Harvesting Event Chains in Ritual Descriptions using Frame Semantics: Chances and Challenges

Anette Frank  Nils Reiter

Department of Computational Linguistics
Heidelberg University, Germany

1st AMICUS Workshop, Vienna, September 30, 2010
Overview

Motivation
  - Studying Rituals
  - Detecting Structure

Processing Ritual Descriptions
  - Data Collection and Annotation
  - Challenges: Domain Adaptation
  - Chances: Aggregation and Exploitation

Conclusions
Rituals

Offer a lamp and speak the mantra tejo ’si.

Shower pieces of fruits from the measuring vessel on the head of the boy with yāḥ phalini.
Investigating the Structure of Rituals

- Many rituals contain re-occurring building blocks
- Ritual elements are used to compose rituals
- The composition follows rules
- Is there a “Ritual Grammar”?  
  - Can we identify the “ritual elements”?  
  - Can we identify rules of composition?  
  - Are there culture-independent rules or elements?

Project Aims

- Investigating the event structure of rituals
- Detecting regularities, similarities and differences
- Using empirical, computational linguistics methods
What does CL have to offer?

Event Annotation using Frame Semantics

- Frames: Prototypical events (e.g., TAKING)
- Frame Elements: Frame-specific semantic roles (e.g., AGENT, THEME, SOURCE)
- Frame Relations: Form a network of frames through inheritance, precedence, etc.
- Scenario Frames: Sequences of events forming a “scenario”

Sense Annotation of Frame Elements

- Offer a lamp.
- OIL LAMP > LAMP > DEVICE .. > OBJECT
- Computing selectional preferences and divergences
Harvesting Event Chains from Ritual Descriptions

*Frame semantics as an abstraction layer for event chains*

Burchardt et al. (2005)

Detecting structural elements

- patterns of (partial) event sequences
- similarities and differences across event role fillers

Hellwig (2009)

Using sequence alignment and semantic distance measures
Overview

Motivation

Processing Ritual Descriptions

Data Collection and Annotation
- Ritual Descriptions: Handbooks vs. Observations
- Annotation: From pre-processing to semantic annotation

Challenges: Domain Adaptation
- Linguistic characteristics of ritual descriptions
- Adaptation techniques: Tagging, Chunking
- Domain Adaptation: Frame Annotation

Chances: Aggregation and Exploitation

Conclusions
Data Sources

<table>
<thead>
<tr>
<th></th>
<th>Handbook (prescriptive)</th>
<th>Observation (descriptive)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(practitioners)</td>
<td>(researchers)</td>
</tr>
<tr>
<td>Initiation</td>
<td>12</td>
<td>8</td>
</tr>
<tr>
<td>Death and Ancestor</td>
<td>1</td>
<td>19</td>
</tr>
</tbody>
</table>

**Table:** Collected ritual descriptions

Michaels and Gutschow (2005; 2008)
Annotation Pipeline

- Frame Semantics
  - Concepts
  - Anaphora
  - Parsing
  - Chunking
  - POS-tagging
  - Tokenization
Annotation Pipeline

Frame Semantics
  Concepts
  Anaphora
  Parsing
  Chunking
  POS-tagging
  Tokenization

Heuristic
Annotation Pipeline

Frame Semantics
  Concepts
  Anaphora
  Parsing
  Chunking
  POS-tagging
  Tokenization

OpenNLP
Morton et al. (2005)
Annotation Pipeline

- Frame Semantics
- Concepts
- Anaphora
- Parsing
- Chunking
- POS-tagging
- Tokenization

Manual Annotation
Overview

Motivation

Processing Ritual Descriptions
  Data Collection and Annotation
    Ritual Descriptions: Handbooks vs. Observations
    Annotation: From pre-processing to semantic annotation
  Challenges: Domain Adaptation
    Linguistic characteristics of ritual descriptions
    Adaptation techniques: Tagging, Chunking
    Domain Adaptation: Frame Annotation
  Chances: Aggregation and Exploitation

Conclusions
Text Characteristics

Ritual Descriptions have very special characteristics

- Foreign terms
- Fixed expressions
- Imperatives
- Complex sentence structure
- Interpretations

Reiter et al. (2010)
Text Characteristics

Ritual Descriptions have very special characteristics

- Foreign terms
- Fixed expressions
- Imperatives
- Complex sentence structure
- Interpretations

