

The Case for Universal Dependencies

Christopher Manning
Stanford University
Department of Computer Science
manning@cs.stanford.edu

Universal Dependencies is a recent initiative to develop a linguistically informed, cross-linguistically consistent dependency grammar analysis and treebanks for many languages, with the goal of enabling multilingual natural language processing applications of parsing and natural language understanding. I outline the needs behind the initiative and how some of the design principles follow from these requirements. I suggest that the design of Universal Dependencies tries to optimize a quite subtle trade-off between a number of goals: an analysis which is reasonably satisfactory on linguistic grounds, an analysis that is reasonably comprehensible to non-linguist users, an analysis which can be automatically applied with good accuracy, and an analysis which supports language understanding tasks, such as relation extraction. I suggest that this is best achieved by a simple, fairly spartan lexicalist approach, which focuses on capturing a level of analysis of (syntactic) grammatical relations, something that can be found similarly defined in many theories of syntax. We take hope from the fact that already many people, coming from quite different syntactic traditions, have felt that Universal Dependencies is near enough to right that they can join the effort and contribute. However, the current proposal is certainly not perfect, and I will also touch on some of the thorny issues and how the current standard might yet be improved.