Çağri Çöltekin and Ben Campbell: TüBa-DZ to UD conversion and UD error detection

This presentation covers two studies related to Universal Dependencies. The first part of this presentation will summarize the automatic conversion of the TüBa-DZ constituency treebank into Universal Dependencies. The conversion process will be discussed, including dependency extraction, POS conversion, head-finding rules, topological fields, as well as unusual cases that were unable to be properly converted into the Universal Dependencies annotation scheme.

The second part covers error detection for Universal Dependencies treebanks for three languages: English, German, and Finnish. The approach used was to compare like strings in the treebanks and see where discrepancies in the dependency relations occurred. In general there were three types of results: 1) false positives, ie. discrepancies that turned out not be errors when taking into account the larger context of the sentence, 2) true positives, ie. clear error(s) in the annotation of at least one of the sentences, and 3) borderline or ambiguous cases where two or more of the differing dependency relations could be argued to be correct and stricter guidelines would be necessary in order to make a clear determination of what the correct dependency relation is.