

io-port 06539039

Tiedemann, Jörg; Agić, Željko

Synthetic treebanking for cross-lingual dependency parsing.

J. Artif. Intell. Res. (JAIR) 55, 209-248 (2016).

Summary: How do we parse the languages for which no treebanks are available? This contribution addresses the cross-lingual viewpoint on statistical dependency parsing, in which we attempt to make use of resource-rich source language treebanks to build and adapt models for the under-resourced target languages. We outline the benefits, and indicate the drawbacks of the current major approaches. We emphasize *synthetic treebanking*: the automatic creation of target language treebanks by means of annotation projection and machine translation. We present competitive results in cross-lingual dependency parsing using a combination of various techniques that contribute to the overall success of the method. We further include a detailed discussion about the impact of part-of-speech label accuracy on parsing results that provide guidance in practical applications of cross-lingual methods for truly under-resourced languages.

doi:10.1613/jair.4785