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

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
- \rightarrow 13 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, chunking, dependency parsing, word sense disambiguation, semantic role labeling, coreference resolution
- Creates an integrated discourse representation for each nar-
- Domain adaptation techniques to cope with text characteristics (Frank et al., 2012; Reiter, 2014; Reiter, Hellwig, et

Alignment Algorithms

- Sequence Alignment (Needleman & Wunsch, 1970)
- Graph-Based Predicate Alignment (Roth, 2014; Roth & Frank, 2012)

Unsupervised

Event Alignment

The thakāli should bring the boy holding his hand and make him sit on a svastika.

TEXT_CREATION

CAUSE_MOTION

Cause_fluidic_motionause_fluidic_motion

Cause_fluidic_motion

Cause_fluidict_metion

Unconnected entites:

Sprinkle water from the arghyapātra reciting the devasya tvā.

The boy should worship the kalasa by throwing rice.

hand#n#1

Events

Participants

Coreferences

Word senses

reciting the agnir murdha divah and the trataram indram.

Offering of oil reciting the kāṇḍātkāṇḍāt or the dīrghāyutvāya.

Put grains of unhusked rice on the head of the boy reciting the

Scatter the material from the siphã on the head of the boy reciting the

Wave the thāybhū on which a candramaṇḍala is drawn reciting the

Worship the hands of the maternal uncle, sister of the father and

Hand over the plate with the candramandala together with a small

The preceeding mantra is for pouring warm water by the father on

Divide the hair with a porcupine tail reciting again the dirghayutvaya.

Evidence: "In the east above the forehead, a piece of the Banyan

tree, and in the south above the right ear, a piece of the country

udumbara, in the south above the right ear, a piece of the asvattha.

and in the west also a piece of the fig tree. In the middle a golden coin together with darbha grass at the back; in the middle, on the place of

The thakāli should pour cold water reciting the svasti na indro.

Put the copper pot together with **curd** and grinded grey **sesame** paste

The razor should be handed over to the maternal uncle.

The dīrghāyutvāya and the savitrā prasūta daivyā...

the head of the boy before shaving the hair.

measuring vessel and keys reciting the ausraghnam.

Perform ārati with waving a lamp on the siphā.

Frame element ID Token II

Do the same below i.e. on the feet.

Incense etc., svagã, blessing.

Again ārati reciting the tejo 'si

on the fire.

pratistha reciting the mano jutir.

oLemma ○Frame name ○Frame ID

Bringing

CAUSE_FLUIDIC_MOTION

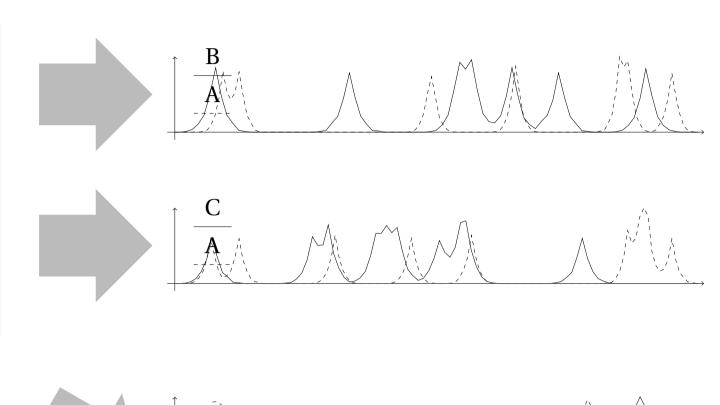
water#n#1

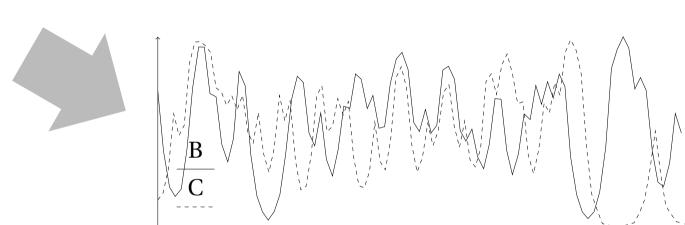
• Bayesian Model Merging (Finlayson, 2012; Stolcke & Omohundro, 1993)

