

Coreference annotation in the Russian Clinical Pear Stories Corpus: annotation scheme and research potential

S. Toldova, M. Bergelson, M. Khudyakova, E. Ivtushok

Coreference is binding of all the noun phrases related to a specific entity. It is one of the most researched discourse features, providing coherence, assigning prominence, and distribution of attention by the members of discursive communication.

A vast majority of research dedicated to coreference analysis is concentrated on written discourse. However, oral discourse is different from the written one, as it can contain various hesitation markers (hesitation pauses, fillers like well, etc.). Those markers can also include name selecting (e.g. Muzhchina sobiraet avocado, net, grushy — The man gathers avocados, no, pears). These features can make coreference annotation of oral discourse more complicated and require a thorough analysis.

The main purpose of this study is a discourse analysis of these features. The analysis is done on oral retellings taken from Russian Clinical Pear Stories corpus (Russian CliPS), a multimedia corpus of narratives produced by speakers with aphasia and right hemisphere damage, as well as neurologically healthy speakers of Russian. The corpus contains retellings of the Pear film, created in 1975 to elicit comparable speech samples with a clear story line and to study the flow of discourse. The nature of the narratives in the corpus is a perfect source for a comparative investigation of discourse micro-structural features (grammatical and lexical phenomena).