Coptic Scriptorium – Entity Annotation Guidelines

Amir Zeldes[1] & Lance Martin[2]

1 Georgetown University
2 The Catholic University of America

Version: 1.2.0_2024-11-12

1 Introduction

1.1 Preamble

Entity annotation concerns the annotation of **referring expressions** in a text, i.e. spans of text that refer to things in the world, and their classification into **entity types**. The purpose of entity annotation in Coptic Scriptorium is to facilitate searches which include specific entity types (e.g. finding a certain epithet using linguistic annotations, such as ⲟⲩⲁⲁⲃ ‘holy’, but only when applied to a PERSON), to inventorize entities (find all cases of e.g. places mentioned in the Apophthegmata Patrum), and to function as a gateway for entity linking, enabling searches for specific persons (“John the Baptist”), regardless of the exact expression used to mention them. The latter task of entity linking is left outside of the scope of the current guidelines.

Entity annotation can be applied to three types of referring expressions:

- Named entities, which are headed by a proper noun (e.g. “Apa Papnoute”)
- Non-named entities, headed by a common noun (e.g. “the angel”)
- Pronouns – these are currently not annotated by our schema (e.g. “she” is a person)

1.2 Referring expressions

Almost all nouns and proper nouns correspond to referring expressions, with the exception of non-referring nouns, such as:

- ⲁϩⲉ ⲣⲁⲧ.. - “stand, set foot” - does not actually refer to the foot of a person
- ϩⲛ ⲟⲩ ⲙⲉ - “truly” - does not actually introduce a referenceable ‘truth’

One test for referentiality is whether a pronominal or nominal subsequent mention is possible/plausible. For example, the following sounds odd:

- ?? ⲁϥⲁϩⲉⲣⲁⲧϥ ⲁⲩⲱ ⲡⲉⲓⲣⲁⲧ …“he stood on foot, and this foot…”

For more examples, also see the section “Non-referring expressions” below.

2 Entity Types

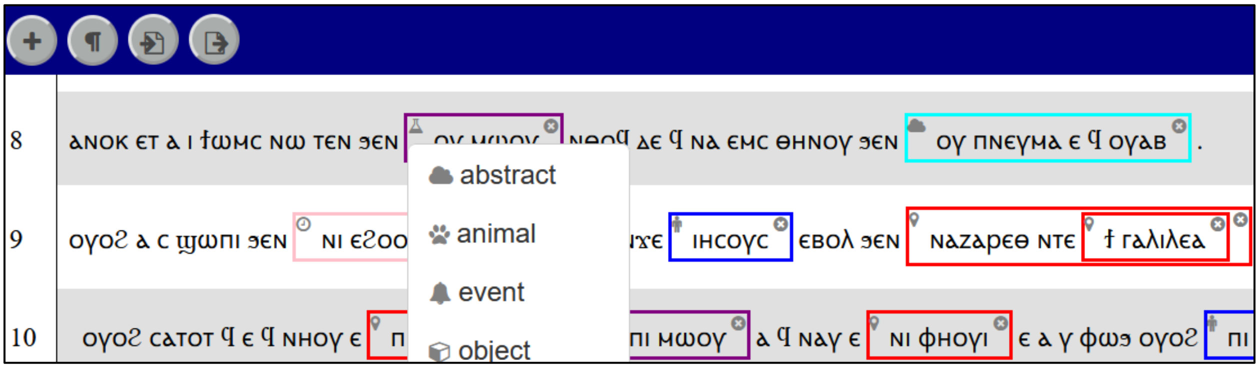
We distinguish 10 entity types:

- ABSTRACT - intangible entities not covered by other classes (incl. ideas, emotions)
- ANIMAL – dog, fish, …
- EVENT – an occurrence, e.g “the death of the king”, “the arrival of a monk”
- OBJECT – concrete inanimate object not belonging to other categories
- ORGANIZATION – organized body of people, e.g. ⲧⲉⲕⲕⲗⲏⲥⲓⲁ, ⲧⲉⲥⲧⲣⲁⲧⲉⲓⲁ
- PERSON – references to humans, loose groups of humans (ⲙⲏⲏϣⲉ ‘crowd’), deities
- PLACE – towns, countries, but also ad-hoc places (behind the house, outside)
- SUBSTANCE – mass noun indicating a material, e.g. sand, water, wine
- TIME – date terms, durations like ‘year’, ‘day’, terms like ‘moment’

