The proposed poster is concerned with the influence of frequency-based and probabilistic factors on the placement of filled as well as unfilled pauses and discourse markers.

Recent studies have established links between frequently used sequences of elements and the places of occurrence of hesitations (Bybee 2007; Erman 2007; Fox, Maschler and Uhmann 2010; Kapatsinski 2005; Shriberg and Stolke 1996; Tily et al. 2009). My current project investigates on a theoretical level in how far probabilistic factors compete with other attractors of pauses and discourse markers, such as syntactic boundaries, and on a methodological level whether more complex probabilistic scores such as the Mutual Information Score (MI) and the upcoming Lexical Gravity G (Daudaravičius and Marcinkevičienė 2004; Gries and Mukherjee forthcoming), which takes type as well as token frequencies into account, are more accurate predictors than simpler measures such as transitional probability and bigram frequency. The proposed poster shows first results from a pilot study based on Switchboard NXT.

References


