# Designing a Corpus Database to Enhance the Learning of Old English Language

**Authors :** Raquel Mateo Mendaza, Carmen Novo Urraca

**Abstract :** The current paper presents the elaboration of a corpus database that aligns two different corpora in order to simplify the search of information both for researchers and students of Old English. This database comprises the information contained in two main reference corpora, namely the Dictionary of Old English Corpus (DOEC), compiled at the University of Toronto, and the York-Toronto-Helsinki Parsed Corpus of Old English (YCOE). The first one provides information on all surviving texts written in the Old English language. The latter offers the syntactical and morphological annotation of several texts included in the DOEC. Although both corpora are closely related, as the YCOE includes the DOE source text identifier, the main problem detected is that there is not an alignment of texts that allows for the search of whole fragments to be further analysed in terms of morphology and syntax. The database proposed in this paper gathers all this information and presents it in a simple, more accessible, visual, and educational way. The alignment of fragments has been done in an automatized way. However, some problems have emerged during the creating process particularly related to the lack of correspondence in the division of fragments. For this reason, it has been necessary to revise the whole entries manually to obtain a truthful high-quality product and to carefully indicate the gaps encountered in these corpora. All in all, this database contains more than 60,000 entries corresponding with the DOE fragments annotated by the YCOE. The main strength of the resulting product is its research and teaching implications in the study of Old English. The use of this database will help researchers and students in the study of different aspects of the language, such as inflectional morphology, syntactic behaviour of given words, or translation studies, among others. By means of the search of words or fragments, the annotated information on morphology and syntax will be automatically displayed, automatizing, and speeding up the search of data.

**Keywords :** alignment, corpus database, morphosyntactic analysis, Old English
**Conference Title :** ICCDC 2020 : International Conference on Corpus Design and Compilation
**Conference Location :** Paris, France
**Conference Dates :** December 28-29, 2020