
A stand-off XML-TEI representation of reference annotation

Aria Adli¹, Eric Engel¹, Laurent Romary², Fahime Same¹
¹*Universität zu Köln*, ²*INRIA*

{aria.adli|eric.engel|f.same}@uni-koeln.de, laurent.romary@inria.fr

In this poster, we present an XML-TEI conformant stand-off representation of reference in discourse, building on the seminal work carried out in the MATE project (Poesio, Bruneseaux & Romary 1999) and the earlier proposal on a reference annotation framework in Salmon-Alt & Romary (2005).

We make a three-way distinction between markables (the referring expressions), discourse entities (referents in the textual or extra-textual world), and links (relations that hold between referents, e.g., part-whole). Our approach differs from previous suggestions in that (i) inherent properties of the referent itself (e.g., animacy) are disentangled from the expressions used to refer to that referent, (ii) existing annotations from other layers such as morphosyntax are cleanly separated from the annotation of reference, but can be combined in queries and (iii) our proposal is integrated into the larger structure of existing TEI-ISO standards, thereby allowing for compatibility with existing TEI-encoded corpora and data sustainability.

The workflow of adding reference annotations to an existing corpus will be demonstrated with concrete examples from ongoing work in the SFB 1252 (subprojects C01 and INF), where this representation of reference is the backbone for the annotation of (sentence) topic chains in dialogue data and for queries of topics in various grammatical constructions.

References: • Poesio, M., F. Bruneseaux & L. Romary. 1999. The MATE meta-scheme for coreference in dialogues in multiple languages. In *ACL'99 Workshop Towards Standards and Tools for Discourse Tagging*, 65–74. College Parc, US. • Salmon-Alt, S. & L. Romary. 2005. The Reference Annotation Framework: A case for semantic content representation. In H. Bunt (ed.), *IWCS-6, Tilburg, Netherlands: ACL SIGSEM*.