

Using NLP Methods for the Analysis of Rituals

Nils Reiter[♠], Oliver Hellwig[♦], Anand Mishra[♦], Anette Frank[♠], Jens Burkhardt[♠]

♠ Dept. of Computational Linguistics and ♦ South Asia Institute, University of Heidelberg

Ritual structure research

- Rituals exist in all human cultures across all times.
- Similarities between rituals from very distant cultures can be observed.
- Is there an underlying universal structure of rituals? A ritual grammar?

Project research topics

- A formal-linguistic, data-driven and quantitative approach for detecting ritual structures
- Frame semantics for the formalization of events and semantic roles
- Frame ontology allows for abstraction of events, roles and participants
- Ritual-specific knowledge expressed in ritual descriptions is linked to a ritual ontology
- Search functionalities for ritual scientists to verify hypotheses against semantically annotated data

Heterogeneous domain

- Terminology from different fields (history, theology, ...) is partially controversial
- Systematic exploration of the domain by extracting terms from scientific literature *about* rituals
- Data-driven approach for acquiring domain vocabulary

Alternative approaches for domain term extraction

► tf * idf

- Comparison corpora BAWE and BNC
- $freq_{t,d}$: frequency of term t in corpus d
- df_t : number of corpora in which t appears
- D : # of corpora (= 3)

$$tf_{t,d} = \frac{freq_{t,d}}{\max_{t'} freq_{t',d}}$$

$$tfidf_{t,d} = tf_{t,d} * \log\left(\frac{D}{df_t}\right)$$

Buitelaar & Sacaleanu (2001)

► χ^2

- Calculation of raw χ^2 values based on lemma and pos
- Sum over both non-domain corpora
- Normalized and scaled to [0,1]

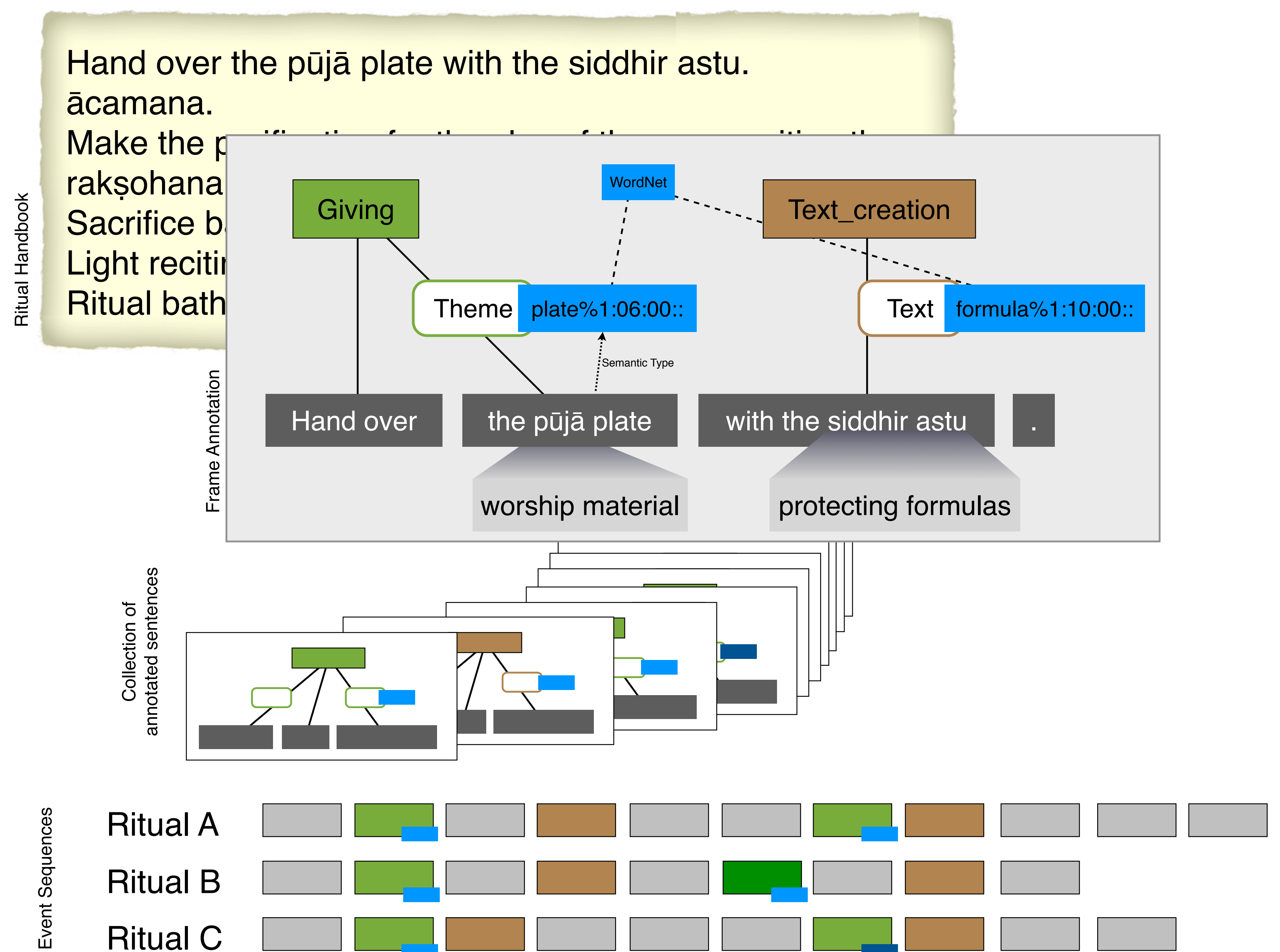
Agirre et al. (2001)

