Adapting NLP Tools and Frame-Semantic Resources for the Semantic Analysis of Ritual Descriptions

Nikolina Koleva

UdS

22. February 2012

Outline

- interdisciplinary subject, a bunch of challenges

- interdisciplinary subject, a bunch of challenges
- Is there underlying structure of rituals?

- interdisciplinary subject, a bunch of challenges
- Is there underlying structure of rituals?
- Does "ritual grammar" exist?

Motivation

- interdisciplinary subject, a bunch of challenges
- Is there underlying structure of rituals?
- Does "ritual grammar" exist?
- Is it universal?

Motivation

- interdisciplinary subject, a bunch of challenges
- Is there underlying structure of rituals?
- Does "ritual grammar" exist?
- Is it universal?
- There is evidence for variances and similarities in rituals wrt.

Motivation

- interdisciplinary subject, a bunch of challenges
- Is there underlying structure of rituals?
- Does "ritual grammar" exist?
- Is it universal?
- There is evidence for variances and similarities in rituals wrt.
  - culture

- interdisciplinary subject, a bunch of challenges
- Is there underlying structure of rituals?
- Does "ritual grammar" exist?
- Is it universal?
- There is evidence for variances and similarities in rituals wrt.
  - culture
  - time

- detection of complex event sequences

- detection of complex event sequences

- finding participants, objects, places and times involved in the events

- detection of complex event sequences

- finding participants, objects, places and times involved in the events
  How?

- detection of complex event sequences

- finding participants, objects, places and times involved in the events
  How? → using NLP tools

The Tasks

- detection of complex event sequences

- finding participants, objects, places and times involved in the events
  How? $\rightarrow$ using NLP tools

- recognition of variations and regularities of rituals

The Tasks

- detection of complex event sequences

- finding participants, objects, places and times involved in the events
  How? $\rightarrow$ using NLP tools

- recognition of variations and regularities of rituals

- focus on discourse semantic aspects due to complex event sequences

Difficulties

- no all-encompassing theoretical framework for ritual analysis

- thus recurrent structures in event sequences are unknown

- descriptions of rituals have different text features than texts used for the development of the used NLP tools
  $\rightarrow$ Need for adaptation

Resources and Tools for Analysis of Ritual Structure

- Frame Semantics
  - powerful framework: concept of scenario frames connected by frame relations and role inheritance

- Lexical Ontology, e.g. WordNet, for variation analysis in characteristics of events across rituals

- Semantically annotated corpora and reference ontology enables reasoning with external knowledge resources

Main steps

1. corpus creation and annotation
   - contains descriptions of different cultures
   - annotated with linguistic and ritual-specific tags

2. analysis of the ritual structure
   - deployment of logical and statistical methods for the detection of recurring structures and systematic variances in ritual descriptions based on semantic annotation

Ritual Descriptions

- collected from different sources

- collected from different sources
  - Hindu rituals from Nepal

Ritual Descriptions

- collected from different sources
  - Hindu rituals from Nepal
  - Middle East

Ritual Descriptions

- collected from different sources
  - ► Hindu rituals from Nepal
  - ► Middle East

- textual sources

Ritual Descriptions

- collected from different sources
  - ▶ Hindu rituals from Nepal
  - ▶ Middle East

- textual sources
  - ▶ theory-oriented studies by ritual researchers that deal with religious, ethnologic and social rituals (used to build ritual specific ontology)

Ritual Descriptions

- collected from different sources
  - Hindu rituals from Nepal
  - Middle East

- textual sources
  - theory-oriented studies by ritual researchers that deal with religious, ethnologic and social rituals (used to build ritual specific ontology)
  - descriptions of rituals

Ritual Descriptions

- collected from different sources
  - Hindu rituals from Nepal
  - Middle East

- textual sources
  - theory-oriented studies by ritual researchers that deal with religious, ethnologic and social rituals (used to build ritual specific ontology)
  - descriptions of rituals
    1. ethnographic observations (How rituals are performed in modern times?)

