Combining corpora to improve the learning of Old English language

Carmen Novo Urraca
University of La Rioja, Spain

Raquel Mateo Mendaza
University of La Rioja, Spain

Resumen

This paper focuses on the elaboration of a corpus database in Old English to improve the learning experience of undergraduates of English Studies at university. Nowadays, the introduction of corpus linguistics in learning has involved a step forward in the study of languages and language description (Aston 2001). Corpora comprise relevant data on language use and show it in an automatic way allowing the user to make searches on different grounds. Corpus linguistics has also been applied to historical languages and several corpora of Old English and English diachrony are available for students and researchers. Many authors have pointed out the pedagogical benefits that the use of corpus linguistics in language learning, specially, those activities conducted under the Data-Driven learning approach (Johns 1991). This direct approach allows students to be the main users of the corpus and to establish their own findings, whereas the teacher is seen as a mediator or guide. With this purpose, the corpus presented to the students should be as complete and useful as possible. Regarding Old English, the most relevant corpora are the Dictionary of Old English Corpus (DOEC; Healey et at. 2009) and the York-Toronto-Helsinki Parsed Corpus of Old English (YCOE; Taylor et al. 2003). The former contains most of the surviving manuscripts written in Old English, whereas the second one provides the morphosyntactic annotation of some of the texts displayed by DOEC. In this respect, the aim of this work is to design a tool which combines both corpora to facilitate searches where all the information is displayed together. With this purpose, a corpus database for in-class academic purposes has been created in which the DOEC fragments are semi-automatically aligned to the tagged information of the YCOE. In this way, students can be assigned advanced tasks, based on the Data-Driven Learning approach, that let them understand how Old English language works and, at the same time, reach their own conclusions on frequency, morphology, syntax, semantics and diachrony issues.

Palabras clave: Corpus linguistics; Data-driven learning; Historical linguistics; Old English; Teaching.

Referencias

Aston, G. (2001). Learning with corpora: An overview. In G. Aston (ed.), *Learning with corpora* (pp. 7–45). Houston: Athelstan.

Healey diPaolo, A., Price Wilkin, J., X. Xiang (2009) (Ed.). *The Dictionary of Old English Web Corpus.* Toronto. Dictionary of Old English Project.

Johns, T. (1991). Should you be persuaded – two samples of data-driven learning materials. *Classroom Concordancing: ELR Journal, 4,* 1-16.

Taylor, A., Warner, A., Pintzuk, S., F. Beths (2003) (Ed.). *The York–Toronto–Helsinki Parsed Corpus of Old English Prose*. England: University of York.