WORSHOP SUBMISSION: Recent advances in computational linguistics: New methods and results

Treebank of Old English with Universal Dependencies. Categories, relations and enhanced dependencies

Universal dependencies (UD) is a framework of morphological and semantic-syntactic annotation that allows for cross-linguistic comparison on an explicit and consistent basis (Nivre et al. 2016). This framework has already been applied to over one hundred languages (Nivre et al. 2020), of which there are around two hundred available treebanks, or databases of structured linguistic data with morphological annotation and syntactic parsing (Taylor et al. 2003). The annotation system of UD consists of universal part-of-speech tags (UPOS), language-specific part-of-speech tags (XPOS), universal morphological features (Feats), lemmas, dependency heads and universal dependency labels. Against this background, the aim of this paper is to lay the foundations of the application the framework of UD to Old English. This aim raises the following research questions: (i) which lexical and morphological categories of the UD inventory are relevant to Old English?; (ii) which dependency relations of the UD taxonomy can be found in Old English sentences?; and (iii) is it necessary to put forward enhanced dependencies in order to account for Old English semantic-syntactic relations? The data will be drawn from *ParCorOEv2. An open access annotated parallel corpus Old English-English* (ParCorOEv2), which comprises around 110,000 records. The analysis will check the dependencies of the UD framework against the information available from ParCorOEv2: token (inflectional form), tagging (file number, lemma, lexical category, inflectional category and gloss), a concordance showing the prefield and the postfield of the concorded term, the translation of the fragment into Present-Day English and the sources of the Old English and the Present-Day English texts. Conclusions are expected as regards the relevance of the inventories of lexical and grammatical categories. Conclusions will also be drawn on the applicability of the UD inventory of dependency relations and the need for enhanced dependencies, given changes on the cline of grammaticalisation like the generalisation of prepositional government, the inflection of the past participle in passives, the auxiliarisation of preterite-present verbs, the development of the telic function in verbal compounds and derivatives, and the status of formal subjects.

References

Martín Arista, J., S. Domínguez Barragán, L. García Fernández, E. Ruíz Narbona, R. Torre Alonso and R. Vea Escarza (comp.). 2021. *ParCorOEv2. An open access annotated parallel corpus Old English-English*. Nerthus Project, Universidad de La Rioja, www.nerthusproject.com.

Nivre, J., M de Marneffe, F. Ginter, Y. Goldberg, J. Hajič, D. C. Manning, R. McDonald, S. Petrov, S. Pyysalo, N. Silveira, R. Tsarfaty and D. Zeman. 2016. Universal Dependencies v1: A Multilingual Treebank Collection. In D. Zeman (ed.), *Proceedings of the Tenth International Conference on Language Resources and Evaluation (REC'16)*. Marseille: European Language Resources Association. 1659-1666.

Nivre, J., M. de Marneffe, F. Ginter, J. Hajič, C. D. Manning, S. Pyysalo, S. Schuster, F. Tyers and D. Zeman. 2020. Universal Dependencies v2: An Evergrowing Multilingual Treebank Collection. In N. Calzolari, F. Béchet, P. Blache, K. Choukri, C. Cieri, T. Declerck, S. Goggi, H. Isahara, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk and S. Piperidis (eds.),

Proceedings of the 12th Language Resources and Evaluation Conference. Marseille: European Language Resources Association. 4034-4043.

Taylor, A., M. Marcus and B. Santorini. 2003. The Penn Treebank: An overview. In A. Abeillé (ed.), *Treebanks: Building and Using Parsed Corpora*. Dordrecht: Kluwer Academic Publishers. 5-22.