3 Annotation interface

3.1 Typed entity spans

The annotation interface uses colorful, possibly nested boxes to represent entity spans. You can switch between a running text view and a sentence-by-sentence view using the ¶ button (second button in the blue toolbar below). In sentence mode, you can also hover over the sentence number (8, 9 or 10 in the image below) to see the translation of the current sentence (if available in a ‘translation’ column in the GitDox spreadsheet).



8 ⲁⲛⲟⲕ ⲉⲧ ⲁ ⲓ ϯⲱⲙⲥ ⲛⲱ ⲧⲉⲛ ϧⲉⲛ ⲟⲩ ⲙⲱⲟⲩ ⲛⲑⲟϥ ⲇⲉ ϥ ⲛⲁ ⲉⲙⲥ ⲑⲏⲛⲟⲩ ϧⲉⲛ ⲟⲩ ⲡⲛⲉⲩⲙⲁ ⲉ ϥ ⲟⲩⲁⲃ .

9 ⲟⲩⲟϩ ⲁ ⲥ ϣⲱⲡⲓ ϧⲉⲛ ⲛⲓ ⲉϩⲟⲟ ⲓϫⲉ ⲓⲏⲥⲟⲩⲥ ⲉⲃⲟⲗ ϧⲉⲛ ⲛⲁⲍⲁⲣⲉⲑ ⲛⲧⲉ ϯ ⲅⲁⲗⲓⲗⲉⲁ

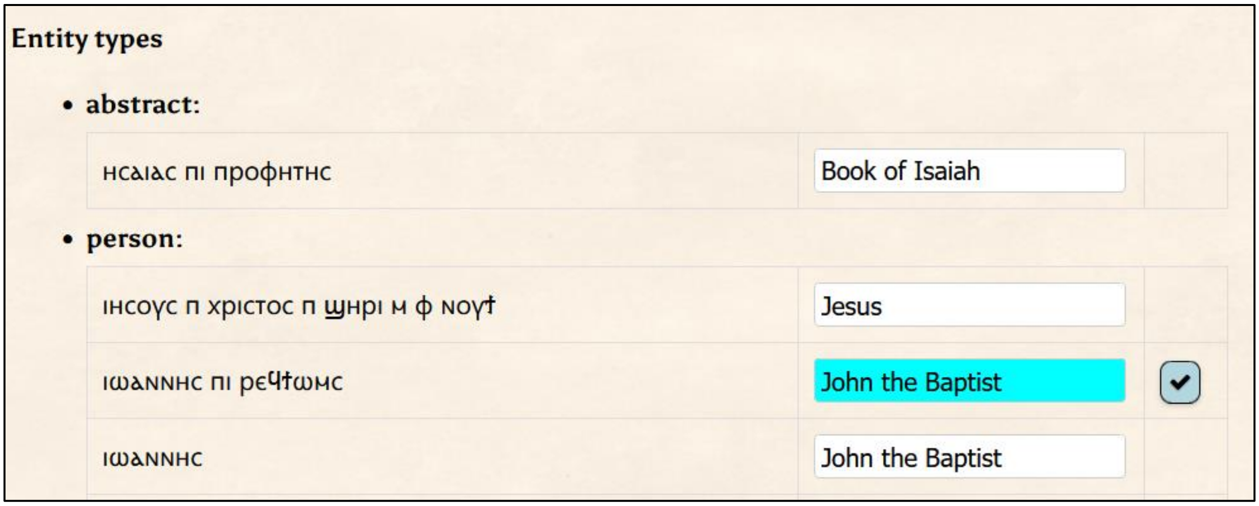
10 ⲟⲩⲟϩ ⲥⲁⲧⲟⲧ ϥ ⲉ ϥ ⲛⲏⲟⲩ ⲉ ⲡ ⲡⲓ ⲙⲱⲟⲩ ⲁ ϥ ⲛⲁⲩ ⲉ ⲛⲓ ⲫⲏⲟⲩⲓ ⲉ ⲁ ⲩ ⲫⲱϣ ⲟⲩⲟϩ ⲡⲓ