Example

*Hand over the worship materials.*

Reiter et al. (2010)
Text Characteristics

Ritual Descriptions have very special characteristics

- Foreign terms
- Fixed expressions
- Imperatives
- Complex sentence structure
- Interpretations

Example

*Hand over the worship materials.*

Frequency of imperative verbs

(BNC) 2% ↔ 70% (Ritual texts)

Reiter et al. (2010)
Domain Adaptation

In general: Transfer models trained on one domain onto another

Standard Case

Source Domain Standard domain like newspaper texts;
Lots of training data available

Target Domain New domain like ritual descriptions;
No training data available

Annotating a small amount of target domain data is often feasible

Daumé III et al. (2010), Daumé III and Marcu (2006)
## Domain Adaptation: Combining Data Sets

- Retraining of model(s) on various data sets

<table>
<thead>
<tr>
<th>Data</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>baseline</td>
<td></td>
</tr>
<tr>
<td>WSJ</td>
<td>The Wall Street Journal</td>
</tr>
<tr>
<td>Rit</td>
<td>Ritual descriptions</td>
</tr>
<tr>
<td>data set</td>
<td></td>
</tr>
<tr>
<td>WSJ + Rit</td>
<td>Union of WSJ and Rit</td>
</tr>
<tr>
<td>WSJ + Rit↑</td>
<td>Oversampling Rit</td>
</tr>
<tr>
<td>WSJ ↓ + Rit</td>
<td>Undersampling WSJ</td>
</tr>
<tr>
<td>features</td>
<td></td>
</tr>
<tr>
<td>WSJ × Rit</td>
<td>Augmented features (Daumé III 2007)</td>
</tr>
<tr>
<td>WSJ × Rit↑</td>
<td>Oversampling Rit</td>
</tr>
<tr>
<td>WSJ ↓ × Rit</td>
<td>Undersampling WSJ</td>
</tr>
</tbody>
</table>

**Table:** Overview of domain adaptation approaches
Augmenting the Feature Space

- One feature space for each domain
- A combined feature space for the “general domain”

Daumé III (2007)
Augmenting the Feature Space

- One feature space for each domain
- A combined feature space for the “general domain”

Daumé III (2007)

Example

- The monitor is flickering
- We monitor daily news
Augmenting the Feature Space

- One feature space for each domain
- A combined feature space for the “general domain”

Daumé III (2007)

Example

| DT | The          |
| DNS | monitor      |
| VB  | is           |
| VBG | flickering   |

| DT | The          |
| DNS | monitor      |
| VB  | is           |
| VBG | flickering   |

| PRP | We           |
| VB  | monitor      |
| JJ  | daily        |
| NNS | news         |

The
monitor
is
flickering

We
monitor
daily
news
Augmenting the Feature Space

- One feature space for each domain
- A combined feature space for the “general domain”

Example

The monitor is flickering.
We monitor daily news.

Daumé III (2007)
# Results for POS-Tagging

<table>
<thead>
<tr>
<th>Training data</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>WSJ</td>
<td>94.01 %</td>
</tr>
<tr>
<td>Rit</td>
<td>95.47 %</td>
</tr>
<tr>
<td>WSJ + Rit</td>
<td>97.32 %</td>
</tr>
<tr>
<td>WSJ + Rit ↑</td>
<td>97.59 %</td>
</tr>
<tr>
<td>WSJ ↓ + Rit</td>
<td>96.97 %</td>
</tr>
<tr>
<td>WSJ × Rit</td>
<td>97.19 %</td>
</tr>
<tr>
<td>WSJ × Rit ↑</td>
<td>97.40 %</td>
</tr>
</tbody>
</table>

**Table:** Part of speech tagging results with different models. (Tested on 408 sentences using 10-fold cross validation)
## Results for Chunking