► pagerank

- One node for each candidate term
- Edges between two co-occurring cand. terms
- Calculate PageRank (Brin & Page, 1998)
- Highly connected nodes get higher weight

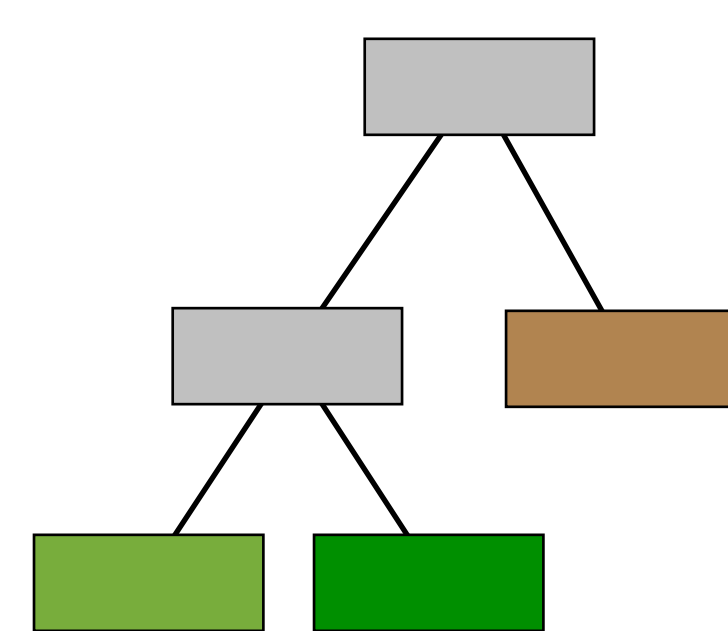
Yang et al. (2009)

This research is funded by the German Research Foundation (DFG) and part of the collaborative research center **SFB 619: Ritual Dynamics**.



► Frames & Event Sequences

- Ritual Element ?
- Are and related?



► Roles & Semantic Types

- How similar are and ?
- Can be substituted by ?

Evaluation

- 2 ritual experts evaluated 20 extracted terms as *yes*, *no* or *maybe*

pos	All	N	V	A
κ	0.35	0.22	0.35	0.49

- Approaches score differently for different parts of speech
- Percentage of partially correct extracted terms

Approach	A1=yes, A2=maybe	A1=no, A2=maybe
	Overall N/V/A	Overall N/V/A
tf * idf	23.33% 20/30/20	56.67% 70/40/60
χ^2	56.67% 30/60/80	30% 60/30/0
PageRank	20% 30/0/30	60% 50/60/70

References

Eneko Agirre, Olatz Ansa, David Martínez, and Eduard Hovy. Enriching WordNet Concepts with Topic Signatures. In *Proceedings of the Workshop on WordNet and Other Lexical Resources at NAACL*. 2001.
Sergey Brin and Lawrence Page. The anatomy of a large-scale hypertextual Web search engine. *Computer networks and ISDN systems*, 30(1-7):107–117. 1998.
Paul Buitelaar and Bogdan Sacaleanu: Ranking and Selecting Synsets by Domain Relevance. In *Proceedings of the Workshop on WordNet and Other Lexical Resources at NAACL*. 2001.
Charles J. Fillmore, Christopher R. Johnson, and Miriam R.L. Petruck. Background to FrameNet. *International Journal of Lexicography*, 16(3):235–250. 2003.
Yuhang Yang, Tiejun Zhao, Qin Lu, Dequan Zheng, and Hao Yu. Chinese term extraction using different types of relevance. In *Proceedings of ACL and AFNLP*. 2009.