a single word may encompass enough morphemes to convey a complete sentence. Upon initial

inspection of a printed Coptic text, such as the example shown in (1),<sup>1</sup> one might conjecture that Coptic exhibits characteristics of a polysynthetic language.

- (1)  $\chi\epsilon\epsilon\kappa\epsilon\mu\epsilon\rho\epsilon\pi\epsilon\tau\chi\iota\tau\omicron\gamma\omega\kappa$   
*če-e-k-e-mere-p-et-hi-touô-k*  
 COMP-OPT-2SG.M-OPT-love.PNOM-DEF.SG.M-on-bosom-2SG.M  
 $\mathfrak{n}\mathfrak{g}\mathfrak{m}\mathfrak{e}\mathfrak{s}\mathfrak{t}\mathfrak{e}\mathfrak{p}\mathfrak{e}\mathfrak{k}\mathfrak{x}\mathfrak{a}\mathfrak{x}\mathfrak{e}$   
*n-g-meste-pe-k-čače*  
 CONJ-hate.PNOM-POSS.SG.M-2SG.M-enemy.M  
 ‘(you have heard) that you shall love your neighbour and you shall hate your enemy’  
 (NT Matthew 5:43 from Wilmet (1958: 958))

In (2)–(4), I provide examples from languages that are representative of the polysynthetic, synthetic, and analytic typological categories. These serve as a point of comparison for the Coptic text previously discussed (see (1)). Fortescue (2013: 252) categorises Classical Nahuatl as polysynthetic due to its propensity for incorporating numerous morphemes into single words, see (2).<sup>2</sup>

- (2) Example of a polysynthetic language  
 o mitzmoteochihuilitzino  
 $\bar{o}=\emptyset\text{-mitz-mo-te}\bar{o}\text{-chihui-lih-tzin-oh}$   
 PST=3.SBJ-2SG.OBJ-REFL-god-make-APPL-HON-VBLZ.PST  
 (Classical Nahuatl (Uto-Aztecan))  
 ‘He blessed you’  
 (Camino del Cielo, folio 107v (de León 1611), annotated and translated by Mitsuya Sasaki)

Old Nubian is identified as synthetic,<sup>3</sup> particularly in its verbal morphology, which van Gerven Oei (2014: 171) details extensively, see (3).<sup>4</sup>

<sup>1</sup>The romanization of Coptic is following Grossman & Haspelmath (2014).

<sup>2</sup>The hyphenated version is a linguistically interpreted version by a Nahuatl linguist, and the original is in Classical Nahuatl. <h> is not written in the original but it should be written in the linguistically interpreted text.

<sup>3</sup>In Old Nubian, a superlinear stroke on a consonant always means /i/ before the consonant, see van Gerven Oei (2022: 38).

<sup>4</sup>A different morphological interpretation with regard to morpheme boundaries and functions of morphemes was also proposed by Satzinger (2010: 751).

(3) Example of a synthetic language

εἷτοϥ οὐελλο      Δῖπποϥ οὐελλα    ΔΟΥᾶρα      ἄλεξανδρεν  
*eittou ouel-Ø-lo      dippou ouel-l-a    dou-ar-a      aleksandre-n*  
 woman one-NOM-FOC village one-DIR EXIST-PST1-PRED Alexandria-GEN  
 ʔḱroϥla      (Old Nubian (Nilo-Saharan))  
*ṣik-gou-la*  
 district-ground-DAT

‘There was a woman living in a village, in the district of Alexandria.’

(Miracles of St. Mina, p.1,

ll. 5-8, Browne (1994: 5), annotation following van Gerven Oei (2022: 67), a different morphological interpretation regarding morpheme boundaries and functions of morphemes was proposed by Satzinger (2010: 751))

By contrast, Classical Chinese exemplifies a highly analytic structure, surpassing even Modern Mandarin — a language often cited as a paradigm of analyticity — given its minimal use of inflectional morphemes, see (4).

(4) Example of an analytic language

不 尚 賢 使 民 不 爭      (Old Chinese (Sino-Tibetan))  
*bù shàng xián shǐ mín bù zhēng*  
 NEG respect clever CAUS people NEG conflict

‘If you don’t respect the clever, you never let people be in conflict’

(Laozi, *Tao Te Ching*, 3 (Hachiya 2008), with Modern Mandarin pronunciation)

The concept of a word boundary (WB henceforth) is crucial for determining the index of synthesis in a language. For the languages previously mentioned, WBs are inferred based on modern counterparts.

Additionally, the absence of spaces in traditional Coptic manuscripts complicates the task of identifying WBs in Coptic texts.

The question arises, then: What do the spaces in our printed Coptic texts signify? Takla (1998) provides insight into the history of word division in Coptic literature:

The first attempt to divide the words was probably done by the scholars in Europe as early as [the] 17th and 18th centuries. Foremost among them is the Coptic Raphael al-Tukhi, residing in the Vatican. Eventually the same system was employed by Copts when they published the first printed texts during the days of Pope Cyril IV or shortly after. (Takla 1998: 121)

Following Takla (1998)'s account, the spaces found in modern Coptic texts are a relatively recent development and may not accurately reflect authentic word boundaries. It is worth noting that instances of segmentation exist in Coptic manuscripts predating the 17th century, such as the *Macquarie Magical Papyrus*.<sup>5</sup> Thus, it is probable that the segmentation approach employed by European scholars was influenced by an existing Coptic tradition of word division.

## 2 Typology of spacing on Coptic texts

In the study of Coptic texts, scholars have adopted various strategies for segmenting morpheme groups, particularly concerning the placement of spaces. For my analysis, these practices have been classified into four types, see Figure 1.

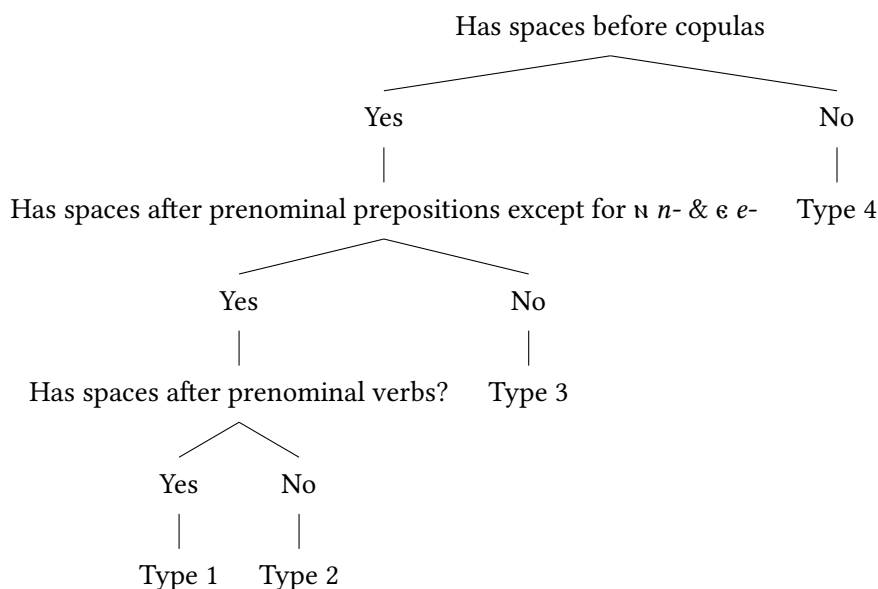


Figure 1: Typology of Coptic spacing

The first classification makes a distinction based on the presence or absence of spaces preceding copulas. When a space does not precede a copula, the text conforms to what I refer to as Type-4 spacing. This approach to spacing is consistent with the standards set forth by Kuhn (1956a), see (5), and further supported by Wilmet (1958), see (6).

<sup>5</sup>Choat & Gardner (2014); Example (10) from the *Macquarie Magical Papyrus* shows that the bound groups were divided by upper dots.

- (5) Type 4: (Kuhn 1956a: 12) with the copula *ne pe* attached to the word *naï pai* before it

ⲉⲧⲧⲉⲡⲁⲓⲉ	ⲛⲉⲥⲛⲏⲩ	ⲟⲩⲁⲓⲁⲑⲟⲛ	
<i>etbe-pai-ce</i>	<i>ne-snêu</i>	<i>ou-agat<sup>h</sup>on</i>	
because-DEM.SG.M-therefore DEF.PL-brother.PL INDEF.SG-good			
ⲛⲁⲛⲡⲉ	ⲉⲧⲣⲉⲛⲧⲟⲥⲏ̅	ⲉⲡⲛⲟⲩⲧⲉ	[...]
<i>na-n-pe</i>	<i>e-tre-n-toen</i>	<i>e-pnoute</i>	[...]
DAT-1PL-COP.SG.M DIR-CAUS-DEF.PL-cleave DIR-DEF.SG.M-god [...]			

‘So therefore, brethren, it is good for us to cleave to God [...]

(Besa’s *Letters and Sermons*, ‘On Faith, Repentance, and Vigilance II,’ 1, (Kuhn 1956b: 11))

- (6) Type 4: (Wilmet 1958) with the copula *ne pe* attached to the word *naï pai* before it

ⲁⲩⲱ ⲁⲓⲣⲙⲛⲧⲣⲉ	ⲭⲉⲡⲁⲓⲡⲉ	
<i>auô a-i-r-mntre</i>	<i>çe-pai-pe</i>	
and PST-1SG-do.PNOM-witness COMP-DEM.SG.M.ABS-COP.SG.M		
ⲡⲥⲱⲧⲡⲓ	[...]	
<i>p-sôtp</i>	[...]	
DEF.SG.M-choice [...]		

‘and I witnessed that this one is the choice [...]

(NT John 1:34 (Wilmet 1958: 377))

The typology of spaces in relation to copulas and other morphological markers provides the basis for further classification. Type-3 segmentation is characterised by a space before a copula coupled with the absence of spaces following prenominal prepositions. This segmentation pattern is prominent in the field of Coptology; for example, Bentley Layton adopts this approach in his reference grammar of the Sahidic dialect of Coptic, where he also provides a theoretical framework for it (Layton 2011: 25–26).

In contemporary scholarship, Layton’s Bound Group (BG henceforth) model is frequently cited, delineated in his authoritative grammar work and represented as Type 3 in Figure 1. The largest corpus of Coptic text, the Coptic SCRIPTORIUM,<sup>6</sup> has implemented this spacing typology. Bentley Layton’s formulation of

<sup>6</sup>Coptic Scriptorium: Digital Research in Coptic Language and Literature (<https://copticcriptorium.org/>, last accessed 13 January 2024). See Schroeder & Zeldes (2016).

a BG is rooted in a prosodic framework, positing that a BG encapsulates a single stress point. Martin Haspelmath expands on this by characterising a BG as a “stress group” (Haspelmath 2014).

