09 Coreferences

Coreference occurs when several elements within a text—be it words, phrases, or entire sentences—point to the same entity in the real world, outside of language itself. This entity, known as the referent, can represent a wide array of things, including but not limited to people, animals, plants, inanimate objects, places, events, roles, and even abstract concepts.
Identifying coreferenceis complex yet critical for information extraction. It ensures clear understanding of entity relationships within a text, enabling advanced natural language processing applications.

- Introduction to Coreference Resolution
- Annotation Guidelines
- References and Links

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

Annotation Guidelines

This chapter summarizes the annotation guidelines for coreference resolution as applied to Slovene texts. The guidelines are arranged from the latest, up-to-date version to the oldest version.

Version 1.6

Project Development of Slovene in a Digital Environment
ŽITNIK, Slavko, ARHAR HOLDT, Špela, ROBIDA, Nejc in BLAGUS, Neli, 2023: *Smernice za označevanje koreferenčnosti v slovenskem jeziku*: Različica 1.6. Čistopis za projekt Razvoj slovenščine v digitalnem okolju. [DOCX] [PDF] - only in Slovene

References and Links

This chapter compiles relevant references and provides links to projects where coreference resolution has been developed and applied to Slovene texts.

Projects, in which the system has been developed:

ReLDI

Development of Slovene in a Digital Environment

References:

RELDI: Uputstvo za anotiranje koreferenci, Verzija 1.1, Januar 2018.

Martha Palmer, Will Styler, Kevin Crooks, Tim O'Gorman: *Richer Event Description (RED) Annotation Guidelines* v.1.7.

https://github.com/timjogorman/RicherEventDescription/blob/master/guidelines.md

M. Ogrodniczuk, M. Zawisławska, K. Głowińska, and A. Savary, Coreference Annotation Schema for an Inflectional Language, in *Proceedings of the 14th International Conference on Intelligent Text Processing and Computational Linguistics (CICLing 2013)*, 2013, pp. 394–407.

M. Ogrodniczuk, K. Głowińska, M. Kopeć, A. Savary, and M. Zawisławska, Interesting Linguistic Features in Coreference Annotation of an Inflectional Language, in *Proceedings of the 12th China National Conference on Computational Linguistics (CCL 2013) and the First International Symposium on Natural Language Processing Based on Naturally Annotated Big Data (NLP-NABD 2013)*, 2013, pp. 97–108.

S. Pradhan, A. Moschitti, N. Xue, O. Uryupina, and Y. Zhang, “CoNLL-2012 Shared Task: Modeling Multilingual Unrestricted Coreference in OntoNotes,” in Proceedings of the Joint Conference on EMNLP and CoNLL: Shared Task, 2012, pp. 1–40.

M. Recasens, M. A. Martí, and C. Orasan, Annotating Near-Identity from Coreference Disagreements, *Proceedings of LREC 2012*, pp. 165–172, 2012.

M. Recasens, *Coreference: Theory, Annotation, Resolution and Evaluation*, PhD dissertation, 2010.