Ritual Descriptions

- collected from different sources
  - Hindu rituals from Nepal
  - Middle East

- textual sources
  - theory-oriented studies by ritual researchers that deal with religious, ethnologic and social rituals (used to build ritual specific ontology)
  - descriptions of rituals
    1. ethnographic observations (How rituals are performed in modern times?)
    2. ritual manuals (translations of original manuals that prescribe a ritual)

Ritual Descriptions

- collected from different sources
  - Hindu rituals from Nepal
  - Middle East

- textual sources
  - theory-oriented studies by ritual researchers that deal with religious, ethnologic and social rituals (used to build ritual specific ontology)
  - descriptions of rituals
    1. ethnographic observations (How rituals are performed in modern times?)
    2. ritual manuals (translations of original manuals that prescribe a ritual)
  - not trivial alignment of manuals that mention only relevant part of the events to an exhaustive (possibly subjective description)

Text Characteristics

- foreign terms

Text Characteristics

- foreign terms
  - He sweeps the place for the sacrificial fire with **kuśa**.

- foreign terms
  - ▶ He sweeps the place for the sacrificial fire with **kuśa**.
    → He sweeps the place for the sacrificial fire with <**grass *kuśa**>.

Text Characteristics

- foreign terms
    - He sweeps the place for the sacrificial fire with **kuśa**.
      $\rightarrow$ He sweeps the place for the sacrificial fire with **<grass *kuśa>**.
- fixed expressions that often have to be spoken

Text Characteristics

- foreign terms
  - He sweeps the place for the sacrificial fire with **kuśa**.
    → He sweeps the place for the sacrificial fire with <**grass *kuśa**>.
- fixed expressions that often have to be spoken
  - Salutation to Kubera reciting the mantra **arddha-mãsãh**.

Text Characteristics

- foreign terms
  - He sweeps the place for the sacrificial fire with **kuśa**.
    → He sweeps the place for the sacrificial fire with **<grass *kuśa>**.
- fixed expressions that often have to be spoken
  - Salutation to Kubera reciting the mantra **arddha-mãsãh**.
    →replace with placeholders during processing and reinsert them afterwards

Text Characteristics

- foreign terms
  - ▶ He sweeps the place for the sacrificial fire with **kuśa**.
    → He sweeps the place for the sacrificial fire with <**grass *kuśa**>.
- fixed expressions that often have to be spoken
  - ▶ Salutation to Kubera reciting the mantra **arddha-mãsãh**.
    →replace with placeholders during processing and reinsert them afterwards
- Imperatives, PPs and nested sentences

Text Characteristics

- foreign terms
    - He sweeps the place for the sacrificial fire with **kuśa**.
      $\rightarrow$ He sweeps the place for the sacrificial fire with **<grass *kuśa>**.
- fixed expressions that often have to be spoken
    - Salutation to Kubera reciting the mantra **arddha-mãsãh**.
      $\rightarrow$replace with placeholders during processing and reinsert them afterwards
- Imperatives, PPs and nested sentences
    - very common and difficult to process

Text Characteristics

- foreign terms
  - He sweeps the place for the sacrificial fire with **kuśa**.
    $\rightarrow$ He sweeps the place for the sacrificial fire with **<grass *kuśa>**.
- fixed expressions that often have to be spoken
  - Salutation to Kubera reciting the mantra **arddha-mãsãh**.
    $\rightarrow$replace with placeholders during processing and reinsert them afterwards
- Imperatives, PPs and nested sentences
  - very common and difficult to process
    $\rightarrow$ chunks are sufficient for semantic role labeling

Text Characteristics

- foreign terms
  - He sweeps the place for the sacrificial fire with **kuśa**.
    $\rightarrow$ He sweeps the place for the sacrificial fire with **<grass *kuśa>**.
- fixed expressions that often have to be spoken
  - Salutation to Kubera reciting the mantra **arddha-mãsãh**.
    $\rightarrow$replace with placeholders during processing and reinsert them afterwards
- Imperatives, PPs and nested sentences
  - very common and difficult to process
    $\rightarrow$ chunks are sufficient for semantic role labeling
- comments occurence: not each sentence describes the ritual