Nonetheless, it is crucial to acknowledge that there is no direct evidence of stress in morphs from the period of active Sahidic Coptic usage. Despite this absence, Layton’s propositions on stress patterns find corroborative evidence, albeit indirectly, through Prince’s analysis of Coptic pronunciation in liturgical contexts from the 20th century (Prince 1902).

While Layton’s BG theory provides a prosodic and phonological rationale for the cohesion of non-stressed morphs, it stops short of thoroughly addressing their morpho-syntactic interconnectedness, suggesting a potential avenue for future exploration.

It is also important to note that Layton (2011) employs a special hyphenation, where most morphemes, except for articles and possessive articles, are separated by hyphens. This hyphenation-based segmentation differs from the many other Type-3 editions.

Despite this difference, Layton’s work remains a fundamental reference for the study of Coptic grammar and provides valuable insights into the language’s structure and morpho-syntax. The hyphenation-based segmentation used by Layton serves to highlight the morphological composition of Coptic words and phrases, while the Type-3 segmentation focuses more on the prosodic and syntactic units of the language.

In this chapter, we primarily focus on the Type-3 segmentation as a basis for analysing word boundaries and calculating the morpheme-to-word ratio, as Type-3 segmentation aligns more closely with the concept of the bound group and provides a suitable framework for cross-linguistic comparison. However, we acknowledge the importance of Layton’s work and the alternative perspective offered by his hyphenation-based segmentation.

The final distinction shown in Figure 1 hinges on the spacing following verbs which are in the prenominal state. The absence of a space after such verbs denotes Type-2 segmentation, see (7). Conversely, if there is a space following prenominal verbs, the text is categorised as Type 1, a style utilised by scholars such as Till (1942), Steindorff (1883), and Quecke (1984).

(7) Type 2: Layton (2011: 219)

ḿ-ḏNOK	ḏN	ḿMḏTḒ	ḿḒ	ḏLLḏ	ḏNOK	Nḿ-ḿIḐT
<i>n-anok</i>	<i>an</i>	<i>m-mate</i>	<i>pe</i>	<i>alla</i>	<i>anok</i>	<i>nm-p-iôt</i>
NEG-1SG	NEG	LOC-very	COP.SG.M	but	1SG.M	with-DEF.SG.M-father

ენტ-αφ-ταουο-ει

*ent-a-f-taouo-ei*

REL-PST-3SG.M-send-1SG

‘It is not a matter of Me alone, but of Me and the Father who sent Me’

(NT John 8:16 from Layton (2011: 219))

(8) presents one instance of Type-1 segmentation, a style characterised by the insertion of spaces after prenominal verbs.

(8) Type 1: Till (1942: 51)

πχοεις

ναφωπε

εογα

πε

αγω πεφραν

*p-choeis*

*na-šôpe*

*e-oua*

*pe*

*auô pe-f-ran*

DEF.SG.M-Lord FUT-appear CIRC-one COP.SG.M and POSS.SG.M-3SG.M-name

εογα

πε

*e-oua*

*pe*

CIRC-one COP.SG.M

‘The Lord will become one and his name is one’

(NT Zechariah 14:9)<sup>7</sup>

Till’s method conforms to Type 1, yet it is distinguished by its systematic use of spacing to differentiate homonyms across different parts of speech. Such an approach, while methodical, could be considered more prescriptive or artificial compared to other segmentation practices, as it intentionally modifies the text structure to clarify ambiguity in homonymy.

In conclusion, the various approaches to word segmentation and spacing in Coptic are synthesised in Figure 1. This visualization provides a systematic overview of the classification scheme applied to Coptic text segmentation.

While the typology of Coptic spacing provides valuable insights into the various approaches to segmentation, it is crucial to determine which type of spacing most accurately represents word boundaries in the language. For the purposes of this study, we argue that Layton’s bound group (Type 3) is the most suitable representation of word boundaries in Coptic from a typological perspective.

Layton (2011)’s bound group is characterised by a single stress and often corresponds to a grammatical word, aligning with Haspelmath (2023)’s definition of a word as a “minimal form that can express a complete grammatical word”. By treating bound groups as words, we can better capture the morpho-syntactic units of Coptic and analyze their properties in relation to cross-linguistic patterns.

<sup>7</sup>“Der Herr wird werden indem er eins ist und sein Name indem er eins ist” (Till 1942: 51).



Furthermore, Haspelmath (2023)’s definitions of an affix as a “bound form that is not a root and that cannot occur alone” and a clitic as a “bound form that is not an affix but still depends on another form” provide a useful framework for distinguishing between these elements in Coptic. Applying these definitions to the various segmentation types, we find that Layton (2011)’s bound group (Type 3) strikes a balance between representing the prosodic unity of Coptic words and capturing the grammatical independence of clitics.

While other segmentation types, such as Type 1 and Type 2, may offer alternative perspectives on word boundaries, we maintain that Layton's bound group (Type 3) provides the most typologically sound basis for analysing Coptic morpho-syntax and calculating the Morpheme-per-Word (M/W henceforth) ratio, as discussed further in Section 8.

### 3 Punctuation and diacritics

On late antique manuscripts, Coptic was originally written in *scriptio continua*, i.e., without spaces. In determining the segmentation of words in Coptic texts, we may rely on certain punctuation marks and diacritical signs that suggest boundaries. For instance, upper-dots (UD henceforth) typically signal the termination of sentences, clauses, or phrases, as exemplified in (9 and 10).

- (9) Use of UD (“|” indicates a line break)

· πεχεπαγγελ|λосπ̑χοεις|ναϙ ·  
UD *peče-p-aggelos-m-p-čoeis-na*<sup>f</sup> UD  
UD said.PNOM-DEF.SG.M-angel-DIR-DEF.SG.M-lord-DAT-3SG.M UD  
· ჯემპრ|ჯოთეზახარიაс ·  
*če-mpr-r-hote-zak<sup>h</sup>arias* UD  
COMP-PROH-do.PNOM-fear-Zachariah UD  
· |ჯეაუსôtმე|пексопċ ·  
*če-a-u-sôtm-epe-k-sops* UD  
COMP-PST-3PL-listen-DIR-POSS.SG.M-2SG.M-prayer UD

‘UD The angel of the Lord said to him, UD “Do not fear, Zachariah, UD because your prayer was heard UD [...]’

(P. Palau Rib. inv. 181 = NT Luke 1:13)

- (10) Diverse use of UD

· αγγελος · ετουααβ · ηνσομ · πωτ  
UD *aggelos* UD *et-ouaab* UD *hn-com* UD *p-iôt*  
UD angel UD REL-be holy.STA UD in.PNOM-power UD DEF.SG.M-father

·	ḡñ śom	·	noḡḡre	·	ḡñśom	·	peḡneḡma
UD	<i>hn-com</i>	UD	<i>n-šēre</i>	UD	<i>hn-com</i>	UD	<i>pe-pneuma</i>
UD	in.PNOM-power	UD	DEF.PL-son	UD	in.PNOM-power	UD	DEF.SG.M-pirit
·	etouaav	·	ḡñśom	·	neḡaggelos	·	
UD	<i>et-ouaab</i>	UD	<i>hn-com</i>	UD	<i>ne-f-aggelos</i>	UD	
UD	REL-be_holy.STA	UD	in.PNOM-power	UD	DEF.PL-3SG.M-angel	UD	
tḡrou	·						
<i>tēr-ou</i>	UD						
all-3PL	UD						

‘the holy angel in power, the father in power, the sons in power, the holy spirit in power, all his angels’

(P. Mac. Inv. 375, p. 11, l.4–8)

In Quecke (1984), a distinction is made between different placements of upper dots: Very-high upper dots are typically found at the boundaries of sentences, while standard-height upper dots frequently occur at clause boundaries. This is particularly noticeable preceding or following the complementizer particle  $\alpha\epsilon\ \check{\epsilon}$ , and before  $\alpha\gamma\omega\ au\hat{o}$  ‘and’. Furthermore, in certain manuscripts, upper dots are also utilised to delineate smaller linguistic units.

Furthermore, apostrophe-like markers (ALM henceforth) are prevalent in Coptic manuscripts, serving as indicators of micro-level textual divisions. These markers are particularly evident in manuscripts associated with Shenoute.

Shenoute is a prominent figure in Coptic literature and monasticism. Shenoute, also known as Shenoute of Atripe, was a 5th-century Coptic abbot who led the White Monastery in Upper Egypt. He is renowned for his extensive corpus of writings, which significantly influenced Coptic literature and provide valuable insights into the language and religious practices of the time. The consistent use of apostrophe-like markers in Shenoute's manuscripts suggests a systematic approach to text organisation and punctuation, setting a standard that may have influenced other Coptic writers.

By highlighting the prevalence of these markers in Shenoute's works, we can better understand their role in structuring Coptic texts and their potential impact on the wider Coptic literary tradition.

(11) Use of ALM

ἀνναγάρετα|γαπή      ἡγάγῃ      ἡγῇ|τῆεζοῦν  
*a-n-nau-gar-e-t-agapê* ALM *n-hah* ALM *nhê-tn-e-houn* ALM  
 PST-1PL-see.ABS      PRT      DIR-DEF.SG.F-love ALM LK-many ALM

ⲉⲡⲏⲟⲩⲧⲧⲉ

*e-p-noute*

IN.PNOM-2PL DIR-inside

‘but we saw the abundant love in you toward God’

(Vienna K 925, l. 16–19)

In addition to the markers discussed previously, Coptic manuscripts feature a range of other punctuation marks that are less common. These include the colon (:), the *diplē* sign (Ⲅ),<sup>8</sup> the period (.), and the comma (,), among others.