To obtain initial entity predictions, you can click on the [Auto NER] button below the document window. Predicted entity spans can be adjusted by dragging their left or right boundaries and releasing over the word which should be at the left/right end of the entity span. Entities may not span across sentence boundaries. Additional controls:

- To change the entity type, click the icon at the top left of the box and choose the entity type
- To delete an entity, click the little [x] button at the top right of its box
- To add a new entity, click to select a token (or drag to select multiple tokens) and press the [+] button, or hit Enter. You can also make a single token span by selecting just one token first, then adjust the span, instead of selecting the exact span first.

3.2 Entity identity linking

After completing corrections for entity spans, click the button [List named entities] to display a list of named entity candidates for entity linking (a.k.a. Wikification). Clicking the button will display a textbox for each unique mention text which is headed by a word tagged as NPROP in each entity category. Note that multiple mentions with exactly the same text (e.g. repeated mentions of ⲓⲏⲥⲟⲩⲥ) will receive only one box for entity linking.



Entity types

- **abstract:**

ⲏⲥⲁⲓⲁⲥ ⲡⲓ ⲡⲣⲟⲫⲏⲧⲏⲥ	Book of Isaiah	

- **person:**

ⲓⲏⲥⲟⲩⲥ ⲡ ⲭⲣⲓⲥⲧⲟⲥ ⲡ ϣⲏⲣⲓ ⲙ ⲫ ⲛⲟⲩϯ	Jesus	
ⲓⲱⲁⲛⲛⲏⲥ ⲡⲓ ⲣⲉϥϯⲱⲙⲥ	John the Baptist	✓
ⲓⲱⲁⲛⲛⲏⲥ	John the Baptist	

The objective in this task is to enter the title of the Wikipedia article corresponding to each entity text, if such an article exists; otherwise, enter “(pass)”. You can optionally click on [Guess identities] to receive suggestions from the system, which will be highlighted in blue, and must be confirmed individually by clicking on the checkmark next to them. If you type a title yourself, the system will also provide auto-complete suggestions from titles already in the system.

If you see ‘mangled’ entity texts (e.g. incomplete phrases, or some superfluous words), please check the entity boxes in the entity annotation to ensure boundaries are correct. If one of the suggestions is not a named entity, please check the POS tagging layer in the spreadsheet mode (you can safely switch between spreadsheet and entities mode, as long as you click save after editing the entities).

4 Markable selection guidelines

4.1 Appositions

Repeated mentions of the same entity in apposition are considered a single span, and do not contain more mentions of the same entity:

- [ⲓⲱϩⲁⲛⲛⲉⲥ ⲡ ⲃⲁⲡⲧⲓⲥⲧⲏⲥ] “[John the Baptist]”
- [ⲡ ⲣⲣⲟ ⲍⲏⲛⲱⲛ] “[King Zeno]”
- [ⲡⲉⲛ ⲡ ⲉⲧ ⲟⲩⲁⲁⲃ ⲕⲁⲧⲁ ⲥⲙⲟⲧ ⲛⲓⲙ ⲁⲡⲁ ⲕⲩⲣⲟⲥ ⲡ ⲉⲛⲧ ⲁ ϥ …] “[Our Holy One in every way, Apa Cyrus, who has …]”

Although outwardly similar, appositions must be distinguished from dislocations, in which a pronominal subject or object is repeated separately. For personal pronouns, the pronoun is simply left out of the nominal span:

- [ⲡⲉϥ ⲉⲓⲱⲧ] ϥ ⲛⲁⲩ ⲉⲣⲟ ⲟⲩ - “[his father], he sees them”
- ϥ ⲛⲁⲩ ⲉⲣⲟ ⲟⲩ ⲛϭⲓ [ⲡⲉϥ ⲉⲓⲱⲧ] - “he sees them, that is [his father]”