<table>
<thead>
<tr>
<th>Training data</th>
<th>Precision</th>
<th>Recall</th>
<th>$F_{\beta=1}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>WSJ</td>
<td>87.72 %</td>
<td>87.23 %</td>
<td>87.47</td>
</tr>
<tr>
<td>Rit</td>
<td>91.09 %</td>
<td>89.85 %</td>
<td>90.47</td>
</tr>
<tr>
<td>WSJ + Rit</td>
<td>90.18 %</td>
<td>89.44 %</td>
<td>89.80</td>
</tr>
<tr>
<td>WSJ + Rit ↑</td>
<td>91.07 %</td>
<td>89.62 %</td>
<td>90.33</td>
</tr>
<tr>
<td>WSJ ↓ + Rit</td>
<td>91.46 %</td>
<td>90.34 %</td>
<td>90.89</td>
</tr>
<tr>
<td>WSJ × Rit</td>
<td>88.98 %</td>
<td>88.15 %</td>
<td>88.56</td>
</tr>
<tr>
<td>WSJ × Rit ↑</td>
<td>91.75 %</td>
<td>90.24 %</td>
<td>90.99</td>
</tr>
<tr>
<td>WSJ ↓ × Rit</td>
<td>91.49 %</td>
<td>90.44 %</td>
<td>90.96</td>
</tr>
</tbody>
</table>

**Table:** Chunking results with different models  
(Tested on 408 sentences using 10-fold cross validation)
Annotation Procedure

Stage 1

- Identify frequent verbs (predicates)
- Group them into semantic fields
  - cast, drag, draw, hang, lay, move, place, put, raise, remove, throw, ...
- Find appropriate frames
  - Placing and Moving
Annotation Procedure

Stage 1

- Identify frequent verbs (predicates)
- Group them into semantic fields
  - cast, drag, draw, hang, lay, move, place, put, raise, remove, throw, . . .
- Find appropriate frames
  - Placing and Moving

Stage 2

- Automatic pre-annotation of frames, based on established verb-frame lexicon
- Train automatic role labeler on manual role annotations
- Correct (unsure) role assignments manually
Annotation Process

Pre-annotation of frames

- Offer\textsubscript{Giving} a lamp and speak\textsubscript{Speaking} the mantra tejo ’si.
Annotation Process

Pre-annotation of frames

- Offer\textsuperscript{Giving} a lamp and speak\textsuperscript{Speaking} the mantra tejo 'si.

Manual validation of frames and role annotation

- Offer\textsuperscript{Giving} \[ a \text{ lamp} \]_{\text{Theme}} and speak\textsuperscript{Speaking} \[ the \text{ mantra tejo 'si} \]_{\text{Text}}.

- Frame assignment: Between 68.8\% and 91.4\% precision
Annotation Process

Pre-annotation of frames

- $\text{Offer}_{\text{Giving}}$ a lamp and $\text{speak}_{\text{Speaking}}$ the mantra tejo 'si.

Manual validation of frames and role annotation

- $\text{Offer}_{\text{Giving}}$ [a lamp]$_{\text{Theme}}$ and $\text{speak}_{\text{Speaking}}$ [the mantra tejo 'si]$_{\text{Text}}$.
- Frame assignment: Between 68.8% and 91.4% precision

Semi-automatic role (and co-reference) annotation

- Role assignment: Between 50% and 100% precision depending on frequency
Overview

Motivation

Processing Ritual Descriptions
- Data Collection and Annotation
  - Ritual Descriptions: Handbooks vs. Observations
  - Annotation: From pre-processing to semantic annotation
- Challenges: Domain Adaptation
  - Linguistic characteristics of ritual descriptions
  - Adaptation techniques: Tagging, Chunking
  - Domain Adaptation: Frame Annotation

Chances: Aggregation and Exploitation

Conclusions
Aggregation and Exploitation

Abstraction and normalization of event chains
- Based on annotated frames, roles and coreference chains

Capturing statistics about “ritual elements”
- Co-occurrence of ritual objects/participants with ritual actions
- Identify overlapping (partial) event chains across rituals

Interpretation is up to ritual experts
- Omissions may disclose implicit knowledge about ritual acts
- Enable search over annotated descriptions to validate hypothesis about “ritual grammar”
Harvesting Event Chains: A Show Case

R1: Offer a lamp (with a burning wick and the mantra) tejo ’si.

R2: Worship of the lamp, the wooden measuring vessel and the key (reciting) agnir mūrdhā divaḥ (and) trātāram indram. One should wave with lamp, wooden measuring vessel and key.

R3: Shower pieces of fruits (etc.) from the measuring vessel (on the head of the boy with) yāḥ phalini. (Make this) three (times). Show (and offer) the lamp (to the boy with) tejo ’si.

