About this Guide

This annotation style guide was created by Chiara Palladino and Farnoosh Shamsian for the Ugarit Project. The languages involved are Ancient Greek and English.

This guide addresses the structural linguistic differences between Ancient Greek, a highly inflected language, and English, a more synthetic language. They are designed specifically to address the alignment of Ancient Greek texts and their English translations, but parts of them can be repurposed for other inflected languages.

The style guide was created as follows:
- The authors wrote a draft version of the Guidelines through multiple meetings and discussions: the sources of variation and linguistic differences were identified and sorted based on their previous years-long experience with aligning Ancient Greek and English with Ugarit.
- The authors manually aligned three text samples, taken from poetry and prose, which included: Plato's *Crito* (1777 words); the *Iliad* (2351 words); Xenophon's *Cyropaedia* (1523 words). With these examples, they tested the general consistency and feasibility of the style guide.
- For each new or unpredicted variation, there was a brief discussion and a preferred annotation style or improvement in the style guide was agreed upon. Each change was incorporated into the final version of the style guide.

General Principles

These guidelines follow the definition of Lambert et al.: "the correspondence between two lexical units (single words or expressions) should involve on both sides as few words as possible, but as many words as necessary, with the requirement that the linked words or groups bear the same meaning [...]. If two single words match, they should be linked together, but if there is no correspondence at a single word level, groups of words should be linked together. Groups of words linked together should be as small as possible." In addition, "the only valid elements in an alignment are single words and indivisible groups of words" (p.275): groups of words are linked together when the meaning of the group is distinct from that of the sequence of each word's meaning, and single words cannot be separated from the rest of the group without changing their meaning (= indivisible lexical unit).

Links allowed in Ugarit are one-to-one (1-1), many-to-many (N-N), one-to-many (1-N) and many-to-one (N-1). Currently it is not possible to distinguish between ambiguous and unambiguous, complete or partial. Links in Ugarit do not consider lack of alignment: words that do not correspond are simply left unaligned.

IMPORTANT NOTE: these guidelines can be used to perform alignments between various texts in Ancient Greek and English and to create gold standard corpora. However, they are not

project-specific. If you have a particular research question and you need specific results for the purpose of language research, pedagogy, or data mining, you should create your own guidelines (you can reuse any part of these guidelines).

Guidelines

1. Punctuation

Punctuation is never aligned.

Exception: the apostrophe in the Saxon genitive, when aligned with a corresponding word/group of words in Ancient Greek.

2. Words omitted or incorrectly translated

If a group of words only appears in one language whose meaning has no correspondence in the other, it is not aligned.

If a group of words is an incorrect translation of a group of words in the source language, it is not aligned.

3. Phrasal Construction and Idioms

These are expressions that constitute an indivisible lexical unit, where each word singularly does not bear the same meaning, but the group of words together is a semantic equivalent in both languages. They are aligned with a N-N link if there is an English equivalent. Examples: dative of possession: 'ἔστιν αὐτοῖς' - 'they have'; 'Ἔτι καὶ νῦν' - 'even to this day'; 'Ἐκ χειρός' - 'in a hand-to-hand encounter'.

4. Repetition

When a word is repeated in a language but not in the other, only the first instance in each language is aligned.

5. Negation

Whenever possible, negation is aligned with a 1-1 link. However, in two cases that won't be possible. In such cases, an N-N strategy is recommended.
   1. When the negative English morpheme is abbreviated (-n't):
        a. When it cannot be separated from the word group, it is aligned N-N. Example: 'ἔστιν οὐ' - 'it isn't'.
        b. When it only consists of an auxiliary, e.g. 'don't', it may be aligned with the negative particle, and the verb is aligned with the corresponding English word. Example: μὴ ἄλλως πόιει - don't do otherwise: 'μὴ' - 'don't', 'πόιει' - 'do', 'ἄλλως' - 'otherwise'.

2. When the English morpheme is not abbreviated, e.g. 'it is not', 'do not', etc., the auxiliary is not aligned and the verb is aligned with the corresponding English word with a 1-1 link wherever possible.
3. When the structure is changed in English: <u>Example</u>: 'οὐδὲ δυναμένους' - 'were unable'; 'μή πείθεσθαι' - 'to refuse obedience'.

6. Verbs

6.1. Auxiliaries, Verbal Groups

Verbal groups are groups formed by more than one verb: this includes auxiliaries, passivization, translations of tense and aspect, etc.
All verbal groups are aligned as a whole regardless of the number of words. The type of link can be N-N, N-1, or 1-N. <u>Example</u>: 'τελέεσθαι' - 'will be brought to pass'.

6.2. Verbs and Subjects

Since Ancient Greek is a highly inflected language, often the subject of a verb is left implicit and only expressed by means of the personal ending. However, in English the subject of a verb is almost always explicit.
1. When the subject is a separate noun or pronoun in both languages, it is aligned with a 1-1 link.
2. When the subject is not explicit in both languages, only the verb is aligned.
3. When the subject is explicit in English, but implicit in Ancient Greek:
    a. If the English subject is a pronoun, it is aligned together with the verb, with a 1-N link. <u>Example</u>: 'ἔσαν' - 'they were'.
    b. If the English subject is a noun or proper noun that is not expressed in Ancient Greek, it is not aligned and only the verb is aligned.