If the pronoun is a substitutive demonstrative (ⲡⲁⲓ, ⲧⲁⲓ, ⲛⲁⲓ), then two spans are annotated:

- [ⲡⲉϥ ⲉⲓⲱⲧ] [ⲡⲁⲓ] ⲛⲁⲩ ⲉⲣⲟ ⲟⲩ - “[his father], [this one] sees them”
- [ⲡⲁⲓ] ⲛⲁⲩ ⲉⲣⲟ ⲟⲩ ⲛϭⲓ [ⲡⲉϥ ⲉⲓⲱⲧ] - “[this one] sees them, [that is his father]”

But note that it is also possible for a substitutive demonstrative to stand in true apposition to a noun without dislocation, in which case a single span is annotated as for any apposition:

- ⲁ ⲓ ⲛⲁⲩ ⲉ [ⲡⲉϥ ⲉⲓⲱⲧ , ⲡⲁⲓ ⲉⲧ ⲙⲉⲣⲓⲧ ⲥ] - “I saw [their father, the one who loves her]”

See the UD Coptic guidelines for more information on identifying dislocation vs. apposition.

4.2 Expanded Relative Constructions

The relative construction expanding an article is annotated as an entity:

- [ⲡ ⲉⲧ ⲟⲩ ⲥⲱⲧⲙ ⲉⲣⲟ ϥ] “the one they listened to” (person)

However, if the ⲡ is tagged as a copula, that part of the construction is not part of the entity span, since it is part of a predication. In these instances, we view the predicate noun phrase as an entity, and the relative clause as a subject clause (compare the Universal Dependencies annotation guidelines):

- [ⲡ ⲛⲟⲩⲧⲉ] ⲡ/COP ⲉⲛⲧ ⲁ ϥ ⲁⲩⲝⲁⲛⲉ “It is God who made them grow”

In this example, “God” receives a span, but “who made them grow” is considered a subject clause (i.e. ‘who made them grow is God’), which is not nominal and hence not annotated. Note that according to the tagging guidelines, the second ⲡ should be tagged as COP and lemmatized ⲡⲉ in this sentence. Therefore there is only a single entity mention in this case, [ⲡ ⲛⲟⲩⲧⲉ].

Nominal sentences with a relative construction expanding the article should not be mistaken for the copula construction although similar in form. Compare the following, in which the ⲡ in the second phrase is tagged ART:

- [ⲡ ⲛⲟⲩⲧⲉ] [ⲡ/ART ⲉⲛⲧ ⲁ ϥ ⲁⲩⲝⲁⲛⲉ] ⲡⲉ/COP “[God] is [the one who made them grow]”
- [ⲡⲉⲛ ϫⲟⲉⲓⲥ] [ⲡ/ART ⲛⲟⲩⲧⲉ ⲉⲧ ⲥⲱⲧⲙ ⲉⲣⲟ ⲛ] ⲡⲉ/COP “Our Lord is the God who hears us”
- [ⲡⲉⲛ ϫⲟⲉⲓⲥ] ⲡⲉ [ⲡ ⲛⲟⲩⲧⲉ ⲉⲧ ⲥⲱⲧⲙ ⲉⲣⲟ ⲛ] “It is our Lord, the God who hears us”

In the first example, both “God” and “the one who made them grow” receive spans. “God” is the subject and “the one who made them grow” is the predicate, forming a nominal sentence with the copula ⲡⲉ. The ⲡ after ⲛⲟⲩⲧⲉ is part of the predicate noun phrase in this instance: an article heading the relative clause which is the predicate for the copula at the end. It functions similarly to the relative clause expanding the nouns, i.e., “the God who hears us” in the subsequent examples.