Generally, such punctuation is employed to denote boundaries at the more macro-level compared to the upper dots and apostrophe-like markers. These signs are instrumental in demarcating larger textual units, such as sentences and paragraphs. Among Coptic diacritical marks, superlinear strokes are particularly intriguing among the various punctuation marks, offering valuable insights into the word division in Coptic manuscripts. J. Martin Plumley has insightfully characterised the features of superlinear strokes as follows:

The unbroken succession of consonants in Coptic MSS makes word division a matter of extreme difficulty. What is to be made of such a group as ⲛⲧⲛⲧⲙⲛⲧⲉⲱⲧ [ⲛⲧⲛⲧⲙⲛⲧⲉⲱⲧ], in which only one vowel is clearly discernable? How is such a succession of consonants to be divided into syllables? Fortunately the writers of Sahidic MSS were aware of this difficulty, and invented a simple method to aid the reader: the Superlinear Stroke, or Syllable Marker. By placing a stroke over the letters thus ⲃ̄ [b̄], ⲗ̄ [l̄], ⲙ̄ [m̄], ⲛ̄ [n̄] and ⲡ̄ [r̄], and less frequently ⲕ̄ [k̄], ⲥ̄ [s̄], ⲙ̄̄ [š̄], ⲙ̄̄̄ [m̄̄], and ⲕ̄̄ [h̄], the correct division into syllables is indicated. Thus in good MSS, ⲛⲧⲛⲧⲙⲛⲧⲉⲱⲧ [ⲛⲧⲛⲧⲙⲛⲧⲉⲱⲧ] would appear as ⲛ̄ⲧ̄ⲛ̄ⲧ̄ⲙ̄̄ⲛ̄ⲧⲉⲱⲧ [n̄t̄n̄t̄m̄̄n̄t̄eiōt], indicating the syllabic division ⲛ̄.ⲧ̄ⲛ̄.ⲧ̄ⲙ̄̄ⲛ̄.ⲧⲉⲱⲧ [n̄.t̄n̄.t̄m̄̄n̄t̄.eiōt].

Thus, superlinear strokes can divide syllable units but not word units. Summarising the above, punctuation marks mainly divide clauses and sentences, and superlinear strokes are hints at how to divide syllables. They can be clues for us to divide words in Coptic texts. However, they are incomplete for that purpose and no marks seem to have been designed for marking word boundaries consistently.

<sup>8</sup>For the history and functions of the *diplē* sign, see Miyagawa (2022: 84–89).

## 4 Clitics

The role of clitics in syntactic structure is a subject of keen interest to classicists. Specifically, Wackernagel clitics or second-position (P2 henceforth) discourse clitics, such as  $\Delta\epsilon$  *de* ‘but, and, on the other hand,’  $\gamma\alpha\pi$  *gar* ‘for, because,’ and  $\sigma\epsilon$  *ce* ‘then, therefore, so’ in Coptic, invariably occupy the second position in a sentence.<sup>9</sup> This consistent placement not only marks the boundary between the first and second syntactic elements but also provides insight into the sentence’s prosodic structure. However, these clitics do not necessarily correspond to a single word; they may also attach to phrases, indicating the boundary between the phrase and the following syntactic element. While clitics are dependent on their host words or phrases for pronunciation, they still function as separate grammatical units within the larger syntactic structure.

While Layton’s concept of the bound group is a well-accepted prosodic construct characterised by a single stress point, it is not synonymous with the linguistic definition of a word. In linguistics, a word is typically defined as the smallest unit of the language that can stand alone and convey a complete meaning. It is a grammatical unit that can be moved around within a sentence and that can take inflectional or derivational morphology (Aronoff & Fudeman 2023). This definition emphasises the syntactic and semantic independence of a word, as well as its potential for morphological modification.

By contrast, Layton’s bound group is primarily concerned with prosodic unity, focusing on the stress pattern within a group of morphemes. While a bound group may often correspond to a single word, it can also encompass clitics or other elements that are prosodically dependent but grammatically distinct. Therefore, it is essential to differentiate between the prosodic concept of the bound group and the linguistic definition of a word when analysing the structure of Coptic.

Cross-linguistic evidence shows that certain words lack inherent stress and are referred to as clitics. In the case of Modern Japanese, a language like Coptic that traditionally eschews spaces in writing, there has been considerable debate over the categorisation of adpositions, particles, and converbs as either words or affixes. Contemporary linguistic research, following the trajectory of Arnold Zwicky’s influential work (Zwicky 1977, 1985, Zwicky & Pullum 1983), leans towards classifying these elements as clitics rather than affixes. This perspective is supported by studies that focus on the distinction between clitics and affixes,

---

<sup>9</sup>It is interesting that Coptic P2 means the position after the first phonological word, while Ancient Greek P2 is the position after the first morphological word.

such as Anderson (2005), Spencer & Luís (2012), and Haspelmath (2015). These researchers argue that clitics exhibit greater syntactic flexibility and independence compared to affixes, which are more tightly bound to their host words.

For instance, Anderson (2005) emphasises the syntactic independence of clitics, noting that they can attach to various parts of speech and are not restricted to a specific morphological host. Spencer & Luís (2012) further explore the differences between clitics and affixes, highlighting the role of clitics in marking grammatical relations and their ability to scope over larger syntactic constituents. Haspelmath (2015) provides a cross-linguistic perspective, demonstrating the wide range of functions that clitics can serve across different languages.

In the Japanese example (12), linguistic elements such as postpositions, verbal particles, copulas, auxiliary verbs, topic markers, and complementizers exhibit prosodic adherence to their preceding elements.

(12) Clitics in Japanese

では	みなさんは、	そう	いう	ふうに	川だと
<i>dewa</i>	<i>mina-san=wa,</i>	<i>soo</i>	<i>yu-u</i>	<i>fuu-ni</i>	<i>kawa=da=to</i>
then	all-HON=TOP	thus	say-ADN	manner-ADVL	river=COP=COMP
云われたり、	乳の	流れた	あとだと		
<i>iw-are-tari,</i>	<i>chichi=no</i>	<i>nagare-ta</i>	<i>ato=da=to</i>		
say-PASS-CONV	milk=GEN	flow-PST.ADN	trace=COP=COMP		
云われたりしていた		この	ぼんやりと		
<i>iw-are-tari=shi-te=i-ta</i>		<i>kono</i>	<i>bon'yari=to</i>		
say-PASS-CONV=do-CONV=PROG-PST.ADN		this.ADN	vague=ADVZ		
白い	ものが	ほんとうは	何か	ご承知ですか	
<i>shiro-i</i>	<i>mono=ga</i>	<i>hontoo=wa</i>	<i>nani=ka</i>	<i>go-shoochi=desu=ka</i>	
white-ADN	thing=NOM	real=TOP	what==Q	HON-knowing=COP.HON=Q	
				(Japonic)	

‘So, do you know what this vague white thing that was said to be a river or the remains of flowing milk really is?’

Despite this prosodic bond, the diverse potential for these elements to attach to various hosts categorises them as clitics. A clitic functions as a grammatical word, a unit that operates independently within syntactic structures. However, from a phonological or prosodic perspective, it does not constitute a standalone word. Dixon & Aikhenvald (2002: 25) articulate this concept by distinguishing between a clitic’s prosodic dependency and its grammatical autonomy.

Therefore, although a clitic may exhibit phonological characteristics akin to an affix, it is, in essence, a separate word. Modern linguistic theory analyses morphs on two distinct planes: the morpho-syntax and the phonology (prosody). In notation, the juncture between two clitics or between a clitic and its host word is denoted by an equal sign (=), as systematised in the *Leipzig Glossing Rules* developed by the *Max Planck Institute for Evolutionary Anthropology*. Haspelmath & Sims (2010) have formulated six robust criteria to differentiate between affixes and clitics, as shown in Table (1).

Table 1: Criteria to distinguish between clitics and affixes (Haspelmath & Sims 2010: 202)

Clitics	Affixes
freedom of host selection	no freedom of stem selection possible
freedom of movement	no freedom of movement
less prosodically integrated	more prosodically integrated
may be outside the domain of a phonological rule	within the domain of a phonological rule
do not trigger/ undergo morphophonological or suppletive alternations	may trigger/ undergo morphophonological or suppletive alternations
clitic-host combinations...	
do not have idiosyncratic meanings	do not have arbitrary gaps
affix-base combinations...	
may have idiosyncratic meanings	may have arbitrary gaps

Within the spectrum of criteria for distinguishing clitics, the principle of Freedom of Host Selection (FHS henceforth) stands out as a particularly definitive factor in determining a morph’s morpho-syntactic independence. For instance, the English abbreviated form of ‘is,’ which can attach to an entire noun phrase, exemplifies a clitic that exhibits FHS, thereby demonstrating its syntactic autonomy from any single host word, see (13).

- (13) enclitic =’s (is) in English
- a. The Coptic parchment’s beautiful.

- b. The Coptic parchment I saw's beautiful.
- c. The Coptic parchment I saw yesterday's beautiful.

The morpheme =’s, which can affix to nouns, verbs, and adverbs. It demonstrates the property of Freedom of Host Selection (FHS) by its ability to attach to various syntactic categories: nouns like ‘parchment,’ verbs in their past-tense form like ‘saw,’ and even adverbs like ‘yesterday.’ This versatility confirms that =’s functions as a clitic, as it maintains its syntactic role across different host words. Furthermore, applying the principle of FHS to Coptic, the conjunction or complementizer  $\chi\epsilon$  *čē-* can be identified as a clitic due to its ability to attach freely to different syntactic units, indicating its morpho-syntactic independence.

- Negative particle:  $\chi\epsilon\mu\pi\epsilon\mu\epsilon\lambda\lambda$   $\text{COOYN AN}$  *čē-m-p-hmhal sooun an* (COMP-NEG-DEF.SG.M-slave know.ABS NEG) ‘that the slave doesn’t know’
- Interrogative pronoun:  $\chi\epsilon\text{OY}$   $\pi\epsilon$   $\pi\gamma\alpha\mu\text{OC}$  *čē-ou pe p-gamos* (COMP-what COP.SG.M DEF.SG.M-marriage) ‘what is the honorable marriage’ (Abraham.YA525-530 in Coptic SCRIPTORIUM)
- Verb:  $\chi\epsilon\text{NOY}\chi\epsilon$   $\epsilon\text{BO}\lambda$   $\eta\tau\epsilon\iota\mu\epsilon\lambda\lambda$  *čē-nouče e-bol n-tei-hmhal* (COMP-throw.ABS DIR.ABS-outside ACC-this.SG.F-slave) ‘cast out this slave’ (Abraham.YA518-520 in Coptic SCRIPTORIUM)
- Demonstrative pronoun:  $\chi\epsilon\text{PAI}$   $\epsilon\tau\mu\mu\alpha\gamma$   $\epsilon\tau\epsilon\iota\varsigma\mu\alpha\eta\lambda$  *čē-pai et-m-mau ete-ismaêl* (COMP-DEM.SG.M-REL-LOC-there-REL-Ishmael) ‘that that one who is Ishmael [...]’ (Abraham.YA518-520 in Coptic SCRIPTORIUM)
- Auxiliary:  $\chi\epsilon\alpha\gamma\text{COTI}$   $\varsigma\eta\alpha\gamma$   $\epsilon\tau\pi\epsilon\gamma\omega\pi\epsilon$   $\eta\alpha\beta\text{P}H\eta$   $\alpha\gamma\omega$   $\eta\chi\eta\text{P}\alpha$  *čē-a-u-sotp snau e-tre-u-šôpe n-acrên auô n-khêra* (COMP-PST-3PL two-DIR-CAUS -3PL-be LOC-barren and LOC-widow) ‘since they chose to be barren women and widows?’ (Abraham.YA525-530 in Coptic SCRIPTORIUM)
- Article:  $\chi\epsilon\pi\iota\text{OC}$   $\eta\alpha\text{P}\mu\epsilon\lambda\lambda$   $\mu\pi\kappa\text{OYI}$  *čē-p-noc na-r-hmhal m-p-koui* (COMP-DEF.SG.M-great FUT-do.PNOM-servant ACC-DEF.SG.M-lesser) ‘the great will serve the lesser’ (Abraham.YA518-520 in Coptic SCRIPTORIUM)
- Noun:  $\chi\epsilon\iota\alpha\kappa\omega\beta$   $\alpha\iota\mu\epsilon\text{PIT}\gamma$  *čē-iakôb a-i-merit-f* (COMP-Jacob PST-1SG-love.PPRO -3SG.M) ‘as for Jacob, I loved him’ (Abraham.YA518-520 in Coptic SCRIPTORIUM)

