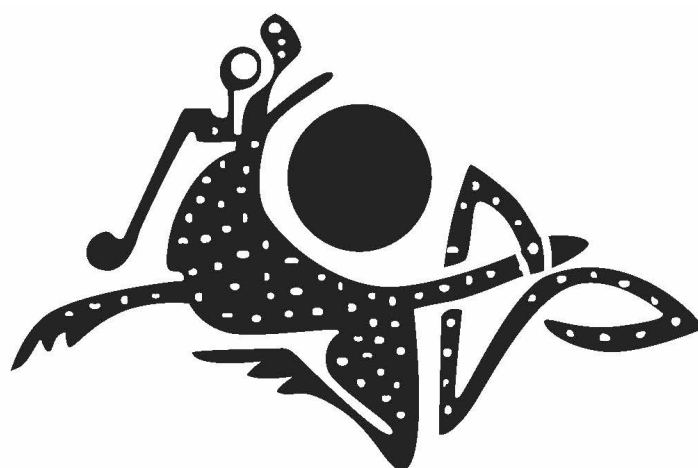# ParCorOEv2
An open access annotated
parallel corpus Old English-English
2021

MANUAL 2: Style guide
Version 1.0

Martín Arista, Javier, Sara Domínguez Barragán, Laura García Fernández,
Esaúl Ruíz Narbona, Roberto Torre Alonso & Raquel Vea Escarza (comp.)

This document contains the edition manual of ParCorOEv2.

Titles
Titles and subtitles of text sections (books, chapters, chapter parts, etc.) are included into the Anglo-Saxon version.

Characters
The characters used in the corpus include the Roman alphabet plus the ash and the eth, both small and capital.
In the corpus, the letter eth subsumes the letters eth and thorn.
In the corpus, the letter ash represents the letters æ and ę.
Capital and small letters are rendered as in the original.
Small capitals are given as small case.
Vocalic length is not marked.
Palatalisation is not marked.
Other non-linguistic symbols, such as a cross, are disregarded.

Numbers
Roman numbers are used as in the text of reference.
The dots that embrace Roman numbers are deleted.
The Arabic numbers in the chronicles are kept as in the original.
Roman numbers are given as in the original, either capital letters or small case. Roman numbers is small capitals are converted into small case.

Word separation
There is no hyphen between the two parts of a compound or a prefix and its base of derivation.
Prefixed (ge-, etc.) or compound words given as two separate words in the Anglo-Saxon edition are rendered as one word. In some editions, the space between the prefixed element and the base is narrower than the usual space between two independent words. This can guide the re-union of prefixes (or adjuncts of compounding) and bases. A comparison with the DOEC or YCOE may identify the discrepancies as to word separation. Ultimately, the word separation in Nerthus is the only that guarantees the automation of the annotation process and should be followed. A note must be included explaining that an alternative source has been used for word separation and specifying, on a general basis, the different types of correspondence between the Anglo-Saxon edition of reference and the alternative edition on which word separation is based.

Abbreviations
As a general rule, no abbreviations are used (& for and, þ for þæt, etc.). They are replaced with the corresponding full forms.
The full forms replacing abbreviations are rendered in lower-case unless they appear between fully capitalised words or immediately follow a full stop or a question mark

Punctuation
Punctuation is kept as in the original in the fields Fragment and Translation.

The high dot that some editions use to represent the comma is replaced by a comma. This should be registered in the field Notes.

Punctuation is removed in the fields ConcTerm and Gloss.

The field ConcTerm cannot contain any punctuation mark. If a punctuation mark precedes the ConcTerm, it is inserted into the field Prefield, on the right end; if a punctuation mark follows the ConcTerm, it is inserted into the field Postfield, on the left end.

Latin intercalations

Short fragments in Latin are registered, unless they duplicate the Anglo-Saxon text. Latin intercalations are rendered in italics in the fields Fragment and Translation.

Added text

Text added by editors and indicated by means of angles, brackets or inverted commas is rendered without such marks. The addition (between angles or brackets, as in the original) is included into the field Notes on tokenisation.

Text intercalations

Exceptionally, text can be added to the fragment from a source other than the reference edition. This is not marked in the field Fragment, but in the field Text intercalation with the following format:

||added text|| (editor year: page). The editions from which additional text is inserted must be uploaded to the layout Text information.

Translation intercalations

Exceptionally, translation selections can be added to the translation from a source other than the reference edition. This is not marked in the field Translation, but is registered in the field Translation intercalation with the following format:

||added translation|| (editor year: page). The editions from which additional text are inserted must be uploaded to the layout Text information.

Explanations for translation

If translators add explanations to the English version that do not have a textual correlate in the Anglo-Saxon version (this is usually indicated by means of italics, brackets or inverted commas), such explanations are excluded from the translation into English. A note must be included explaining, on a general basis, how these explanations have been handled.

Content discrepancies

If discrepancies arise between the Anglo-Saxon version and the modern one as to chapter or section number, year, or any other aspect, the original version is not modified on the grounds of the updated information provided by the translation. This means that certain items of information may not coincide between the concordance and the translation.

Glosses

A gloss is a key to a text written in a language of which potential readers do not have enough linguistic competence. In the context of a parallel corpus, the gloss represents

an intermediate step between the original text and the translation. The gloss is a version of the translation that is aligned word for word. Moreover, the gloss guarantees the alignment at word level, since the translation can only guarantee the alignment at text or token (fragment) level. In practical terms, the gloss should allow potential users to read the Anglo-Saxon text, at least, for gist. The sorted glosses give rise to the parallel text.

As editors, we are not constrained by the translation. The aim of glosses is not to reproduce the translation slot by slot. This is impossible given the differences between the source language and the target language. If we bear in mind that a translation is a version of the original text, it is easier to follow the translation with some freedom.

As editors, we are not constrained by the glossary, either. The glossary provides extremely useful information, but we are free to supply information missing in the glossary or to improve it, particularly when deciding on meaning in context.

Two types of quantitative asymmetry between the source text and the gloss can be expected:

-The source text takes up more slots than the gloss.
-The source text takes up fewer slots than the gloss.

This means that some gloss slots may be empty or contain more than one word. These instances are listed in the ASYMMETRY BANK.

The focus of glosses is on lexical meaning. Nouns, pronouns, adjectives, verbs and adverbs must be included in the gloss so that the text can be read without the help of the glossary. Conversely, the focus of glosses is not on grammatical meaning. Grammatical categories may be left out without the understanding of the text being dramatically diminished.