The expansion of an article can stand in apposition if it repeats an entity mention. The repeated mention is annotated with one span, like all appositions, whereas subject and predicate spans are considered separate mentions:

- [ⲡ ⲛⲟⲩⲧⲉ , ⲡ ⲉⲛⲧ ⲁ ϥ ⲁⲩⲝⲁⲛⲉ] [ⲡ ⲁⲅⲁⲑⲟⲥ] ⲡⲉ “God, the one who made them grow, he is the good one”

Here, “the one who made them grow” is again headed by an article and stands in apposition to “God.” The final ⲡⲉ is the copula in the sentence and marks the noun phrase before it as a predicate, “the good one.”

The position of the copula does not change the meaning and is not annotated. Note also that in first and second person, there is no copula:

- ⲁⲛⲅ [ⲟⲩ ⲭⲣⲓⲥⲧⲓⲁⲛⲟⲥ] “I am [a Christian]”

Occasionally, however, the copula can interrupt a referring expression as in the example below where the head noun ⲡⲛⲟⲩⲧⲉ is separate from the relative clause by the copula, ⲡⲉ. In such cases, it should be included in the span (see 3.3 for more detail).

- [ⲡⲉⲛ ϫⲟⲉⲓⲥ] [ⲡ ⲛⲟⲩⲧⲉ **ⲡⲉ** ⲉⲧ ⲥⲱⲧⲙ ⲉⲣⲟ ⲛ] "Our Lord is the God who hears us"

4.3 Interrupted spans

Entity expressions interrupted e.g. by a copula or particle are spanned to **contain** the copula or particle. For example, the following span includes the intervening copula:

- [ⲛⲉϥ ⲁⲡⲟⲥⲧⲟⲗⲟⲥ **ⲛⲉ** ⲉⲧⲟⲩⲁⲁⲃ]$_{\text{PERSON}}$ 'it is his holy apostles' (literally [*his apostles are which holy*], with intruding 'are')

Similarly:

- [ϥⲧⲟⲟⲩ ⲇⲉ ⲛ ϩⲟⲟⲩ]$_{\text{TIME}}$ – but four days (lit. '[four but days]')
- [ⲟⲩ ϣⲃⲏⲣ · **ϩⲱⲱ ⲕ ⲟⲛ** ⲛⲧⲉ ⲡ ⲛⲟⲩⲧⲉ]$_{\text{PERSON}}$ – but also for your part a friend of God

Non-adjacent relative clauses are included, **unless the interruption contains the verb controlling the head noun** (this prevents some possibly very long 'hermeneutical' relatives inside mentions):

- [ⲣⲱⲙⲉ ⲛⲓⲙ ⲟⲛ ⲉⲧ ⲥⲱⲧⲙ] - and also [any man who hears] (note the interruption 'ⲟⲛ', and inclusion of the relative clause)

But do not include a clause past the verb controlling the head of the span:

- ⲉⲣϣⲁⲛ [ⲧ ⲃⲁϣⲟⲣ]$_{\text{ANIMAL}}$ ⲁϣⲕⲁⲕ ⲉⲃⲟⲗ ⲁⲛ **ⲉⲧⲉ ⲛⲧⲟⲕ ⲡⲉ** … - it is not when [the fox] barks, which is you, … (postponed hermeneutic relative clause in bold is not included, because it appears after the verb ⲁϣⲕⲁⲕ, which controls 'fox' as a subject)

In this case the interruption by the verb ⲁϣⲕⲁⲕ 'bark' which is the predicate of 'fox' triggers the guideline to omit the relative clause. Otherwise, the mention could potentially cover the entire clause, in this case: ⲧ ⲃⲁϣⲟⲣ ⲁϣⲕⲁⲕ ⲉⲃⲟⲗ ⲁⲛ ⲉⲧⲉ ⲛⲧⲟⲕ ⲡⲉ ⲡ ϩⲙϩⲁⲗ ⲙ ⲡ ⲙⲁⲙⲙⲱⲛⲁⲥ ϩⲛ ϩⲉⲛ ϩⲣⲟⲟⲩ ⲉ ⲩ ⲟϣ ….