- Personal pronoun  $\chi\epsilon\tau\omicron\tau\alpha\pi\epsilon\pi\epsilon\varsigma\omega\tau\eta\rho$  *če-ntof pe pe-n-sôtêr* (COMP-2SG.M COP.SG.M POSS.SG.M- 1PL-savior) ‘because he is our savior’ (Abraham.YA535-540 in Coptic SCRIPTORIUM)
- Conjunction  $\chi\epsilon\tau\omicron\tau\alpha\pi$  *ere-p-col na-šače* (COMP- whenever CIRC-DEF.SG.M- liar FUT-speak) ‘whenever a liar speaks’ (Abraham.YA54-50 in Coptic SCRIPTORIUM)

The ability of the Coptic complementizer conjunction  $\chi\epsilon$  *če-* to attach to a wide range of parts of speech exemplifies its substantial FHS, a characteristic that classifies it as a clitic rather than an affix. Similarly, the relative marker  $\epsilon\tau$  *et-* demonstrates a broad FHS, as it can be found preceding various grammatical elements in a sentence, further supporting its identification as a clitic.

- Verb:  $\epsilon\tau\chi\eta\zeta$  *et-sêh* (REL-write.STA) ‘that is written’ (Abraham.YA535-540 in Coptic SCRIPTORIUM)
- Preposition:  $\epsilon\tau\chi\iota\lambda\omega\pi$  *et-hiçô-n* (REL-over-1PL) ‘who is over us’ (Abraham.YA535-540 in Coptic SCRIPTORIUM)
- Adverb:  $\epsilon\tau\mu\mu\alpha\gamma$  *et-mmau* (REL-there) ‘who is there’ (Abraham.YA518-520 in Coptic SCRIPTORIUM)

Based on the criterion of FHS, the Coptic relative marker  $\epsilon\tau$  *et-* is categorised as a clitic. This is due to its syntactic flexibility in attaching to various grammatical constituents, distinguishing it from an affix, which typically has a more fixed position.

Similarly, Coptic articles display characteristics that align with the behavior of clitics. The definite articles in Coptic, coding for gender and number, include the masculine singular  $\pi(\epsilon)$ - *p(e)-*, the feminine singular  $\tau(\epsilon)$ - *t(e)-*, and the plural  $\nu(\epsilon)$ - *n(e)-*. The indefinite articles, coding for number, include the singular  $\omicron\gamma$ - *ou-* and the plural forms  $\chi\epsilon\pi$ - *hen-* /  $\chi\eta$ - *hn-*. The variation in form and the ability to attach to different noun phrases suggest that Coptic articles may also be considered clitics.

- Relative marker:  $\pi\epsilon\tau\eta\lambda\alpha\upsilon\omicron\gamma$  *p-et-nanou-f* (DEF.SG.M- REL-be\_good-3SG.M) ‘the good one’ (*Letter to Aphthonia* in Coptic SCRIPTORIUM)
- Noun:  $\nu\epsilon\varsigma\eta\eta\gamma$  *ne-snêu* (DEF.PL-brother.PL) ‘the brothers’ (*Letter to Aphthonia* in Coptic SCRIPTORIUM)



- Definite article: ππεθουγ *p-p-et-<sup>h</sup>ou* (DEF.SG.M- DEF.SG.M- REL-be\_bad.STA) ‘the bad one’ (*Letter to Aphthonia* in Coptic SCRIPTORIUM)
- Causative auxiliary: πτρετκαμα *p-tre-tn-ka-ma* (DEF.SG.M- CAUS - 2PL-leave. PNOM-place) ‘you leaving’ (*Letter to Aphthonia* in Coptic SCRIPTORIUM)
- Adverb: γενεβολ ηλαος ηαρχαιος *hen-ebol n-laos n-ark<sup>h</sup>aaios* (INDEF.PL-out LOC.PNOM-people LK.PNOM-ancient) ‘those who are from the ancient people’ (Shenoute, *Abraham Our Father* (Abraham\_YA) 547–50 in Coptic SCRIPTORIUM)

Consequently, the significant FHS exhibited by articles in Coptic positions them as clitics, not affixes. Their ability to freely associate with various noun phrases, irrespective of the latter’s syntactic role, underscores their clitic nature in the language structure.

## 5 Prenominal state

In Coptic, various words representing different parts of speech have a “state” (see Figure 2). There are three states: the absolute, the prenominal, and the prepronominal states. The absolute state always has an accent and can be a free form. The prenominal state has no accent, and its vowel is often weakened to a schwa or a zero vowel. Only nominals or noun phrases can stand after a word in a prenominal state. Various parts of speech in Coptic have prenominal states, such as prepositions, transitive verb infinitives and imperatives, body-part nouns, auxiliary verbs, and so-called “converters”.

State↓ Part of speech →	Noun	Intransitive verb (no direct obj.)	Transitive verb (direct obj.)	Participium conjunctum	Preposition	Auxiliary verb (conjugational base)	Converter
absolute form	accented	accented	accented	–	–	–	–
prenominal form	unaccented	–	unaccented	unaccented	unaccented	unaccented	unaccented
prepronominal form	accented	–	accented	–	accented	unaccented	unaccented

Figure 2: Different state forms according to parts of speech (Miyagawa 2023: 566)

If the word is a transitive verb or a preposition, the following nominal or noun phrase is the complement of the word in the prenominal state. If the transitive

verb is in the absolute state, the object marker *n-* *n-* is needed. For example, in *сѣтпоуѣ setp-ou-rôme* (choose.PNOM-INDEF.SG-man) ‘choose a man’, *сѣтп setp-* is in the prenominal state; but in *сѣтп ноуѣ sôtp n-ou-rôme* (choose.ABS OBJ-INDEF.SG-man) ‘choose a man’, *сѣтп sôtp* is the absolute state. The unaccented *ε e* is pronounced as an unaccented schwa, but *ω ô* in *сѣтп sôtp* has an accent since *ω ô* is always accented in the Sahidic dialect of Coptic.

*сѣтпоуѣ setp-ou-rôme* is one phonological word but *сѣтп ноуѣ sôtp n-ou-rôme* is two phonological words. We can also consider that the absolute state marks its complement with the complement/object marker before the complement (dependent marking). However, the prenominal state marks its complement with the vowel weakening on the verb (head marking). The prenominal-state verbs can take a noun with a definite or indefinite marker as their complement, such as *сѣтпоуѣ setp-ou-rôme*. Here, *оу ou-* is an indefinite article. The prepositions, auxiliary verbs, and converters only have prenominal and prenominal states, but no absolute states, whereas transitive verbs can appear in all three states.

In Coptic, articles exhibit a degree of syntactic flexibility that is characteristic of clitics. They can attach to a variety of syntactic elements, including prepositions, verbs, nouns, and even other articles. This behavior suggests that Coptic articles function as clitics rather than affixes.

However, the ability of adjectives to intervene between articles and nouns in Coptic raises questions about the status of articles as clitics or affixes. In some cases, adjectival elements can appear between the article and the noun, as in the construction article-adjective-noun (ART-ADJ-N). This flexibility in word order indicates that Coptic articles do not form a tight morphological unit with the nouns they modify, supporting their analysis as clitics.

It is also important to note that the behavior of adjectives in Coptic is complex and varies depending on the type of adjective and the specific construction. Some adjectives may follow the noun in an article-noun-adjective (ART-N-ADJ) order, while others may precede the noun. The variability in adjective placement suggests that the relationship between articles, adjectives, and nouns in Coptic requires further investigation to fully understand the nature of the articles as clitics or affixes.

- (14) *п-нѣ н-ноѣ p-êi n-noc* (DEF.SG.M- house LK-big) / *п-ноѣ н-нѣ p-noc n-êi* (DEF.SG.M- big LK-house) ‘the big house’
- (15) *п-ѡнѣ ѡнм p-šêre šêm* (DEF.SG.M-boy little) but \**п-ѡнм ѡнѣ p-šêm šêre* (DEF.SG.M-little boy) ‘the little boy’

While Coptic adjectives can indeed function as nouns, their behavior in the article-adjective-noun construction is complex. An attributive preposition is typically inserted before the noun in this construction, and both the adjective and the noun can be interchangeable, as seen in (14). Consequently, the clarity of Dryer's explanation that Coptic adjectives intervening between articles and nouns justify classifying articles as clitics becomes somewhat questionable.

Despite this, the Coptic definite article exhibits the ability to attach to a variety of syntactic elements, including prepositions, verbs, nouns, and even to another definite article, suggesting a degree of FHS. This flexibility extends to indefinite articles as well, supporting the view that Coptic articles function as clitics.

If we accept that the definite article behaves as a clitic, it follows that prenominal prepositions should also be considered clitics. This leads us to two possible interpretations of  $\text{hm-p-ran}$  (in.PNOM-DEF.SG.M-name) 'in the name': either as three separate words  $hm=p=ran$  or as two words  $hm-p=ran$ .

Let us consider the latter interpretation, a head-marking solution, where  $\text{hm-p}$  is treated as a single unit attached to the noun  $\text{ran}$ . If this were the case, we would expect  $\text{hm-}$  to be a prefix that can attach directly to the noun, allowing for the form  $\text{hm-ran}$ . However, this creates a contradiction, as  $\text{p}$  (the definite article) is obligatory and cannot be omitted. The fact that  $\text{hm-ran}$  is not a viable form suggests that  $\text{p}$  is not merely a host for the prefix  $hm-$ , but rather an independent element.

