

From Prague Structuralism to Treebank Annotation

Eva Hajičová, Charles University in Prague, Czech Republic

Abstract

When contemplating about the relationships between treebanking and linguistic theory, we believe it is useful to look a little bit further back and to consider (also) the roots of linguistic science rather than to restrict the attention to the theoretical aspects of what might be viewed as contemporary theoretical developments. In our contribution, we focus on three tenets of the structural and functional approach of the Prague School of Linguistics as reflected in the writings of Praguian linguists of the classical period, namely (i) the concept of markedness and the opposition of the core and periphery of language system, (ii) the functional viewpoint, and (iii) the frequently assumed distinction between the “grammatical” and “semantic” structure of the sentence, with the emphasis on the underlying syntactic relations with the verb as the nucleus. These three aspects are discussed both in their original, fundamental treatments as well as in view of the possibility and usefulness of their application to treebank annotation on different layers. In this connection, such questions are posed as what linguists (be they theoretically, computationally, or corpus- oriented) have expected to get from treebank annotation, in which respects these expectations have been met and in which we have failed or at least temporarily failed, which blind alleys we have fallen in on the way, which unexpected problems we have encountered. In the end, we will consider what we assume can(not) be done without treebanking (tree representations) and we will try to summarize what the treebanks are good for in spite of all the limits, insufficiencies or even (temporary) failures.