4.4 Possessive Constructions

The possessive article construction, e.g. ⲡⲁ ⲡⲁⲩⲗⲟⲥ, forms two spans, as follows, with the entity type being decided based on meaning:

- [ⲡⲁ [ⲡⲁⲩⲗⲟⲥ]$_{\text{PERSON}}$]$_{\text{PERSON}}$ - the ones (=people) belonging to Paul

But note that regular possessive articles are not annotated with spans, just as other pronouns are not annotated:

- [ⲡⲉϥ ⲏⲓ]$_{\text{PLACE}}$ – his house

4.5 Groups and other quantity constructions

Semantically 'empty' heads such as quantity nouns (compare English 'a number of people', which is not both 'a number' and 'people'; similarly 'a lot of', 'the majority of' etc.) are only given one span, for example:

- [ϩⲁϩ ⲛ ⲥⲟⲡ]$_{\text{TIME}}$ – a lot of times

Groups of entities are generally interpreted as the entity type of their constituents, for example, a herd of animals is of the type animal:

- [ⲟⲩ ⲁⲅⲉⲗⲏ ⲛ ϣⲟϣ]$_{\text{ANIMAL}}$ - a herd of buffaloes

Note that there is no nested entity for 'buffalo' in this case, since there is no distinct entity being mentioned (the herd consists exactly of all buffaloes being discussed). This is different in cases where the nested entity is not identical in reference, e.g. '[the houses of [the city]]', where 'the city' can be said to contain more than just the houses.

An exception to the guideline that groups are classified as their constituent type is cases of people who form an organization, e.g. ⲥⲩⲛⲁⲅⲱⲅⲏ, ⲥⲧⲣⲁⲧⲉⲩⲙⲁ etc. are 'organization', not 'person'.

4.6 Idiomatic Expressions

Idiomatic expressions should be annotated as fully as possible even when certain components have low referentiality.

- [ⲛ ϩⲁⲗⲁⲧⲉ ⲛ [ⲧ ⲡⲉ]$_{\text{PLACE}}$]$_{\text{ANIMAL}}$ "the birds of the sky (Mark 4:32)

In this example, the entire phrase may be a way of referring to birds in general, and "sky" therefore is unlikely to be referred to again. However, since it is possible that it could be referred to again and since it passes the referentiality test (1.2), it should be annotated.

4.7 No reference inside compounds

In morphologically complex items containing a verb inside a larger token, that noun cannot be annotated:

- ⲁ ϥ ϫⲓⲃⲁⲡⲧⲓⲥⲙⲁ - he received-baptism

In this case baptism cannot be annotated as an entity, since it’s part of an incorporated verb ‘to baptize’, and receives the part of speech V in Coptic Scriptorium guidelines.

4.8 Coordination

Do not mark coordinate entities in addition to their constituents:

- [ⲓⲱϩⲁⲛⲛⲏⲥ] ⲙⲛ [ⲁⲛⲧⲱⲛⲓⲟⲥ]

In this case we do not also annotate [ⲓⲱϩⲁⲛⲛⲏⲥ ⲙⲛ ⲁⲛⲧⲱⲛⲓⲟⲥ] as a third mentioned entity.

4.9 Container and substance

Container and substance form two entities, for example:

- [ⲟⲩ ⲡⲩⲅⲏ ⲙ [ⲙⲟⲟⲩ]$_{\text{SUBSTANCE}}$]$_{\text{PLACE}}$ - a fountain of water

The fountain can be a PLACE or OBJECT in context, but the water is SUBSTANCE, and both can be referred to separately later on.

4.10 Numeral entities

Numbers standing in for a phrase can be entities, most commonly:

- [ⲟⲩⲁ]$_{\text{PERSON}}$ - ‘one (person)’

4.11 Distributive entities

Repeated distributive noun constructions are interpreted as single entity mentions:

- [ⲡ ⲟⲩⲁ ⲡ ⲟⲩⲁ]$_{\text{PERSON}}$ - ‘one by one’, ‘each man’

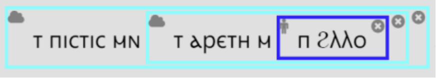
The rationale is that these are like a plural reference, rather than two mentions of individuals (in this case there can be more than two people, and they do not map neatly onto the two numerals).