- *Salutation to Kubera reciting the mantra **arddha-mãsãh**.*

- *Salutation to Kubera reciting the mantra **arddha-mãsãh**.*
  - indicates event happening during the ritual performance: factual

Sentence Classification

- *Salutation to Kubera reciting the mantra **arddha-mãsãh**.*
  - ▶ indicates event happening during the ritual performance: factual

- *The involvement of the nephews can be understood as a symbolic action to address those of the following generation who do not belong to lineage of the deceased.*

- *Salutation to Kubera reciting the mantra **arddha-mãsãh**.*
  - ▶ indicates event happening during the ritual performance: factual

- *The involvement of the nephews can be understood as a symbolic action to address those of the following generation who do not belong to lineage of the deceased.*
  - ▶ another level of information, clear interpretation or comment: interpretative /ignored for the frame annotation/

Sentence Classification

- *Salutation to Kubera reciting the mantra **arddha-mãsãh**.*
  - ▶ indicates event happening during the ritual performance: factual

- *The involvement of the nephews can be understood as a symbolic action to address those of the following generation who do not belong to lineage of the deceased.*
  - ▶ another level of information, clear interpretation or comment: interpretative /ignored for the frame annotation/

- *The wife of the chief mourner [...] will carry a symbolic mat that represents the bed of the deceased[...].*

Sentence Classification

- *Salutation to Kubera reciting the mantra **arddha-mãsãh**.*
  - ▸ indicates event happening during the ritual performance: factual

- *The involvement of the nephews can be understood as a symbolic action to address those of the following generation who do not belong to lineage of the deceased.*
  - ▸ another level of information, clear interpretation or comment: interpretative /ignored for the frame annotation/

- *The wife of the chief mourner [...] will carry a symbolic mat that represents the bed of the deceased[...].*
  - ▸ ambiguous wrt. these classes, or contain both

- many special non-English characters: ś, or even tokens: **Geṇeśa**

- many special non-English characters: ś, or even tokens: **Geṇeśa**
    - employ rule-based tokenizer that uses Unicode character ranges

Linguistic Processing

- many special non-English characters: ś, or even tokens: **Geṇeśa**
  - employ rule-based tokenizer that uses Unicode character ranges
- poor results for PoS-Tagging and Chunking

Linguistic Processing

- many special non-English characters: ś, or even tokens: **Geṇeśa**
  - ▶ employ rule-based tokenizer that uses Unicode character ranges
- poor results for PoS-Tagging and Chunking
  - ▶ a lot of unseen tokens

Linguistic Processing

- many special non-English characters: ś, or even tokens: **Geṇeśa**
  - employ rule-based tokenizer that uses Unicode character ranges
- poor results for PoS-Tagging and Chunking
  - a lot of unseen tokens
  - many rare uncommon constructions

- many special non-English characters: ś, or even tokens: **Geṇeśa**
  - employ rule-based tokenizer that uses Unicode character ranges
- poor results for PoS-Tagging and Chunking
  - a lot of unseen tokens
  - many rare uncommon constructions
  - $\rightarrow$ experiment with different scenarios for the domain adaptation

- two sources

- two sources
  - 532 sentences of ritual descriptions, manually annotated with PoS and Chunks

Experimental Design for Adaptation

- two sources
  - 532 sentences of ritual descriptions, manually annotated with PoS and Chunks
  - Wall Street Journal corpus

Experimental Design for Adaptation

- two sources
  - 532 sentences of ritual descriptions, manually annotated with PoS and Chunks
  - Wall Street Journal corpus
- 10-fold cross-validation for evaluation

Evaluation of models for PoS-Tagging

Training data	Accuracy (%)
WSJ	90.90
RIT	94.82
WSJ + RIT	95.72
WSJ + RIT↑	**96.23**
WSJ↓ + RIT	95.25
WSJ × RIT	**96.86**
WSJ × RIT ↑	**96.85**
WSJ ↓ × RIT	95.92

Evaluation of models for Chunking

Training data	F-measure(%)
WSJ	86.6
RIT	85.7
WSJ + RIT	86.6
WSJ + RIT↑	**88.1**
WSJ↓ + RIT	83.1
WSJ × RIT	74.4
WSJ × RIT ↑	**81.3**
WSJ ↓ × RIT	73.3

Anaphora and Coreference Resolution

1. **The father** should touch the girl [...]. Let **him** give a golden coin as ritual fee[...].
2. Let **the girl** sit on the seat [...]. Let **the girl** wash her face [...].