Therefore, we must discard the head-marking solution and conclude that  $\text{hm=p=ran}$  is the more logical segmentation. This analysis indicates that prenominal prepositions, like the definite article, are indeed clitics that attach to the noun phrase as separate elements, rather than prefixes that attach directly to the noun itself.

However, the categorisation of some prenominal prepositions as clitics becomes challenging when they appear before bare nouns and exhibit high lexicalisation, such as  $\text{e-bol}$  (DIR .PNOM-outside) meaning 'outwardly, away,' or  $\text{n-toot-ou}$  (LOC.PNOM-hand.PPRO-3PL) meaning 'at them.' Similarly, prenominal verbs that precede articles display characteristics of clitics.

The lexicalised patterns that emerge from the combination of prenominal prepositions and nouns in Coptic often result in single words that convey meanings beyond that which the Compositionality Principle would predict. In these instances, the prenominal prepositions function as grammatical or functional morphemes, which is consistent with their status as clitics. As clitics, they are expected to serve grammatical or functional roles within the larger syntactic structure. The highly lexicalised combinations of prenominal prepositions and nouns in Coptic demonstrate the close relationship between these elements, with

the prepositions contributing to the overall meaning of the construction in a way that is characteristic of clitics.

However, it is important to note that the degree of lexicalisation and the specific functional roles played by prenominal prepositions may vary across different constructions. While some combinations may exhibit a high degree of lexicalisation, others may retain a more compositional meaning. The status of prenominal prepositions as clitics does not necessarily imply a complete loss of their original semantic content, but rather highlights their integration into the larger syntactic and semantic unit.

## 6 Prenominal state of transitive verbs

In Coptic Egyptian, verbs have long been recognised as having distinct morphological states, a fact well-established in the literature (e.g., Stern 1880, Steindorff 1951, Polotsky 1960). These states are characterised by differences in their morphological forms and syntactic behavior. Layton's (2011) framework introduces a new terminology to describe these well-known categories, providing a systematic way of referring to the different verb forms.

According to Layton's terminology, the main morphological states of Coptic verbs are the absolute state (ABS), the prenominal state (PNOM), the prepronominal state (PPRO), and the stative form (STA). Each of these states has distinct morphological and syntactic properties that govern their use in Coptic sentences. Additionally, a select number of verbs possess a unique imperative form. For example, the verb *ⲥⲟⲩⲧⲏ* *sōtp* 'choose' has *ⲥⲟⲩⲧⲏ* *sōtp* in the absolute state, *ⲥⲉⲩⲧⲏ* *setp* in the prenominal state, *ⲥⲟⲩⲧⲏ* *sōtp* in the prepronominal state, and *ⲥⲟⲩⲧⲏ* *sōtp* 'be chosen' as the stative form.

The division is primarily based on the position of the verb relative to the subject and object—with a standard order for verbs being subject-verb-object (SVO) and an alternate verb-subject-object (VSO) order for verboids—as well as the application or omission of tense-aspect-mood (TAM) markers, see Figure 3.

Historically, these states represent morphologically distinct forms of the verbal infinitive that are determined by what follows the verb: The prenominal state occurs before a nominal or a noun phrase with the indefinite or definite marker, such as *ⲥⲉⲩⲧⲏⲡⲟⲩⲣⲟⲙⲉ* *setp-ou-rōme* (choose.PNOM-INDEF.SG -man) 'choose a man', the prepronominal state before a personal pronominal suffix, such as *ⲥⲟⲩⲧⲏⲉ* *sōtp-f* (choose.PPRO-3SG.M) 'choose him', and the absolute state is used in other contexts, such as *ⲥⲟⲩⲧⲏ ⲡⲟⲩⲣⲟⲙⲉ* *sōtp n-ou-rōme* (choose.ABS OBJ.PNOM-INDEF.SG-man) 'choose a man' or *ⲥⲟⲩⲧⲏ ⲙⲙⲟⲩ* *sōtp mmo-f* (choose.ABS OBJ-3SG.M) 'choose him'.

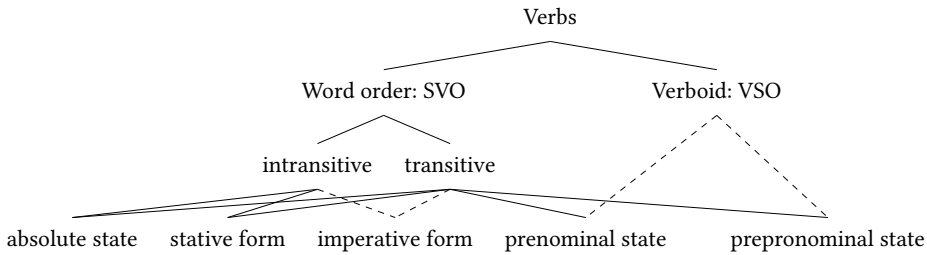


Figure 3: Morphological change of Coptic verbs

Transitive verbs in the imperative mood in Coptic may also appear in the absolute, prenominal, and prepronominal states. A limited number of the verbs have morphologically distinct imperative form, while the other verbs have no special imperative form, but they convey imperative meaning by having no subject. The latter is similar to the English imperative (e.g. ‘Do it!’). Moreover, there is an exceptional verb, namely *ei* ‘come’ which exhibits a fossilised conjugation pattern, which varies according to the subject’s gender and number: Masculine singular imperative *amou*; feminine singular imperative *amê*; plural imperative *amêin*, *amêeitn*, *amôine* (see Crum 1939: 7b).

Martin Haspelmath (2014) analyses the morpho-syntax of Coptic transitive verbs from a typological perspective, describing the absolute state as a free form. He characterises the prenominal state as bound when it precedes a full noun phrase (NP henceforth), and the prepronominal state as bound before a pronominal element.

In the realm of morpho-phonology, the historical change of prenominal states from their absolute counterparts can be categorised into several patterns: 1) weakening or loss of vowels (Types I, II, III, IV, V, VII), 2) weakening of vowels accompanied by the addition of a *τ t* at the end (VII), and 3) no change in form, as with the verb *fī* ‘take’, see further Table 2. Layton (2011: 152) has delineated seven types of regular verbal morphological alterations in Coptic as well as numerous irregular modifications.

The morphological shifts that Coptic transitive verbs undergo, particularly the reduction in vowel strength and their syntactic behavior with following nominal phrases, play a pivotal role in shaping the language’s grammatical framework. While the imperative forms of transitive verbs can manifest in the absolute, prenominal, and prepronominal states, the primary focus of this study will be on the non-imperative forms, as they are more central to the discussion of wordhood and morphological synthesis.

Table 2: Types of morphological changes of Sahidic Coptic verbs

Type	Infinitive			Meaning	Stative	Meaning
	ABS	PNOM	PPRO			
I	ϣⲱⲧⲡ <i>sôtp</i>	ϣⲉⲧⲡ <i>setp-</i>	ϣⲟⲧⲡ <i>sotp-</i>	‘(to) choose’	ϣⲟⲧⲡ <i>sotp</i>	‘be chosen’
II	ⲕⲱⲧ <i>kôt</i>	ⲕⲉⲧ <i>ket-</i>	ⲕⲟⲧ <i>kot-</i>	‘(to) build’	ⲕⲏⲧ	‘be built’
III	ⲡⲱⲟⲛⲉ <i>pôône</i>	ⲡⲉⲛⲉ <i>pene-</i>	ⲡⲟⲟⲛⲉ <i>poone-</i>	‘(to) change’	ⲡⲟⲟⲛⲉ <i>poone</i>	‘be changed’
IV	ϣⲟⲕϣ <i>sols̩</i>	ϣⲟⲕ <i>s̩sl-</i>	ϣⲟⲕⲱⲗ <i>s̩sôl-</i>	‘(to) comfort’	ϣⲟⲕⲱⲗ / ϣⲟⲕⲱⲧ <i>s̩sôl / s̩solt</i>	‘be comforted’
V	ⲧⲁⲕⲟ <i>tako</i>	ⲧⲁⲕⲉ <i>take-</i>	ⲧⲁⲕⲟ <i>tako-</i>	‘(to) destroy’	ⲧⲁⲕⲏⲭ / ⲧⲁⲕⲏⲭⲧ <i>takêu / takêut</i>	‘be destroyed’
VI	ⲭⲟⲗⲟⲥ <i>hloc</i>			‘(to) become sweet’	ⲭⲟⲗⲟⲥ <i>hloc</i>	‘be sweet’
VII	ĉise <i>ĉise</i>	ĉest <i>ĉest-</i>	ĉast <i>ĉast-</i>	‘(to) exalt’	ĉoce <i>ĉose</i>	‘be high’
Irreg. 1	ⲉⲓⲣⲉ <i>eire</i>	ⲣ <i>r-</i>	ⲁⲁ <i>aa-</i>	‘(to) do’	ⲟ <i>o</i>	‘be being’
Irreg. 2	ⲉⲓ <i>ei</i>			‘(to) come’	ⲛⲏⲭ <i>nêu</i>	‘be coming’

To illustrate the application of the absolute, prenominal, and prepronominal states, along with the stative form, this chapter uses the verb ⲕⲱ *kô* ‘to place/leave’ which is a prevalent example of Type-I verbs in Coptic. The analysis will commence with its dictionary form or the absolute state, providing a foundation for understanding its various morphological states within sentence structures.

In the absolute state within Coptic grammar, a direct object is indicated by the presence of an objective marker, such as *n-* ‘of, to, (or object marker)’ or *e-* ‘to, for’, which is prefixed to the noun or noun phrase. For instance, the direct object ⲛⲟⲩⲉⲗⲉⲁⲣⲁ *n-ou-eksedra* includes the object marker *n-* *n-*, see (16).