These rules also apply to infinitives in infinitive+accusative constructs in Ancient Greek.

6.3. Passive verbs translated into active

Verbs are aligned regardless of voice (active, middle, or passive), provided that the English translation bears the same meaning and is correct in the judgment of the annotator. This means that a passive verb in Ancient Greek may be aligned with an active verb in English. <u>Example</u>: 'φάανθεν' - 'shone'.

6.4. Compound verbs with prepositions

Compound verbs (verbs constructed with prepositions in Ancient Greek, e.g. περι-πλέω) are aligned with the corresponding English expression, when they are equivalent in meaning.
<u>Example</u>: 'περιπλέω' - 'I sail around'.

In cases of tmesis (a compound verb that appears with preposition separated at the beginning of phrase), the same rule applies, with a N-N link.

Extra prepositions in English: when an English translation contains an extra preposition that does not explicitly appear in Ancient Greek, the preposition is aligned as part of the verb group, if the English requires it in order to make sense. This principle also applies for auxiliaries and similar verb groups. <u>Example:</u> 'διῄρηνται' - 'divided into'.

7. Participles in Ancient Greek

7.1. General rules

1. Participles translated with an attributive or implicit form, e.g. -ing form in English: the participle is aligned with the corresponding word/s in English. The noun that agrees with the participle is aligned separately, with the corresponding translation in English. <u>Example</u>: 'φέρων' - 'bearing'.
2. Participles translated as a dependent clause in English: the participle is aligned with the whole expression in the English translation (1-N). The noun that agrees with the participle is aligned separately, with the corresponding translation in English. <u>Example</u>: - 'κιὼν' - 'when he had gone'.
3. Substantive and attributive participles: when translated explicitly (mostly as a relative clause), the participle is aligned with the full explicit translation. <u>Example</u>: τοὺς μὲν ἔχοντας - some people who had: 'τοὺς μὲν' - 'some people'; 'ἔχοντας' - 'who had'.
   a. When an article is present, it is aligned together with the participle (N-N), unless the translation does not bear equivalent meaning. Example: πάντες δὲ οἱ καλούμενοι - all who are called: 'πάντες - all'; 'οἱ καλούμενοι' - 'who are called'.

7.2. Genitive absolute

The participle forming a genitive absolute is aligned with the corresponding translation in English, following the general rules of participles (1-1 or 1-N link type). The noun that agrees with the participle is aligned separately, with the corresponding translation in English. <u>Example</u>: ἐμεῦ ζῶντος - while I live: 'ἐμεῦ' - 'I', 'ζῶντος' - 'while live'.

8. Prepositional phrases

Every component of a prepositional phrase is aligned separately with the corresponding English translation. Sometimes a preposition will be translated with an imperfect equivalent in English: such cases should be evaluated by the annotator to decide whether the translation is just partial or incorrect, and aligned accordingly. <u>Example</u>: 'ἐν' - 'in', 'τῇ' - 'the', 'νήσῳ' - 'island'.

9. Case uses

Inflected nouns (including substantivized adjectives, pronouns, etc.) are aligned with the corresponding English grammatical translation, provided that it bears equivalent meaning (1-N link). Example: θυμοῦ - 'of heart'.

Any modifier of the noun (attributes, determiners, etc.) does not obey this rule, since its inflection is considered as originated from agreement with the noun.

There may be cases where the English preposition used is incorrect in the judgement of the annotator: in such cases, the annotator may align just the Ancient Greek word with the lexical translation and omit the preposition. Example: θυμοῦ - against the heart: ø - 'against', ø - 'the', θυμοῦ - 'heart.'

9.1. Vocatives

Nouns in the Vocative case in Ancient Greek are aligned to the corresponding translation in English. The exclamation (Ancient Greek 'ὦ' or Eng. 'oh') is aligned to the corresponding word when present in both languages. When absent in one of the two, it is not aligned.

10. Pronouns and anaphoric structures

Pronouns are aligned as a 1-1 link with the corresponding word in English when they are equivalent in meaning.
Pronouns or other determiners (e.g. possessives, demonstratives, etc.) that replace a noun in one of the two languages but not in the other are not aligned. Example: 'he' and 'Ἀγαμέμνων' are not aligned (even though it's the same person in the context).

