Discovery of Structural Similarities across Narrative Texts

Nils Reiter and Anette Frank
Department of Computational Linguistics, Heidelberg University

Narratives & Structural Similarities
- "Account of connected events, a story" (OED)
  - Narratives consist of events
  - The events are connected
  - Event-level similarities in context
  - (Some) individual events are similar
  - Connections between similar events are similar

Alignment Algorithms
- Sequence Alignment (Needleman & Wunsch, 1970)
- Graph-Based Predicate Alignment (Roth, 2014; Roth & Frank, 2012)
- Bayesian Model Merging (Finlayson, 2012; Stolcke & Omohundro, 2012)

Detection & Inspection of Similarities
- Random-Walk algorithm detects strongly connected events
  - Counts the (average) number of crossings to another sequence
  - Assigns a connectivity score to each event
  - Events can be ranked according to the score

Application Scenarios

Folktales
- Similar story elements appear in different tales
  - Characters (the evil witch)
  - Event subsequences (hero is being tested before he receives a weapon)
- Corpus collection
  - 37 variations of 7 different tales (Uther, 2004)

Rituals
- "Ritual grammar"
  - Rituals consist of reappearing "building blocks"
  - Composition of blocks to rituals follows rules
  - Corpus collection
  - 11 descriptions of 5 ritual types (Gutschow & Michaels, 2005)

Linguistic Processing
- Based on UIMA
- Annotations on many linguistic levels
  - Tokenization, sentence splitting, part of speech tagging, lemmatization, dependency parsing, word sense disambiguation, semantic role labeling, coreference resolution
  - Creates an integrated discourse representation for each narrative
  - Domain adaptation techniques to cope with text characteristics (Frank et al., 2012; Reiter, 2014; Reiter, Hellwig, et al., 2011)

Summary
- Deep linguistic analysis with domain adaptation
- Unsupervised event alignment across narratives
- Targeted manual inspection with visualization

References