- (16) Absolute state: ⲕⲱ *kô* ‘to place / leave’

ⲡⲉⲛⲛⲁⲓⲟⲥ	ⲁⲉ	ⲁⲡⲁ	ⲃⲓⲕⲧⲱⲣ	ⲁⲑⲕⲱ	ⲛⲁⲑ
<i>p-gennaios</i>	<i>de</i>	<i>apa</i>	<i>biktôr</i>	<i>a-f-kô</i>	<i>na-f</i>
DEF.M.SG-noble	PRT	Apa	Victor	PST-3SG.M-place.ABS	DIR-3SG.M

ⲛⲟⲩⲉⲗⲉⲁⲣⲁ	ⲛ̄ⲉⲟⲩⲛ	ⲙⲡⲉⲩⲩⲏ
<i>n-ou-ek'edra</i>	<i>n-houn</i>	<i>m-pe-f-êi</i>
ACC-INDEF.SG-chamber_small	LOC-inside	LOC-POSS.SG.M-3SG.M-house
‘And the noble Victor made for himself a small chamber in his house’		
(Martyrdom of Victor, 6.10 in Coptic SCRIPTORIUM)		

In Coptic, prosodic rules dictate that the accent falls on the vowels  $\omega$   $\hat{o}$ ,  $\eta$   $\hat{e}$ ,  $\circ$   $\circ$ , or any duplicated vowel letter, typically occurring on the ultima (last syllable) or penultima (second-to-last syllable). In the example  $\alpha\kappa\omega$  *afkô*, the accent is on the ultima, which is the vowel  $\omega$   $\hat{o}$ . Other vowels such as  $\alpha$  *a*,  $\epsilon\iota$  *ei*,  $\iota$  *i*,  $\omicron\upsilon$  *ou*, or  $\epsilon$  *e* may be accented or unaccented and are not restricted to the ultima or penultima positions. When  $\omega$   $\hat{o}$ ,  $\eta$   $\hat{e}$ ,  $\circ$   $\circ$ , or a duplicated vowel is present in the ultima or penultima, they automatically receive the accent. However, if the vowel is  $\alpha$  *a*,  $\epsilon\iota$  *ei*,  $\iota$  *i*,  $\omicron\upsilon$  *ou*,  $\epsilon$  *e*, the accent is placed on the penultima if the penultima has the accented vowel letter  $\omega$   $\hat{o}$ ,  $\circ$   $\circ$ ,  $\eta$   $\hat{e}$ , or vowel letter doubling, or on the ultima otherwise.

The prenominal state loses the accent, compared with its absolute state. The vowels in this state are always  $\alpha$  *a* or  $\epsilon$  *e*, the semi-vowels  $\epsilon\iota$  *ei*,  $\iota$  *i*, or  $\omicron\upsilon$  *ou*, or vowels are absent altogether. Verbs in the prenominal state directly precede a noun or a noun phrase with an article, with no intervening elements. The prosodic emphasis, or accent nucleus, for the prenominal state is consistently on the subsequent noun.

- (17) Prenominal state:  $\kappa\alpha$ - *ka*- ‘to place / leave’

ⲁⲩⲱ	ⲛⲉ̅ⲙ̅ⲡⲟⲩⲕⲁⲧⲟⲟⲩⲱ	ⲉⲃⲟⲗ	ⲡⲉ
<i>auô</i>	<i>ne-mp-ou-ka-toot-ou</i>	<i>e-bol</i>	<i>pe</i>
and	PRET-PST.NEG-3PL-place.PNOM-hand.PPRO-3PL	DAT-outside	COP.SG.M
ⲛ̅ⲟⲩ̅ⲛ̅ⲉⲟⲩ̅	ⲧⲏⲣⲟⲩ	ⲉⲧ̅ⲛ̅ⲙ̅ⲡⲏⲩⲩⲉ	
<i>nci-n-com</i>	<i>têr-ou</i>	<i>et-hn-m-pêue</i>	
NOM-DEF.PL-power.F	all-3PL	REL-in-DEF.PL-heaven.PL	

‘And all the powers that are in heavens did not cease being disturbed’  
(Pistis Sophia, 1.1, AQ1 in Coptic SCRIPTORIUM)

In (17), the verb  $\kappa\alpha$  *ka*- is directly followed by the direct object  $\tau\omicron\tau\omicron\upsilon\gamma$  *toot-ou*, which bears the prosodic accent due to its diphthong. The lack of an intervening case marker between  $\kappa\alpha$  *ka*- and  $\tau\omicron\tau\omicron\upsilon\gamma$  *toot-ou* indicates a close syntactic relationship, with  $\kappa\alpha$  *ka*- being phonologically bound to  $\tau\omicron\tau\omicron\upsilon\gamma$  *toot-ou*. This contiguous construction is indicative of the verb’s immediate action upon the direct object.

- (18) Prepronominal state:  $\kappa\alpha\alpha$ = *kaa*- “place / leave”

$\epsilon\omega\chi\epsilon$	$\tau\epsilon\tau\mu\mu\alpha\gamma$	$\mu\pi\kappa\alpha\alpha\varsigma$
<i>ešče</i>	<i>t-et-mmau</i>	<i>mp-f-kaa-s</i>
if	DEF.SG.F-REL-there	NEG.PST-3SG.M-place.PPRO-3SG.F
$\gamma\mu\pi\mu\alpha\rho\alpha\Delta\epsilon\iota\varsigma\omicron\varsigma$	$\epsilon\varphi\mu\alpha\tau\prime\varsigma\omicron$	$\epsilon\rho\omicron$
<i>hm-p-paradeisos</i>	<i>e-f-na-t'so</i>	<i>ero</i>
in-DEF.SG.M-paradise	FOC-3SG.M-FUT-spare	DAT:2SG.F

‘If he didn’t leave the one who is there in paradise, is it you (Aphthonia) that he will spare?’

(Letter to Aphthonia in Coptic SCRIPTORIUM)

The prepronominal state in Coptic refers to the verb form that is immediately followed by a suffix pronoun or personal suffix, which functions as the direct object. In this state, just as with verbs that contain double vowel letters, the accent typically rests on the verb itself.

When a suffix pronoun serves as the direct object, employing the prepronominal state is not mandatory. An alternative construction is permissible, in which the suffix pronoun is joined to an object marker, and the verb appears in its absolute state. For instance, in (18), it is possible to use the form  $\kappa\omega$   $\mu\mu\omicron\varsigma$  *kô mmo-s* (place/leave.ABS ACC-3SG.F) as seen in (16), where  $\mu\mu\omicron\varsigma$  *mmo-s* denotes the third person singular feminine direct object, and the verb  $\kappa\omega$  *kô* is in the absolute state.

It is noteworthy that in Coptic, the absolute, prenominal, and prepronominal states can function as nouns without the need for a nominalising prefix. This multifunctionality allows these forms to be grouped under the term infinitives.

Finally, the stative form of the verb  $\kappa\omega$  *kô* ‘to to place’ or ‘to leave’ is  $\kappa\eta$  *kê*. This stative form encapsulates the resultant state or condition stemming from the action of the verb, providing a nominal or adjectival aspect to the verb’s meaning.

- (19) Stative:  $\kappa\eta$  *kê* ‘to be placed / left’

$\mu\kappa\alpha\iota\rho\omicron\varsigma$	$\mu\tau\mu\epsilon\tau\alpha\mu\omicron\iota\alpha$	$\kappa\eta$	$\mu\alpha\kappa$
<i>p-kairos</i>	<i>n-t-metanoia</i>	<i>kê</i>	<i>na-k</i>
DEF.SG.M-season	GEN-DEF.SG.F-repentance.F	place.STA	DAT-2SG.M

$\epsilon\gamma\mu\alpha\iota$   
*e-hrai*  
 DIR-upper.part

‘The season for repentance hath been set before thee’

(Pseudo-Ephrem, *Asceticon* 2, 4.8 in Coptic SCRIPTORIUM)

The stative in Coptic is used to express a continued state resulting from an action. When applied to transitive verbs, it conveys the ongoing state of being



acted upon, akin to a passive voice; with intransitive verbs, it describes the persistence of the action or state itself. Unlike infinitives, the stative form is distinct in that it never functions nominally.

Coptic also features a subset of verbs known as verboids, which are limited to the prenominal and prepronominal states. These verboids uniquely position the subject immediately after the verb. While most verboids are intransitive, there are some that are transitive, wherein the object, often a pronoun, follows the subject. In constructions where both the subject and the object are represented by suffix pronouns, the object pronoun assumes a specialised form, such as  $\pi\epsilon\chi\alpha\gamma\epsilon\varsigma$  *peča-u-sf* (said.PPRO-3PL.SBJ-3SG.M.OBJ) ‘they said it’. Here,  $\pi\epsilon\chi\alpha$ - *peča-* is a verboid and it has its subject after it.

## 7 Pseudo-noun incorporation and noun incorporation

In their analysis of Coptic through a linguistic typological lens, Grossman & Iem-molo (2013) contend that the structure characterised by a “prenominal state – object” in Coptic can be identified as a form of noun incorporation. Noun incorporation is a morphological phenomenon where transitive objects are integrated into the verb structure, a trait prominently observed in languages across various regions such as the Americas (exemplified by Mohawk and Classical Nahuatl), New Guinea (e.g., in the Yimas language), Northeast Asia (such as in Ainu), and in the Australian languages.

The current most popular orthography of Coptic (Type 3 in Figure 1) reflects phonological unity, which Layton (2011) refers to as bound groups and Haspel-math (2014) as stress groups. It is posited that the stress typically occurs on the first or second syllable within a “stress group”. This observation suggests that the prenominal state of the verb inherently embodies a cohesive phonological unit.

In the prenominal state of Coptic transitive verbs, Miyagawa (2023) observed phenomena that could be interpreted as pseudo-noun incorporation or actual noun incorporation. To arrive at this conclusion, the study employed the criteria set forth in Mithun (1984)’s scale.

The nature of the relationship between the prenominal verb and its object noun raises the question of whether noun incorporation is occurring. This is particularly relevant when considering verboids that have objects, as the tight syntactic bonding in the prenominal state might suggest such an incorporation process. This concept contrasts with the prepronominal state, which presents different syntactic characteristics.

This section provides an analysis of the *Gospel of Thomas* from the Nag Ham-madi Codex II, written in the Sahidic dialect influenced by the Lycopolitan di-alect, focusing on instances of noun incorporation and pseudo-noun incorpora-tion within the text.<sup>10</sup>

Table 3 presents a detailed breakdown of the frequency of various states and forms of verbs and verboids as they appear in the *Gospel of Thomas*.

Table 3: Frequency of the use of verbs in the *Gospel of Thomas*

Verb					Verboid	
ABS	PPRO	PNOM	STA	IMPER	PNOM	PPRO
421	129	104	91	7	128	68

## 7.1 Pseudo-noun incorporation

Pseudo-noun incorporation is well-attested in Polynesian languages such as Ni-uean and Maori.<sup>11</sup> Here is an example of pseudo-noun incorporation from Maori.

- (20) e [ruku~ruku koura nu~nui] ana (Maori)  
T/A [dive~PROG crayfish INT~big PROG  
‘He is diving for big crayfish<sup>12</sup>. (lit. big-crayfish-diving)’  
(Collberg (1997: 39))

In (20), the Maori construction *ruku~ruku* incorporates the noun phrase *koura nu~nui*, indicating a close syntactic relationship akin to incorporation. Maori syntax otherwise generally requires the use of a preposition to express the ob-ject. This syntactic feature of noun incorporation in Maori, often termed pseudo-incorporation, shares similarities with the Coptic prenominal verb + noun phrase constructions such, as (21).