11. Determiners (articles, etc.)

1. When determiners are present in both languages and are equivalent in meaning, they are aligned with a 1-1 link. Examples:
   a. 'τις ἄνθρωπος' transl. 'a man'/'some man'/'any man': 'τις' - 'a/some/any', 'ἄνθρωπος' - 'man'.
   b. 'ὁ ἄνθρωπος' transl. 'the man'/'this man'/'that man': 'ὁ' - 'the/this/that', 'ἄνθρωπος' - 'man'.
   c. Possessive adjectives are included among the options of acceptable translations. Example: τῶν ἵππων - of their horses: 'τῶν' - 'their', 'ἵππων' - 'of horses'.
2. When a determiner is present in both languages, but they do not bear equivalent meaning, they are not aligned and only the noun is aligned. Example: 'ὁ ἄνθρωπος' transl. 'a man': 'ὁ' - ø, 'ἄνθρωπος' - 'man'.
3. When the determiner is present in English but not in Ancient Greek: it is aligned together with the noun to the corresponding Ancient Greek word, provided that they are semantically equivalent (N-1 link). Example: 'ἄνθρωπος' - 'the man'.

4. When the determiner is present in Ancient Greek but not in English: it is not aligned, unless absolutely necessary, i.e. in nouns that require the article in Ancient Greek but not in English. Examples: 'ὁ Κῦρος' - 'Cyrus', 'τῇ Εὐρώπῃ' - 'Europe'. Incl. substantivizations of adjectives and verbs in certain contexts. Examples: 'τὸν αἰσχρόν' - 'moral wrong', 'τὸ μανθάνειν' - 'learning'.
5. Phrasal translations are aligned to the group of determiner + noun, when they bear equivalent meaning. Example: 'τις ἄνθρωπος' - 'someone'.

12. Pronominal uses of articles in Ancient Greek

Sometimes Ancient Greek uses articles with particles like δὲ as pronouns. English translations of this phenomenon differ quite significantly:
1. Pronominal equivalents: the whole expression in Ancient Greek is aligned with the corresponding English if they are equivalent in meaning. Example: 'Οἱ δὲ' - 'they', 'those', 'those people', etc.
2. Pronominal uses + prepositional phrases:
   a. When translated literally, the pronominal expression is aligned with the corresponding English expression, and the prepositional phrase is aligned following the rules of prepositional phrases. Example: Οἱ δὲ ἐν τῇ νήσῳ - Those in the island: 'Οἱ δὲ' - 'those'; 'ἐν' - 'in'; 'τῇ' - 'the'; 'νήσῳ' - 'island'.
   b. When translated with a close semantic equivalent: both groups are aligned, following the rules of idioms, with a N-N link. Example: 'Οἱ δὲ ἐν τῇ νήσῳ' - 'the islanders'.
   c. When translated with the addition of words not present in Ancient Greek: the additional words are not aligned, the rest is aligned according to the guidelines. Example: Οἱ δὲ ἐν τῇ νήσῳ - the inhabitants in the island: 'Οἱ δὲ' - ø; 'ἐν' - 'in'; 'τῇ' - 'the'; 'νήσῳ' - 'island'.

13. Adjectives and Adverbs

Adjectives and adverbs are aligned as individual lexical units, regardless of their function in the context (attributes, predicatives, nouns, etc.).

Comparatives and superlatives, which are usually single words in Ancient Greek, are aligned with the corresponding English translation, with a 1-1 link or, when necessary, a 1-N link. Example: 'κάκιστος' - 'ugliest' / 'the ugliest' / 'the most ugly'.

14. Functional words and Particles

As a general principle, the functional words which are required for grammatical reasons are aligned when they have an equivalent counterpart in the other language. They are aligned with a 1-1 link if there is an equivalent English translation.
Exceptions and special cases:

1. Ἄν (when not equivalent of ἐάν):
    a. + Verbs: if the English translation of this group bears an equivalent meaning, then the group is aligned with a N-N link to the whole English expression. If that is not the case, ἄν is not aligned. <u>Example:</u> 'ἂν δοκῶσι' - 'seem likely'.
    b. In fixed phrasal constructions, e.g. ὅπως ἄν: ἄν is aligned within the phrasal construction (N-1 or N-N link). <u>Example:</u> 'ὅπως ἄν' - 'just as'.
2. τε, μέν, δέ, γάρ, γε, δή et sim.: are aligned 1-1 to the corresponding English translation, if it is explicit, clearly identifiable, and isolated from other words.
    a. τε καί group: when the translation is only one word, only καί is aligned. <u>Example</u>: 'τε καί': 'τε' - ø; 'καί' - 'and'.
    b. Combined uses (e.g. μέν + δέ, οὐδέ + οὐδέ, etc.): particles in combination can be aligned as a group to the corresponding English translation if they cannot be omitted without changing the meaning of the expression, and if the English translation bears an equivalent meaning to their function in combination. <u>Example</u>: 'μέν...δέ' - 'on the other hand'/'on the one hand...on the other hand', 'καί...καί' - 'both...and', 'οὔτε...οὔτε' - neither...or.

Literature

Patrik Lambert, Adrià de Gispert, Rafael Banchs, and José B. Mariño. 2005. Guidelines for word alignment evaluation and manual alignment. *Language Resources and Evaluation.* 39:267-285.

João Graça, Joana Paulo Pardal, Luísa Coheur and Diamantino Caseiro. 2008. Building a golden collection of parallel Multi-Language Word Alignments. *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08).*

I. Dan Melamed, Annotation Style Guide for the Blinker Project. 1998. *IRCS Technical Reports Series*. 53.