R4: (Wave) light (with a burning wick). Now fragrant materials etc. Worship the lamp, the wooden measuring vessel and the key (reciting) agnir mūrdhā (and) trātāram indram. Wave the lamp, the (wooden) measuring vessel and the (iron) key (over the head of the boy reciting) ausraghnām.
R1: *Offer* a lamp (with a burning wick and the mantra) tejo 'si.

R2: *Worship* of the lamp, the wooden measuring vessel and the key (*reciting*) agnir mūrdhā divaḥ (and) trātāram indram. One should *wave* with lamp, wooden measuring vessel and key.

R3: *Shower* pieces of fruits (etc.) from the measuring vessel (on the head of the boy with) yāḥ phalini. (Make this) three (times). *Show (and offer)* the lamp (to the boy with) tejo 'si.

R4: *(Wave) light* (with a burning wick). Now fragrant materials etc. *Worship* the lamp, the wooden measuring vessel and the key (*reciting*) agnir mūrdhā (and) trataram indram. *Wave* the lamp, the (wooden) measuring vessel and the (iron) key (over the head of the boy *reciting*) ausraghnam.
Aligning Sequences

R1: Offer a lamp (with a burning wick and the mantra) tejo 'si.

<table>
<thead>
<tr>
<th>Action 1</th>
<th>Action 2</th>
<th>Vessel</th>
<th>Boy</th>
</tr>
</thead>
<tbody>
<tr>
<td>giving</td>
<td>speaking</td>
<td></td>
<td></td>
</tr>
<tr>
<td>lamp</td>
<td>„tejo‘si“</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Action 2</th>
<th>Vessel</th>
<th>Boy</th>
</tr>
</thead>
<tbody>
<tr>
<td>giving</td>
<td>lamp</td>
<td>„tejo‘si“</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Vessel</th>
<th>Boy</th>
</tr>
</thead>
<tbody>
<tr>
<td>giving</td>
<td>lamp</td>
</tr>
</tbody>
</table>
| „ausra-
ghnam“ |
Aligning Sequences

<table>
<thead>
<tr>
<th>worshipping [-]</th>
<th>speaking [-]</th>
<th>speaking [-]</th>
<th>waving [-]</th>
<th>giving [-]</th>
<th>speaking [-]</th>
</tr>
</thead>
<tbody>
<tr>
<td>lamp, vessel, key</td>
<td>„agnir mūrdhā“</td>
<td>„trātāram indram“</td>
<td>lamp, vessel, key</td>
<td>„tejo‘si“</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>pouring [-]</th>
<th>speaking [-]</th>
<th>showing [-]</th>
<th>giving [-]</th>
<th>speaking [-]</th>
</tr>
</thead>
<tbody>
<tr>
<td>fruit Vessel boy</td>
<td>„yāh phalini“</td>
<td>lamp boy</td>
<td>lamp boy</td>
<td>„tejo‘si“</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>waving [-]</th>
<th>worshipping [-]</th>
<th>speaking [-]</th>
<th>speaking [-]</th>
<th>waving [-]</th>
<th>speaking [-]</th>
</tr>
</thead>
<tbody>
<tr>
<td>lamp</td>
<td>lamp, vessel, key</td>
<td>„agnir mūrdhā“</td>
<td>„trātāram indram“</td>
<td>lamp, vessel, key</td>
<td>„ausra-ghnam“</td>
</tr>
</tbody>
</table>
Combining Sequences

waving [-] lamp [4]
worshipping [-] lamp, vessel, key [2, 4]
speaking [-] “trātāram indram“ [2, 4]
waving [-] lamp, vessel, key [2, 4]
pouring [-] fruit [3], Vessel [3], boy [3]
showing [-] lamp [3], boy [3]
giving [-] lamp [1, 3], boy [3]
Combining Sequences

- Waving [-]
  - Lamp [4]

- Worshipping [-]
  - Lamp, vessel, key [2, 4]

- Speaking [-]
  - "agnir mūrdhā" [2, 4]
  - "yāḥ phalini" [3]

- Speaking [-]
  - "trātāram indram" [2, 4]

- Waving [-]
  - Lamp, vessel, key [2, 4]

- Speaking [-]
  - "tejoʻsi" [1, 3]
  - "ausraghnam" [4]

- Pouring [-]
  - Fruit [3]
  - Vessel [3]
  - Boy [3]

- Showing [-]
  - Lamp [3]
  - Boy [3]

- Giving [-]
  - Lamp [1, 3]
  - Boy [3]
Ritual Element 1?

