

About this Guide

This annotation style guide was created by Valeria Irene Boano (valeriairene.boano01@icatt.it). The languages involved are Latin and English.

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

The style guide was created taking as a starting point the guidelines for Ancient Greek created by Chiara Palladino and Farnoosh Shamsian (2022).

Whenever the guiding principles for Ancient Greek applied also to Latin, they were not modified. Exceptions and guiding principles for different constructions were added *ex novo* whenever necessary.

The author manually aligned a text sample consisting of part of the *Bellum Alexandrinum* of Pseudo-Caesar (559 sentences, 10464 words), following the guidelines for Ancient Greek. Whenever these were not relevant or sufficient, different principles were adopted and written in this 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 Latin 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.

Exceptions:

1. The apostrophe in the Saxon genitive, when aligned with a corresponding word/group of words in Latin.
2. The full stop in the abbreviations is aligned
 - a. with a N-1 link when the English translation is extended. Example: 'M.' - 'Marcus'.
 - b. with a N-N link when the English translation is abbreviated. Example: 'P.' - 'P'.

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.

As few words as possible, as long as they bear the same meaning as a group and cannot be isolated in the translation, are aligned together with a N-N link.

Examples: dative of possession: 'est eis' - 'they have'.

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), if 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: *neque putabat* - and he didn't think: 'neque' - 'and didn't', 'putabat' - 'think'.
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: Example: 'neque potuerunt' - 'they were unable'

6. Coordination

If two Latin nominal conjuncts are translated by other types of constructions, provided that they are clearly identifiable and isolated, and that the translation is correct in the judgement of the annotator, they are linked to the English translation with a 1-1 link. The coordinating

conjunction is not aligned. Example: *contignatione ac materia* - with wooden floorings: '*contignatione*' - 'with floorings'; '*materia*' - 'wooden'; '*ac*' - \emptyset .

7. Enclitics

Whenever in Latin there is an enclitic word fused together with another word (i.e. *-que*, *-ne*, *-cum* ...), the Latin form is linked to the English corresponding multiple words with a 1-N link.

Example: '*seque eiecerunt*' - 'and rushed forth'.

If the value of the enclitic is not expressed in English, the Latin word is aligned anyway with the corresponding English word with a 1-1 alignment.

8. Verbs

8.1. 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. Example: '*credebant*' - 'they did not trust'

8.2. Verbs and Subjects

Since Latin 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 Latin:
 - a. If the English subject is a pronoun, it is aligned together with the verb, with a 1- N link. Example: '*adfirmabat*' - 'he declared'.
 - b. If the English subject is a noun or proper noun that is not expressed in Latin, it is not aligned and only the verb is aligned.

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

Exception: in passive periphrastic construction (see par. 8.5), whenever the English active translation provides a subject, which is not explicitly expressed by a Latin dative of agent (nor by any other Latin word), the English subject is not aligned, even if it is a pronoun.

8.3. Passive verbs translated into active and other changes of morphological features

Verbs are aligned regardless of voice (active, middle, or passive), provided that the English translation bears the same meaning and is correct in the judgement of the annotator. This means that a passive verb in Latin may be aligned with an active verb in English. Example: '*agebatur*' - 'was taking place'.

Exception: when passive periphrastic construction (see par. 8.5) is translated into active, if the English translation provides a subject, which is not explicitly expressed by a Latin dative of agent (nor by any other Latin word), the English subject is not aligned.

As a general principle, whenever a morphological feature, such as person or number, changes in the translation, the verb is aligned only if the meaning is the same according to the judgement of the annotator.

8.4 Compound verbs with prepositions

Compound verbs (verbs constructed with prepositions in Latin, e.g. *circum-vehor*) are aligned with the corresponding English expression when they are equivalent in meaning. Example: '*circumvehitur*' - 'he sails around'.

Extra prepositions in English: when an English translation contains an extra preposition that does not explicitly appear in Latin, 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. Example: '*prohibent*' - 'prevent from'.

8.5 Passive periphrastic

The Latin construction of passive periphrastic is aligned with the English translation with a N-N link. Example: '*esse vincendum*' - 'must win'.

Whenever the English active translation provides a subject, which is not explicitly expressed by a Latin dative of agent (nor by any other Latin word), the English subject is not aligned. This constitutes an exception from the general principle expressed at par. 8.2.

9. Participles in Latin

9.1 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. Example: '*pertinentes*' - 'extending'
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. Example: '*laborantibus*' - 'when they were struggling'
3. Substantive and attributive participles: when translated explicitly (mostly as a relative clause), the participle is aligned with the full explicit translation. Example: '*inceptum*' - 'that had been started'.

9.2 Ablative absolute

The participle forming an ablative 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. Example: *suscepto officio* - when he took up the office: '*suscepto*' - 'when he took up'; '*officio*' - 'the office'.

10. 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. Example: '*ab*' - 'from'; '*rege*' - 'the king'.

11. 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: '*telorum*' - 'of weapons'.

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 judgment of the annotator: in such cases, the annotator may align just the Latin word with the lexical translation and omit the preposition.

11.1 Vocatives

Nouns in the Vocative case in Latin are aligned to the corresponding translation in English.

12. 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: '*eum*' and 'Caesar' are not aligned (even though it's the same person in the context).

Exception: if the pronoun is translated by a general common name, and, according to the judgement of the annotator, this translation is correct outside the specific context, it is aligned. Example: '*nostrī*' - 'our men'.

12.1 Relative pronouns

When the Latin relative pronoun is not aligned with an English relative pronoun:

- a. If the English translation is another type of pronoun or a demonstrative determiner, it is aligned. Example: '*quibus*' - 'these'.
- b. If the Latin relative pronoun is translated as a causal clause, a purpose clause, a consequence clause, or in general by an adverbial clause, provided that the English translation bears the same meaning and is correct in the judgement of the annotator, the pronoun is aligned with the English subordinating conjunction. Example: '*qui*' - 'to'.

13. Articles

Since the article is present in English but not in Latin, it is aligned together with the noun to the corresponding Latin word, provided that they are semantically equivalent (N-1 link). Example: '*aedificia*' - 'the buildings'.

14. 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 Latin, are aligned with the corresponding English translation, with a 1-1 link or, when necessary, a 1-N link. Example: '*celeberrimis*' - 'the most frequented'.

15. Numerals

Whenever a Latin numeral is composed of two or more parts, it is aligned to the English counterpart with a N-1 link. Example: '*tricensima septima*' - 'thirty-seven'.

16. Subordinating conjunctions

If a Latin subordinating conjunction is translated by an English verbal expression, provided that it is correct in the judgment of the annotator, it is aligned with a 1-N link. Example: '*ne*' - 'to prevent from'.

17. 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. *et, aut, vel, neque* et sim.: are aligned 1-1 to the corresponding English translation, if it is explicit, clearly identifiable, and isolated from other words.
 - a. *et + et* group, *vel + vel* group, *aut + aut* group: when it is translated by just one word, only the second Latin word is aligned. Example: '*et... et*': 'et'- ∅; '*et*' - 'and'.
 - b. 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. Example: '*et...et*' - 'both...and'; '*aut...aut*' - 'neither...or'.
2. '*ne...quidem*' group: '*ne*' - 'not'; '*quidem*' - 'even'.

Literature and references

Palladino, Chiara, & Shamsian, Farnoosh. (2022). Translation Alignment: Ancient Greek to English. Annotation Style Guide and Gold Standard. (1.0) [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.7362097>

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.