

Jelke Bloem (Universiteit van Amsterdam):

Corpusonderzoek met behulp van automatisch gegenereerde syntactische annotatie

In deze lezing ga ik in op het gebruik van automatische syntactische annotatie bij Nederlands corpusonderzoek, op basis van ervaringen in mijn onderzoek naar werkwoordsvolgorden. Grote hoeveelheden tekst kunnen automatisch geannoteerd worden, maar de parsecomputer maakt fouten, terwijl correcte annotatie bij taalkundig onderzoek nogal belangrijk is. Hoe erg is dit, en hoe kunnen we een beeld krijgen van de fouten die de parser maakt? Het blijkt dat er een aantal trucs zijn om de kwaliteit van de automatische annotatie voor een specifieke onderzoeksvraag te bepalen, waardoor we met grotere zekerheid uitspraken kunnen doen op basis van automatisch geannoteerde corpusdata.