

Universal Dependencies: Principles and Tools

Daniel Zeman

Institute of Formal and Applied Linguistics
Charles University, Czech Republic
Faculty of Mathematics and Physics

Abstract

Universal Dependencies is an international community project and a collection of morphosyntactically annotated data sets (“treebanks”) for more than 100 languages. The collection is an invaluable resource for various linguistic studies, ranging from grammatical constructions within one language to language typology, documentation of endangered languages, and historical evolution of language. In the tutorial, I will first quickly show the main principles of UD, then I will present the actual data and various tools that are available to work with it: parsers, batch processors, search engines and viewers.

Biography. Daniel Zeman is a senior researcher and lecturer at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University in Prague. His research interests range from natural language parsing and morphology to low-resource language processing and linguistic typology. He is one of the co-founders and main coordinators of the Universal Dependencies project.