09 Koreference

O koreferenčnosti govorimo, ko več elementov znotraj besedila (besed, fraz ali povedi) kaže na isto entiteto v resničnem svetu, izven samega jezika. Ta entiteta, ki ji pravimo referent, lahko predstavlja širok nabor stvari, npr. ljudi, živali, rastline, predmete, pa tudi kraje, dogodke, vloge, abstraktne pojme. Odkrivanje koreferenčnosti je zapletena, vendar ključna naloga pri ekstrakciji informacij. Zagotavlja jasno razumevanje odnosov med entitetami znotraj besedila, kar omogoča napredno uporabo v nalogah s področja obdelave naravnega jezika.

- Predstavitev označevanja koreferenčnosti
- Označevalne smernice
- Reference in povezave

Predstavitev označevanja koreferenčnosti

V tem poglavju je strnjeno predstavljeno označevanje koreferenc. Podrobnejšo predstavitev najdete v smernicah v poglavju Označevalne smernice.

Dele besedila, ki se nanašajo na istega referenta, imenujemo **omenitve** (angl. *mentions*). Omenitve se lahko pojavljajo preko različnih stavkov, povedi, odstavkov ali drugih delov besedila. Pri označevanju jih povežemo v t. i. **koreferenčno verigo** (angl. *coreference chain*). Koreferenčne verige nakazujemo z različnimi barvami, kot na primer v naslednjih povedih:

- **[1.a]** Peter ima dva psa.
- **[1.b]** On se velikokrat igra z njima.

Omenitvi *Peter* in *On* se nanašata na istega referenta, tj. isto konkretno osebo, kar pomeni, da sta koreferenčni. Enako velja za omenitvi *njima* in *dva psa*, ki se nanašata na isti par živali.
Naloga pri označevanju koreferenčnosti je identifikacija in povezovanje omenitev. Označuje se samo omenitve, ki so koreferenčne z drugo omenitvijo. Delov besedila, ki niso koreferenčni z nobenim drugim delom besedila, torej ne označujemo.

Označevanja koreferenčnosti prikazuje še spodnja shema. Prikazane so tri omenitve, ki se sklicujejo na določeno entiteto. Povezave med zaporednimi omenitvami, ki se sklicujejo na isto entiteto, so koreferenčne povezave in skupaj z vsemi omenitvami tvorijo koreferenčno verigo. Vsaka omenitev ima določene še lastne oznake.

Koreferenčna veriga entitete



Dodante oznake omenitev:

Type
Copular-N, Copular-S, Appositive, Zero anaphora, Clause, Regular
Referent *(* only for Copular-S, otherwise NA)*
Subject/Object/NA
Discontinuous
True/False
Ellipsis
True/False

Označevalne smernice

V tem poglavju so zbrane označevalne smernice za odkrivanje koreferenčnosti.

Različica 1.6

projekt Razvoj slovenščine v digitalnem okolju
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]

Reference in povezave

V tem poglavju so zbrane relevantne reference in povezave na projekte, v katerih se je označevalni sistem razvijal ter uporabljal.

Projekti, na katerih se je razvijal označevalni sistem

ReLDI

Razvoj slovenščine v digitalnem okolju

Reference

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.