*1. **The father** should touch the girl [...]. Let **him** give a golden coin as ritual fee[...].*
*2. Let **the girl** sit on the seat [...]. Let **the girl** wash her face [...].*

- BART /machine learning toolkit/; output entire coreference chain
- JavaRAP /rule-based/; generate only anaphor-antecedent pairs

Evaluation of Anaphora and Coreference Resolution

- 26 anaphor-antecedent pairs that correspond to 10 coreference chains

Evaluation of Anaphora and Coreference Resolution

- 26 anaphor-antecedent pairs that correspond to 10 coreference chains

- JavaRAP: 61.5% correct pairs found; evaluation only for anaphora resolution

Evaluation of Anaphora and Coreference Resolution

- 26 anaphor-antecedent pairs that correspond to 10 coreference chains

- JavaRAP: 61.5% correct pairs found; evaluation only for anaphora resolution

Measure		Anaphora & Coreference		Coreference only	
		Baseline	BART	Baseline	BART
MUC	Precision	12.1	39.72	6.2	23.5
	Recall	12.1	70.73	6.6	66.6
	F_1-measure	12.1	50.87	6.4	34.4
B-CUBED	Precision	30.4	11.38	29.6	7.2
	Recall	19.3	65.95	25.9	64.9
	F_1-measure	23.6	19.41	27.7	13.0

- automatic extraction of verbs

Semantic Annotation of Ritual Descriptions

- automatic extraction of verbs
- map the ritual actions to frames in FrameNet

Semantic Annotation of Ritual Descriptions

- automatic extraction of verbs
- map the ritual actions to frames in FrameNet
- 80% of the verbs from the ritual corpus are contained as lexical units

Semantic Annotation of Ritual Descriptions

- automatic extraction of verbs
- map the ritual actions to frames in FrameNet
- 80% of the verbs from the ritual corpus are contained as lexical units
  - some of them are equivalent only at a lexical level, occur in different senses

Semantic Annotation of Ritual Descriptions

- automatic extraction of verbs
- map the ritual actions to frames in FrameNet
- 80% of the verbs from the ritual corpus are contained as lexical units
  - some of them are equivalent only at a lexical level, occur in different senses
- frames of FrameNet are too abstract for annotation of ritual descriptions

Semantic Annotation of Ritual Descriptions

- automatic extraction of verbs
- map the ritual actions to frames in FrameNet
- 80% of the verbs from the ritual corpus are contained as lexical units
  - some of them are equivalent only at a lexical level, occur in different senses
- frames of FrameNet are too abstract for annotation of ritual descriptions $\rightarrow$ new frames have to be designed
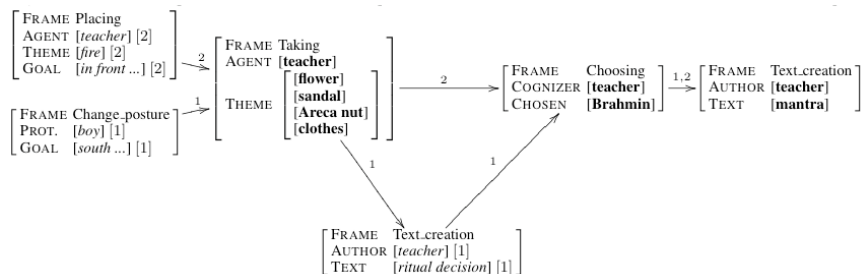
Semantic Annotation of Ritual Descriptions

- automatic extraction of verbs
- map the ritual actions to frames in FrameNet
- 80% of the verbs from the ritual corpus are contained as lexical units
  - some of them are equivalent only at a lexical level, occur in different senses
- frames of FrameNet are too abstract for annotation of ritual descriptions $\rightarrow$ new frames have to be designed
- train a frame-semantic labeler on initial frame inventory and annotated corpus (manually created)