4.12 Coordinated modifiers

In cases where a modifier applies to both nouns in a coordination, both spans will contain the modifier, creating a nested span for the second conjunct, for example:

- [ⲧ ⲡⲓⲥⲧⲓⲥ ⲙⲛ [ⲧ ⲁⲣⲉⲧⲏ ⲙ [ⲡ ϩⲗⲗⲟ]$_{\text{PERSON}}$]$_{\text{ABSTRACT}}$]$_{\text{ABSTRACT}}$ - ‘the faith and virtue of the old man’

Note that since both ‘faith’ and ‘virtue’ are of the old man, based on the interrupted span guideline, the ‘faith’ span contains ‘virtue’ too (otherwise the ‘old man’ would be left out of the first span). Graphically this results in the following configuration:



5 Entity classification guidelines

5.1 Body Parts

Most body parts are marked as objects, since they are tangible:

- [ⲟⲩ ϭⲓϫ]$_{\text{OBJECT}}$ – “a hand”
- [ⲡⲉϥ ⲃⲁⲗ]$_{\text{OBJECT}}$ – “his eye”

However some referential body parts are usually considered abstract, notably ϩⲏⲧ ‘heart’, which usually refers to one’s spirit, emotions, etc., and not the physical organ:

- ϯ ⲛⲁ ⲧⲣⲉ [ⲡⲟⲩ ϩⲏⲧ]$_{\text{ABSTRACT}}$ ⲙⲕⲁϩ – “I will make your [heart] suffer”

Other uses of body parts may be totally figurative or idiomatic (i.e. not referring to anything), in which case they are not annotated – see ‘Non-Referring Expressions’ above.

5.2 Parts of Plants

Parts of plants are marked as plants.

- [ⲡ ⲕⲗⲁⲇⲟⲥ]$_{\text{PLANT}}$ “ the branch”
- [ⲡ ϭⲣⲟϭ]$_{\text{PLANT}}$ “the seed”

5.3 Peoples and demonyms

Pluralized demonyms indicating members of a people are labeled person:

- [ⲛ ϩⲉⲗⲗⲏⲛ]$_{\text{PERSON}}$

However peoples mentioned as a people (not as a group of individuals) are labeled organization:

- [ⲡⲉⲕ ⲗⲁⲟⲥ ⲓⲥⲣⲁⲏⲗ]$_{\text{ORGANIZATION}}$

These cases are usually singular and involve a named people. This guideline does not apply to ad-hoc groups of people who do not form an organized entity, e.g. ⲙⲏⲏϣⲉ ‘crowd’ is still annotated as PERSON.

5.4 Substance and Object

Inanimate entities that can be counted should be marked as object:

- [ⲛ ⲡⲉⲧⲣⲁ]$_{\text{OBJECT}}$ “rocks”
- [ϣⲟⲙⲛⲧ ⲛ ⲟⲉⲓⲕ]$_{\text{OBJECT}}$ “Three breads (i.e., three loaves of bread)

Inanimate entities that cannot be counted should be marked as substance:

- [ⲡ ⲕⲁϩ]$_{\text{SUBSTANCE}}$ “the soil”
- [ⲙⲟⲟⲩ]$_{\text{SUBSTANCE}}$ “water”

5.5 Event and Abstract

Nominalized infinitives are often events:

- [ⲡ ϩⲱⲧⲃ]$_{\text{EVENT}}$ “the murder”

5.6 Multiple entity types in one chain

It is possible for an entity to be referred to in multiple ways, underscoring different aspects of the entity. In such cases, where it seems clear that the entity types are distinct (e.g. due to metonymy, metaphorical extensions, etc.), it is possible to have different entity types

- [ⲡⲉ ⲭⲣⲓⲥⲧⲟⲥ]$_{\text{PERSON}}$ ⲡⲉ [ⲧⲉⲓ ⲥⲛⲧⲉ]$_{\text{ABSTRACT}}$ - ‘this foundation is Christ’ (where ‘foundation’ refers to ‘Christ’ metaphorically)

Mark each entity with its own type, i.e., ‘Christ’ as PERSON and ‘foundation’ as ABSTRACT.

