

Manfred Stede

Text parsing in a multi-level representation framework

Multi-level analysis (MLA) aims at explaining the coherence of a text not by assigning "the" single discourse structure to it but by defining a set of separate levels of analysis, which collectively capture its "textuality". The talk explains the set of levels used for manual annotation in the "Potsdam Commentary Corpus", as well as the software infrastructure enabling this effort. Then, "annotation mining" is introduced as the task of gathering information on a more complex level of description from a set of less complex levels. Finally, the approach is carried over to the problem of automatic text parsing, where it translates to the task of determining constraints and preferences (i.e., possibly-partial results) on the various levels of analysis, and exploiting inter-relationships for achieving the particular application goal.