Combining Referring Expression Generation and Surface Realization: A Corpus-based Investigation of Architectures

Sina Zarriess (Universität Bielefeld)

In recent data-driven generation research, the tasks of linearization or surface realization and referring expression generation (REG) have received increasing attention (Belz and Kow, 2010; Belz et al., 2011).
We suggest a generation task that integrates discourse-level referring expression generation and sentence-level surface realization. We present a data set of German articles annotated with deep syntax and referents, including some types of implicit referents. Our experiments compare several architectures varying the order of a set of trainable generation modules. The results suggest that a revision-based pipeline, with intermediate linearization, significantly outperforms standard pipelines or a parallel architecture.