6 Non-referring expressions

6.1 Interrogatives

No annotations are needed for plain interrogatives (ⲛⲓⲙ, ⲟⲩ), but complex interrogatives including a noun are annotated:

- [ⲛⲓⲙ ⲅⲁⲣ ⲛ ⲣⲱⲙⲉ]$_{\text{PERSON}}$ ⲡ ⲉⲧ ⲥⲟⲟⲩⲛ ⲛ [ⲛⲁ [ⲛ ⲣⲱⲙⲉ]$_{\text{PERSON}}$]$_{\text{PERSON}}$ “[what human] is he who knows [those things which are of humans]]?”
- [ⲁϣ ⲙ ⲡⲓⲣⲁⲥⲙⲟⲥ]$_{\text{EVENT}}$ “[what kind of trial]?”

6.2 Common figurative and other fixed expressions to ignore

The following are considered idiomatic or functional expressions, in which the constituent nouns are not construed as referential and no annotation is needed:

- ⲁϩⲉ ⲣⲁⲧ ϥ – ‘stand, set foot’ - “foot” is not an entity mention
- ⲉ ⲡ ⲉⲥⲏⲧ - ‘down’, lit. ‘to the ground’
- ⲉ ⲡ ⲧⲏⲣϥ meaning ‘at all’ is not referential
- ⲛⲉⲩ ⲉⲣⲏⲩ “themselves”
- ⲙ ⲙⲏⲛⲉ - ‘daily’
- ⲛ ⲟⲩ ⲕⲟⲩⲓ - ‘a little’ (manner adverbial; note ⲕⲟⲩⲓ *can* be referential if referring to a person or thing, e.g. ⲛⲧⲕ [ⲟⲩ ⲕⲟⲩⲓ]$_{\text{PERSON}}$ ‘you are a little one)
- ⲛ ⲟⲩ ϩⲟⲩⲟ - ‘more so’
- ⲛ ⲟⲩⲱⲧ - ‘together’
- ⲛ ϣⲟⲣⲡ - ‘first(ly)’
- ⲛ ⲧⲉ ⲩⲛⲟⲩ - ‘then’
- ⲛ ⲧ ϩⲉ - meaning ‘like’, and similarly Bohairic ⲫ ⲣⲏⲧⲓ
- ⲡⲁϩⲟⲩ ⲙ
- ⲣ ϩⲛⲁ ϥ – ‘want, do one’s will’ - the word ϩⲛⲁ / ϩⲛⲉ ‘will’ is figurative, as this is a fixed expression for ‘desire’
- ϩⲁ ⲉⲟⲟⲩ - ‘glorious’ - the ⲉⲟⲟⲩ is not referential, as the expression is only used adjectivally
- ϩⲓ ⲟⲩ ⲥⲟⲡ - ‘at once’
- ϭⲟⲙ - meaning ‘capable’ in constructions like ⲛⲧⲕ ϭⲟⲙ ⲁⲛ ‘you are not capable’, but ϭⲟⲙ can be referential in, e.g. [ⲧ ϭⲟⲙ ⲙ [ⲡ ⲛⲟⲩⲧⲉ]]$_{\text{ABSTRACT}}$; the same holds for Bohairic expressions with ϣϫⲟⲙ.
- (ϩⲱⲃ) ⲛ ϭⲓϫ – ‘handywork’ - the whole phrase (handywork) is ABSTRACT or OBJECT in context, but ‘hand’ is not a referent
- ϯ ⲧⲟⲟⲧ ϥ – ‘help, give a hand’

6.3 Entities with low referentiality

If it is unclear if a span is referring or non-referring, it should be annotated if it has an article and is not on the list in 5.2. Otherwise, it should not receive annotation:

- ⲁϥⲛⲧⲟⲩ ⲉ [ⲟⲩ ⲥⲁ]$_{\text{PLACE}}$ - “he took them to [a side]”

However, “behind” should not be annotated since it lacks an article:

- ⲉ ⲡⲁϩⲟⲩ ⲙ … - “behind…”