In Coptic, if noun incorporation is recognised, it should be classified not as noun incorporation but rather as pseudo-noun incorporation.

<sup>10</sup>All the examples from the *Gospel of Thomas* are taken from Layton (2004).

<sup>11</sup>For Niuean pseudo-noun incorporation, see Massam (2001).

<sup>12</sup>I.e., *Plecoglossus altivelis*.



(23) Noun incorporation in the Gospel of Thomas

ϣⲛⲁⲱⲧⲡⲉ	ⲁⲛ	ⲙⲡⲙⲟⲩ
<i>f-na-či-tʰpe</i>	<i>an</i>	<i>m-p-mou</i>
3SG.M-FUT-receive.PNOM-taste NEG ACC-DEF.SG.M-death		
“He shall not taste death.”		

(Logos 1, the Gospel of Thomas (Layton 2004))

In (23), the complex verb form  $\chi\iota\tau\epsilon$  *či-tʰpe* is lexicalised, taking an additional object  $\mu\pi\mu\omicron\upsilon$  *m-p-mou* with the object marker, and it is separated from the compound verb  $\chi\iota\tau\epsilon$  *či-tʰpe* by the negative particle  $\alpha\iota$  *an*. This arrangement is a definitive example of noun incorporation, aligning with Type II of noun incorporation as outlined by Marianne Mithun (1984).

There are many examples of lexicalised compound verbs consisting of prenominal-state verbs and object nouns, see (24).

(24) Examples of noun incorporation in Coptic

- a. ⲧⲧⲕⲁⲥ *tʰ-tkas* (give.PNOM-pain) ‘to hurt’ (from Logos 28, the Gospel of Thomas, Layton 2004)
- b.  $\chi\iota\tau\epsilon$  *či-tʰpe* (receive.PNOM-taste) ‘to taste’ (from Logos 1, the Gospel of Thomas, Layton 2004)
- c. ϣⲓⲣⲟⲟⲩⲱ *fī-roouš* (take.PNOM-worry) ‘to worry’ (from Logos 36, the Gospel of Thomas, Layton 2004)
- d. ⲭⲉⲥⲟⲗ *če-col* (say.PNOM-lie) ‘to lie’ (from Logos 6, the Gospel of Thomas, Layton 2004)

These are highly lexicalised since the meaning is not following the Principle of Compositionality, and also since they can take a direct object with the object marker  $\iota$  *n-* after them. Therefore, they are examples of pure morphological noun incorporation.

### 7.3 The verbalizer *p- r-*

Let us consider some cases of the use of *p- r-*, the prenominal state of  $\epsilon\iota\pi\epsilon$  *eire* ‘to do’, especially with Greek loan verbs.

(25) Two uses of the prenominal state *p- r-* ‘to do’

- a. Usage with Greek verbs as the object

ϣⲛⲁⲣⲧⲓⲙⲁ	ⲙⲡⲟⲩⲁ
<i>f-na-r-tima</i>	<i>m=p-oua</i>
3SG.M-FUT-do.PNOM-honour ACC=DEF.SG.M=one	
‘He will honor the one.’	

(Logos 47, the Gospel of Thomas, Layton 2004)

## b. Usage with Coptic noun as the object

ⲕⲣⲟⲣⲱⲗ	ⲙⲡⲙⲁ	ⲧⲏⲣⲕ
<i>f-r-crôh</i>	<i>m=p=ma</i>	<i>têr-f</i>
3SG.M-do.PNOM-need	ACC-DEF.SG.M-place	all-3SG.M

‘He needs all the places.’

(Logos 67, the Gospel of Thomas, Layton 2004)

For example, in (25a) and (25b), the verbs ⲕⲙⲁⲣⲧⲙⲁ *f-na-r-tima* ‘he will do honor’ and ⲕⲙⲁⲣⲟⲣⲱⲗ *f-r-çrôh* ‘he needs’ demonstrate the use of *p-r-* with a Greek verb as an object and a Coptic noun as an object, respectively. Notably, both (25a) and (25b) feature an additional object marked by a contrapositional preposition.

In this *r-OBJ1 n=OBJ2* construction, *r-OBJ1* is a lexicalised compound verb, and *OBJ2* is the direct object of *r-OBJ1*, indicating the transitivity of *r-OBJ1*. The absolute state ⲉⲓⲣⲉ *eire* can take *OBJ1* as *eire n-OBJ1* but cannot take both *OBJ1* and *OBJ2*. Also, *p-r-* assumes a diluted sense of ‘to do’ compared to ⲉⲓⲣⲉ *eire*, serving mainly to index nouns and Greek loan verbs as verbs—a verbalising role. The near absence of semantic load when *p-r-* takes Greek verbs as *OBJ1* is evident.

In the case of *p-r-* + Greek verb, the initial element of this construction is increasingly assuming the role of a verbalizer prefix, particularly evident in the prenominal state of ⲉⲓⲣⲉ *eire* when accompanied by Greek loan verbs. This linguistic phenomenon suggests that within the spectrum of prenominal-state verbs, the verbalizer *p-r-* exhibits properties most akin to an affix.

In the *Gospel of Thomas*, several verbs exhibit a high frequency of occurrence in the prenominal state compared to their absolute forms. These verbs—we can call them light verbs—while not forming a closed set, tend to take on a grammaticalised or semantically bleached meaning when used in the prenominal state, particularly in constructions involving a direct object (see Table 4).

Table 5 highlights the notable frequency of the prenominal state *p-r-* against the absolute state ⲉⲓⲣⲉ *eire* and suggests that its significant usage indicates its grammatical integration as a morphological verbalizer prefix.

## 8 The morphemes-per-word (M/W) ratio

Finally, in order to objectively and quantitatively measure the polysynthetic nature of the Coptic language, the ratio of morphemes per word (M/W ratio henceforth) will be calculated. The M/W ratio is a linguistic index used to determine a language’s level of synthesis. Table 5 shows examples of M/W ratios in various languages. The higher the M/W ratio, the higher the syntheticity of the language.

Table 4: Frequency of prenominal vs. absolute states for the most frequent verbs in the *Gospel of Thomas*

PNOM	ⲡ <i>r-</i> (59)	ⲧ <i>ti-</i> (9)	ⲭⲓ <i>či-</i> (6)	ⲭⲉ <i>če-</i> (4)	ⲕⲓ <i>fi-</i> (3)	ⲙⲉⲥⲧⲉ <i>meste-</i> (3)	ⲛⲉⲭ <i>neč-</i> (2)
ABS	ⲉⲓⲣⲉ <i>eire</i> (7)	ⲧ <i>ti</i> (9)	ⲭⲓ <i>či</i> (3)	ⲭⲟ <i>čô</i> (11)	ⲕⲓ <i>fi</i> (2)	ⲙⲟⲥⲧⲉ <i>moste</i> (3)	ⲛⲟⲩⲭⲉ <i>nouče</i> (7)
Meaning	‘do’	‘give’	‘receive’	‘say’	‘take’	‘hate’	‘throw’

Table 5: M/W ratio of various languages based on Haspelmath &amp; Sims (2010: 6)

Language	Type	M/W ratio
West Greenlandic	Polysynthetic	3.72
Sanskrit	Synthetic	2.59
Swahili	Synthetic	2.55
Old English	Somewhat Synthetic	2.12
Lezgian		1.93
German		1.92
Modern English	Analytic	1.68
Vietnamese	Highly Analytic	1.06

For the purposes of this study, we define a word as a grammatical unit that can stand alone and convey a complete meaning, following the linguistic definition provided by Aronoff & Fudeman (2023).

In the context of Coptic, we will consider Layton’s concept of the bound group as the closest approximation to this definition of a word. Bound groups, as described by Layton, are prosodic units characterised by a single stress and often correspond to grammatical words. However, it is important to note that bound groups may also include clitics and other elements that are prosodically dependent but grammatically distinct.

For the calculation of the M/W ratio, we use the corpus of the *Letter to Aphthonia* written by Besa in the Sahidic dialect of Coptic, available in the Coptic SCRIPTORIUM, which uses spaces between Layton’s bound groups (Type 3 in Figure 1). In this corpus, prenominal-state words, articles, and complementizers

are written together with content words such as nouns and verbs, forming bound groups.

Table 6: The M/W ratio of *Letter to Aphthonia*

Words (bound groups)	Morphemes	M/W ratio
822	1,167	1.42

The resulting M/W ratio of 1.42 for the *Letter to Aphthonia* suggests that Coptic has a relatively low degree of synthesis, placing it closer to analytic languages on the typological spectrum. This finding aligns with the observations of Reintges (2011a,b) and Egedi (2007), who argue that Coptic displays a high degree of analyticity in its grammatical structure.

Furthermore, the results of this study challenge the claims made by Loprieno (1995), who argues for the polysynthetic nature of Coptic. The low M/W ratio indicates that Coptic words are not highly polysynthetic, as they do not exhibit the high number of morphemes per word typically associated with polysynthetic languages.

The M/W ratio also sheds light on the ongoing debate on the synthetic vs analytic nature of Coptic, as exemplified by the differing views of Haspelmath (2014) and Reintges (2011a). While Haspelmath argues for Coptic's synthetic status, the low M/W ratio found in this study lends support to Reintges' assessment of Coptic as an analytic language.

It is important to note that the M/W ratio is just one metric for assessing the synthetic or analytic nature of a language, and other factors, such as morphological and syntactic features, should also be considered. However, the quantitative evidence provided by the M/W ratio serves as a valuable contribution to the ongoing discussion on Coptic's typological classification and helps to substantiate the arguments made by scholars who propose an analytic status for the language.

## 9 Conclusion

In conclusion, perceptions of Coptic synthesis vary among scholars, largely due to the difficulty of defining word boundaries. Spaces are modern constructs, and punctuation and diacritics do not unequivocally indicate word boundaries. The introduction of the linguistic concept of clitics offers a more refined understanding of Coptic morpho-syntax. By evaluating elements such as articles, prenomi-

nal verbs, auxiliaries, and prepositions through the lens of *Freedom of Host Selection*, these are identified as clitics.

The morpho-syntactic analysis of Coptic, focusing on the interaction between clitics and word segmentation, reveals its analytic nature. Through the comprehensive exploration of noun incorporation and the functional dynamics of clitics, the study challenges and refines the traditional understanding of Coptic's grammatical structures.

This linguistic inquiry, although drawing on a limited spectrum of Coptic corpus analyses, ultimately positions the language closer to an analytic typology, characterised by a lower density of morphemes per word. The conclusion of this linguistic investigation is corroborated by the morpheme-per-word ratio derived from Coptic texts, which aligns with the typological features of more analytic languages.

