Introduction to Coreference Resolution

This chapter summarizes the process of coreference resolution in Slovene texts. A more detailed presentation can be found in the guidelines in the Annotation Guidelines chapter.

In a text, elements that point to the same entity are known as mentions. These mentions can span across various clauses, sentences, or paragraphs. During the annotation process, these mentions are connected to form what is termed a coreference chain. To visually distinguish these, coreference chains are often color-coded. Consider the sentences:

- **[1.a]** Peter ima dva psa. ("Peter has two dogs.")
- **[1.b]** On se velikokrat igra z njima. ("He often plays with them.")

Here, 'Peter' and 'On' are coreferential as they refer to the same individual, just as 'njima' and 'dva psa' point to the same pair of animals. The aim of coreference resolution is to identify and link all mentions. Only those mentions that have a coreferential relationship are marked. Text segments that do not share coreference with another segment are not annotated as such.

The annotation process for coreference is further depicted in the diagram below, where three mentions of a specific entity are illustrated. The sequential links between mentions that refer to the same entity represent coreferential links. These links, along with the mentions they connect, constitute the coreferential chain. Additionally, each mention is accompanied by a set of tags.

coref_shema_en.png

Revision #3
Created 2023-11-13 15:16:02 UTC by Tina Munda
Updated 2023-11-13 15:24:25 UTC by Tina Munda