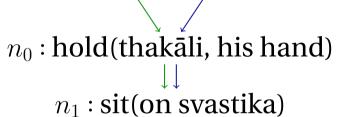
Detection & Inspection of Similarities

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







 n_2 : recite(raksohanam)

 n_5 : recite(adhy avoca | devasya tvā)

 n_4 : sprinkle(water)

 n_6 : salutation n_7 : offer(lamp)

 n_3 : wash(body)

 n_8 : burn(wick) -

sit(on svastika) recite(raksohanam) recite(adhy avoca) burn(wick) sprinkle(water) recite(devasya tvā)

> Description C: hold(thakāli, his hand) recite(raksohanam) orinkle(water) recite(devasya tvā)

Summary

(Reiter, Frank, et al., 2014)

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

- rative
- al., 2011)

References

BCANIMDGEHFLK

Finlayson, Mark Alan. Learning Narrative Structure from Annotated Folktales. PhD thesis. Massachusetts Institute of Technology, Feb. 2012.

Frank, Anette, Thomas Bögel, Oliver Hellwig & Nils Reiter. Semantic Annotation for the Digital Humanities. In: Linguistic Issues in Language Technology 7 (1), Jan. 2012. Gutschow, Niels & Axel Michaels. Handling Death. The Dynamics of Death and Ancestor Rituals Among the Newars of Bhaktapur. Vol. 3. Ethno-Indology. Heidelberg Studies in South Asian Rituals. Wiesbaden: Harrassowitz Verlag, 2005.

Needleman, Saul B. & Christian D. Wunsch. A general method applicable to the search for similarities in the amino acid sequence of two proteins. In: Journal of Molecular

Reiter, Nils. Discovering Structural Similarities in Narrative Texts using Event Alignment Algorithms. Submitted on Aug 13, defended on Nov 27. PhD thesis. Heidelberg

Reiter, Nils, Anette Frank & Oliver Hellwig. An NLP-based cross-document approach to narrative structure discovery. In: Literary and Linguistic Computing 29 (4), 2014, pp. 583-605. eprint: http://llc.oxfordjournals.org/content/29/4/583.full.pdf+html.

Reiter, Nils, Oliver Hellwig, Anette Frank, Irina Gossmann, Borayin Maitreya Larios, Julio Rodrigues & Britta Zeller. Adapting NLP Tools and Frame-Semantic Resources for the Semantic Analysis of Ritual Descriptions. In: Language Technology for Cultural Heritage: Selected Papers from the LaTeCH Workshop Series. Ed. by Caroline Sporleder, Antal van den Bosch & Kalliopi A. Zervanou. Theory and Applications of Natural Language Processing. Berlin, Heidelberg: Springer, 2011, pp. 171–193.

Roth, Michael. Inducing Implicit Arguments via Cross-document Alignment: A Framework and its Applications. PhD thesis. Heidelberg University, 2014.

Roth, Michael & Anette Frank. Aligning Predicates across Monolingual Comparable Texts using Graph-based Clustering. In: 2012, pp. 171–182. Stolcke, Andreas & Stephen Omohundro. Hidden Markov Model Induction by Bayesian Model Merging. In: Advances in Neural Information Processing Systems. Ed. by Steve J.

Hanson, J. D. Jack D. Cowan & C. Lee Giles. Vol. 5. San Mateo, California: Morgan Kaufmann, 1993, pp. 11–18.

Uther, Hans-Jörg. The Types of International Folktales: A Classification and Bibliography. Based on the system of Antti Aarne and Stith Thompson. FF Communications 284–286. Helsinki: Suomalainen Tiedeakatemia, 2004.