The study's findings contribute significantly to the discourse on the degree of synthesis in Coptic morphology, offering new perspectives that could influence future linguistic research and the pedagogy of Coptic language studies. This research provides a vital step towards a more nuanced appreciation of Coptic's place in the landscape of linguistic typology.

## Abbreviations

ABS	absolute state	FOC	focalizer
ADJ	adjective	FUT	future
ADN	adnominal	GEN	genitive
ADVZ	adverbializer	HON	honorific
APPL	applicative	INDEF	indefinite
ART	article	INT	intensifier
CAUS	causative	LK	linker
CIRC	circumstantial	LOC	locative
COMP	complementizer	M	masculine
CONJ	conjunctive	N	noun
CONV	converb	NEG	negative
COP	copula	NOM	nominative
DAT	dative	NT	New Testament
DEF	definite	OBJ	object
DEM	demonstrative	OPT	optative
DIR	directional	PASS	passive
EXIST	existential	PL	plural
F	feminine	PNOM	prenominal



POSS	possessive	REL	relativizer
PPRO	prepronominal state	SBJ	subject
PRED	predicative	SG	singular
PROH	prohibitive	STA	stative
PROG	progressive	T/A	temporal/aspectual
PST	past	TAM	tense-aspect-mood
PRT	particle	TOP	topic
Q	question marker	VBLZ	verbalizer
REFL	reflexive		

## Acknowledgements

This paper is based on my presentations and further discussions at the 148th Conference of the Linguistic Society in Japan in 2014, the Digital Coptic 2 workshop in Washington D.C. in 2015, and the International Summer School on Typology and Lexis (TyLex) in Moscow in 2018. I thank Martin Haspelmath, Eitan Grossman, Amir Zeldes, Caroline Schroeder, Osahito Miyaoka, Tomomi Sato, Syuntarô Tida, the editor, and the two anonymous reviewers for their important comments.

## References

- Anderson, Stephen R. 2005. *Aspects of the theory of clitics (Oxford studies in theoretical linguistics)*. Oxford: Oxford University Press.
- Aronoff, Mark & Kirsten Anne Fudeman. 2023. *What is morphology?* 3rd edn. (Fundamentals of linguistics). London: Wiley-Blackwell.
- Browne, Gerald M. 1994. *The old Nubian miracle of Saint Menas*. Vienna: Mödling.
- Choat, Malcolm & Ian Gardner. 2014. *A Coptic handbook of ritual power*. Turnhout: Brepols.
- Collberg, Sheila Dooley. 1997. Determiners and incorporation in Maori. *Lund University, Dept. of Linguistics Working Papers* 46. 25–44.
- Crum, Walter Ewing. 1939. *A Coptic dictionary*. Oxford: Oxford University Press.
- de León, Martin. 1611. *Camino del cielo en lengua mexicana*. Mexico City: En la Empronta de Diego Lopez Dávalos.
- Dixon, Robert & Alexandra Aikhenvald. 2002. Word: A typological framework. In Robert Dixon & Alexandra Aikhenvald (eds.), *Word: A cross-linguistic typology*, 1–41. Cambridge: Cambridge University Press.

- Egedi, Barbara. 2007. Reconsidering the categorial status of Coptic suffix and conjugation base. In Kata Endreffy & András Gulyás (eds.), *Proceedings of the Fourth Central European Conference of Young Egyptologists*, 109–119. Prague: NKTH.
- Fortescue, Michael. 2013. Polysynthesis in the arctic/sub-arctic. In David A. Peterson Balthasar Bickel Lenore A. Grenoble & Alan Timberlake (eds.), *Polysynthesis in the Arctic/Sub-Arctic: How recent is it?*, 241–264. Amsterdam: John Benjamins.
- Grossman, Eitan & Martin Haspelmath. 2014. The Leipzig-Jerusalem transliteration of Coptic. In Martin Haspelmath, Eitan Grossman & Tonio Sebastian Richter (eds.), *Egyptian-Coptic linguistics in typological perspective* (Empirical Approaches to Language Typology [EALT] 55), 145. Berlin: Mouton De Gruyter.
- Grossman, Eitan & Giorgio Iemmolo. 2013. *A rare case of differential marking on A/S: The case of Coptic*. Paper presented at the 10th Conference of the Association for Linguistic Typology.
- Hachiya, Kunio. 2008. *Rōshi [laotzi]*. Tokyo: Iwanami Shoten.
- Haspelmath, Martin. 2014. A grammatical overview of Egyptian and Coptic. In Eitan Grossman, Tonio Sebastian Richter & Martin Haspelmath (eds.), *Egyptian-Coptic linguistics in typological perspective*, vol. 55 (Empirical Approaches to Language Typology), 103–44. Berlin: De Gruyter Mouton.
- Haspelmath, Martin. 2015. Defining vs. Diagnosing linguistic categories: A case study of clitic phenomena. In Joanna Błaszczak, Dorota Klimek-Jankowska & Krzysztof Migdalski (eds.), *How categorical are categories* (Studies in Generative Grammar), 273–303. Berlin: De Gruyter Mouton.
- Haspelmath, Martin. 2023. Defining the word. *Word* 69(3). 283–297.
- Haspelmath, Martin & Andrea D. Sims. 2010. *Understanding morphology*. 2nd edn. London: Hodder Education.
- Kammerzell, Frank. 2000. Egyptian possessive constructions: A diachronic typological perspective. *Sprachtypol. Univ. Forsch. (STUF)* 53. 97–108.
- Kuhn, Karl H. 1956a. *Letters and sermons of Besa* (Corpus Scriptorum Christianorum Orientalium 158). English translation. Louvain: Imprimerie Orientaliste L. Durbecq.
- Kuhn, Karl H. 1956b. *Letters and sermons of Besa* (Corpus Scriptorum Christianorum Orientalium 157). Coptic text. Louvain: Imprimerie Orientaliste L. Durbecq.
- Layton, Bentley. 2004. *Coptic gnostic chrestomathy: A selection of Coptic texts with grammatical analysis and glossary*. Leuven: Peeters.

- Layton, Bentley. 2011. *A Coptic grammar: With chrestomathy and glossary: Sahidic dialect*. 3rd edn. Wiesbaden: Harrassowitz.
- Loprieno, Antonio. 1995. *Ancient Egyptian: A linguistic introduction*. Cambridge: Cambridge University Press.
- Massam, Diane. 2001. Pseudo noun incorporation in Niuean. *Natural Language & Linguistic Theory* 19(1). 153–197.
- Mithun, Marianne. 1984. The evolution of noun incorporation. *Language* 60(4). 847–894.
- Miyagawa, So. 2022. *Shenoute, Besa and the Bible: Digital text reuse analysis of selected monastic writings from Egypt*. Göttingen: SUB Göttingen. DOI: 10.53846/goediss-9082.
- Miyagawa, So. 2023. Noun incorporation in Coptic. In Diliana Atanassova, Frank Feder & Heike Sternberg-el Hotabi (eds.), *Pharaonen, Mönche und Gelehrte: Auf dem Pilgerweg durch 5000 Jahre ägyptische Geschichte über drei Kontinente. Heike Behlmer zum 65. Geburtstag*, 556–574. Wiesbaden: Harrassowitz.
- Polotsky, Hans J. 1960. The Coptic conjugation system. *Orientalia* 29(3). 392–422.
- Prince, J. Dyneley. 1902. The modern pronunciation of Coptic in the mass. *Journal of the American Oriental Society* 23. 289–306.
- Quecke, Hans. 1984. *Das Johannesevangelium Saïdisch: Text der Handschrift PPalau Rib. Inv.-Nr. 183 mit den Varianten der Handschriften 813 und 814 Chester Beatty Library und Der Handschrift M 569*. Rome & Barcelona: Papyrologica Castrocaviana.
- Reintges, Christoph. 2011a. High analyticity and Coptic particle syntax: A phase-based approach. *The Linguistic Review* 28(4). 533–599.
- Reintges, Christoph. 2011b. Sapirian ‘drift’ towards analyticity and long-term morphosyntactic change in ancient Egyptian. In Ritsuko Kikusawa & Lawrence A. Reid (eds.), *Historical linguistics*, 289–328. Amsterdam: John Benjamins.
- Sato, Tomomi. 1992. Hōgō kara mita hoppō no shogengo [Northern languages from the perspective of noun incorporation]. In Osahito Miyaoka (ed.), *Kita no Gengo: Ruikai to rekishi [Northern languages: Typology and history]*, 141–152. Tokyo: Sanseido.
- Satzinger, Helmut. 2010. Old Nubian: Black African language of most ancient attestation. In Włodzimierz Godlewski & Adam Łajtar (eds.), *Between the cataracts: Part two, fasc. 1: Session papers*, 747–754. Oxford: Archeopress.
- Schroeder, Caroline & Amir Zeldes. 2016. Raiders of the lost corpus. *Digital Humanities Quarterly* 10(2). 1–13.
- Spencer, Andrew & Ana R. Luís. 2012. *Clitics: An introduction*. Cambridge: Cambridge University Press.

- Steindorff, Georg. 1883. *Gesios und Isidoros*. Leipzig: Hinrich.
- Steindorff, Georg. 1951. *Lehrbuch der koptischen Grammatik*. Chicago: University of Chicago Press.
- Stern, Ludwig. 1880. *Koptische Grammatik*. Leipzig: T. O. Weigel.
- Takla, Hany N. 1998. The revival and modernization of the Coptic language. *Bulletin of the Saint Shenouda the Archimandrite Coptic Society* 5. 117–124.
- Till, Walter. 1942. Zur Worttrennung im Koptischen. *Zeitschrift für Ägyptische Sprache und Altertumskunde* 77(1). 48–52.
- van Gerven Oei, Vincent. 2014. Remarks toward a revised grammar of old Nubian. *Dotawo: A Journal of Nubian Studies* 1(1). 165–184.
- van Gerven Oei, Vincent. 2022. *A reference grammar of Old Nubian*. Leuven: Peeters.
- Wilmet, Michel. 1958. *Concordance du nouveau testament sahidique, II. Les mots autochtones*, vol. 2. Louvain: Secrétariat du CorpusSCO.
- Zwicky, Arnold M. 1977. *On clitics* (Phonologica). Bloomington: Indiana University Linguistics Club.
- Zwicky, Arnold M. 1985. Clitics and particles. *Language* 61(2). 283–305.
- Zwicky, Arnold M. & Geoffrey K. Pullum. 1983. Cliticization vs. Inflection: English n't. *Language* 59(3). 502–513.