Semantic Annotation of Ritual Descriptions

- automatic extraction of verbs
- map the ritual actions to frames in FrameNet
- 80% of the verbs from the ritual corpus are contained as lexical units
  - some of them are equivalent only at a lexical level, occur in different senses
- frames of FrameNet are too abstract for annotation of ritual descriptions $\rightarrow$ new frames have to be designed
- train a frame-semantic labeler on initial frame inventory and annotated corpus (manually created)
- manual correction after the automatic assignment of semantic roles

Detecting Ritual Structure |

1. "... the boy sits down$_{Change\_posture}$ south of the teacher. (The teacher) takes$_{Taking}$ flowers, sandal, Areca nut and clothes and declares$_{Text\_creation}$ the ritual decision to select$_{Choosing}$ the Brahmin by saying$_{Text\_creation}$ the mantra ..."

2. "... (the teacher) places$_{Placing}$ (fire in a vessel of bell metal) in front of himself. Having taken$_{Taking}$ flowers, sandal, Areca nut, clothing etc. he should select$_{Choosing}$ a Brahmin. The Brahmin is selected with$_{Text\_creation}$" the mantra ..."

$$
\begin{bmatrix}
\text{FRAME} & \text{Placing} \\
\text{AGENT} & [\textit{teacher}]\,[2] \\
\text{THEME} & [\textit{fire}]\,[2] \\
\text{GOAL} & [\textit{in front} ...]\,[2]
\end{bmatrix}
\begin{bmatrix}
\text{FRAME} & \text{Taking} \\
\text{AGENT} & [\textbf{teacher}] \\
 & [\textbf{flower}] \\
\text{THEME} & [\textbf{sandal}] \\
 & [\textbf{Areca nut}] \\
 & [\textbf{clothes}]
\end{bmatrix}
\begin{bmatrix}
\text{FRAME} & \text{Choosing} \\
\text{COGNIZER} & [\textbf{teacher}] \\
\text{CHOSEN} & [\textbf{Brahmin}]
\end{bmatrix}
\begin{bmatrix}
\text{FRAME} & \text{Text\_creation} \\
\text{AUTHOR} & [\textbf{teacher}] \\
\text{TEXT} & [\textbf{mantra}]
\end{bmatrix}
$$

$$
\begin{bmatrix}
\text{FRAME} & \text{Change\_posture} \\
\text{PROT.} & [\textit{boy}]\,[1] \\
\text{GOAL} & [\textit{south} ...]\,[1]
\end{bmatrix}
$$

$$
\begin{bmatrix}
\text{FRAME} & \text{Text\_creation} \\
\text{AUTHOR} & [\textit{teacher}]\,[1] \\
\text{TEXT} & [\textit{ritual decision}]\,[1]
\end{bmatrix}
$$

- interpret each frame and its roles as atomic symbol
- find global alignment between *n* sequences describing the same ritual

- collecting texts: manuals and descriptions
- analysis of the domain characteristics
- adaptation of preprocessing tools and semantic resources
- detection of ritual structure and possible subsequences

Literature

[1] Nils Reiter, Oliver Hellwig, Anette Frank, Irina Gossmann, Borayin Maitreya Larios, Julio Rodrigues, and Britta Zeller. "Adapting NLP Tools and Frame-Semantic Resources for the Semantic Analysis of Ritual Descriptions". In: C. Sporleder, A. van den Bosch, and K. Zervanou (Eds.), *Language Technology for Cultural Heritage*. Selected Papers from the LaTeCH Workshop Series, Series: Theory and Applications of Natural Language Processing. Heidelberg: Springer, 2011.

[2] Nils Reiter, Oliver Hellwig, Anand Mishra, Anette Frank, Irina Gossmann, Borayin Maitreya Larios, Julio Rodrigues, and Britta Zeller. "Adapting Standard NLP Tools and Resources to the Processing of Ritual Descriptions". *Proceedings of ECAI 2010 workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities.*

Thank you for your attention!