- Waving lamp [4]
- Worshipping lamp, vessel, key [2, 4]
- Speaking "agnir mūrdhā" [2, 4], "yāḥ phalini" [3]
- Speaking "trātāram indram" [2, 4]
- Waving lamp, vessel, key [2, 4]
- Speaking "tejo’si“ [1, 3], "ausraghnām“ [4]
- Pouring fruit [3], Vessel [3], boy [3]
- Showing lamp [3], boy [3]
- Giving lamp [1, 3], boy [3]
Ritual Element 2?

- waving lamp [4]
- waving lamp, vessel, key [2, 4]
- speaking [-]
  - "agnir mūrdhā" [2, 4]
  - "yāḥ phalini" [3]
- speaking [-]
  - "trātāram indram" [2, 4]
- waving [-]
  - lamp, vessel, key [2, 4]
- speaking [-]
  - "tejo‘si" [1, 3]
  - "ausragham" [4]
- pouring [-]
  - fruit [3]
  - Vessel [3]
  - boy [3]
- showing [-]
  - lamp [3]
  - boy [3]
- giving [-]
  - lamp [1, 3]
  - boy [3]
Ritual Element 3?

- Waving 
  - Lamp [4]
- Worshipping 
  - Lamp, vessel, key [2, 4]
- Speaking 
  - "agnir mūrdhā“ [2, 4]
  - "yāḥ phalini“ [3]
- Speaking 
  - "trātāram indram“ [2, 4]
- Waving 
  - Lamp, vessel, key [2, 4]
- Speaking 
  - "tejo‘si“ [1, 3]
  - "ausraghnām“ [4]
- Pouring 
  - Fruit [3]
  - Vessel [3]
  - Boy [3]
- Showing 
  - Lamp [3]
  - Boy [3]
- Giving 
  - Lamp [1, 3]
  - Boy [3]
## Calculating Bigrams

<table>
<thead>
<tr>
<th></th>
<th>give</th>
<th>speak</th>
<th>worship</th>
<th>wave</th>
<th>throw</th>
<th>pour</th>
<th>show</th>
</tr>
</thead>
<tbody>
<tr>
<td>give</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td></td>
<td>1</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>speak</td>
<td>2</td>
<td></td>
<td></td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>worship</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>wave</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>throw</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pour</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>show</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Ritual Element Candidates

Co-occurring Bigrams

- give – speak
- speak – speak
- worship – speak
- speak – wave

Found by Inspection

- worship – speak – speak – wave
- give – speak
Overview

Motivation
Studying Rituals
Detecting Structure

Processing Ritual Descriptions
Data Collection and Annotation
Challenges: Domain Adaptation
Chances: Aggregation and Exploitation

Conclusions
Challenges and Chances

Domain Adaptation for Preprocessing

- Harvesting events depends on *high-quality preprocessing*
- Standard tools perform poorly on the ritual domain
- *Augmenting the feature space and balancing training data* achieves significant performance gains
Challenges and Chances

Domain Adaptation for Preprocessing

- Harvesting events depends on *high-quality preprocessing*
- Standard tools perform poorly on the ritual domain
- *Augmenting the feature space* and *balancing training data* achieves significant performance gains

Semantic Annotation and Automation

- Frame semantics can be used to model ritual event sequences
- A large portion of FrameNet frames can be used, some frames need to be created
- *Challenges for automation*: adapting SRL, concept labeling and coreference resolution to ritual domain
Prospects

'We are not alone…'

- Domain adaptation for event-based semantic processing is crucial for many applications
Prospects

'We are not alone…'

- Domain adaptation for event-based semantic processing is crucial for many applications

Progress on all aspects of annotation

- Domain adaptation – Manual annotation – Automation
Prospects

'We are not alone…’

- Domain adaptation for event-based semantic processing is crucial for many applications

Progress on all aspects of annotation

- Domain adaptation – Manual annotation – Automation

Keep pursuing our aims

- Establish new ways of conducting empirical research in the social sciences using computational linguistic techniques
Acknowledgements

Our collaborators:

Oliver Hellwig, Anand Mishra, Axel Michaels

Our research assistants:

Irina Gossmann, Borayin Maitreya Larios, Julio Cezar Rodrigues, Britta Zeller

SFB 619
Ritualdynamik
Thank you.


