Data-Driven and Ontological Analysis of FrameNet for Natural Language Reasoning

Ekaterina Ovchinnikova¹, Laure Vieu²,³, Alessandro Oltramari², Stefano Borgo², Theodore Alexandrov⁴

¹University of Osnabrück, ²LOA-ISTC-CNR Trento, ³IRIT-CNRS Toulouse, ⁴University of Bremen

May, 20th - LREC, Valetta
Introduction

- **Lexical-semantic** knowledge for reasoning
  - **WordNet** [Morato et al., 2004]
    - search
    - information extraction
    - ...
  - **FrameNet**
    - question answering [Shen and Lapata, 2007]
    - recognizing textual entailment [Burchardt et al., 2009]
    - ...
Introduction

- Shortcomings of FrameNet with regard to NL reasoning
  - low coverage [Shen and Lapata, 2007; Cao et al., 2008]
  - conceptual inconsistency and lack of axiomatization

Our focus
methodology for improving the conceptual structure of FrameNet for the goals of NL reasoning
Outline

1. FrameNet for Reasoning
2. Proposed Methodology
3. Conceptual Problems
4. Data-Driven Analysis
5. Ontological Analysis
6. Case Study
7. Conclusion
Outline

1. FrameNet for Reasoning
2. Proposed Methodology
3. Conceptual Problems
4. Data-Driven Analysis
5. Ontological Analysis
6. Case Study
7. Conclusion
FrameNet for reasoning

NP subject
PP with by
...

to give, to hand over, ...

Giving frame

donor

theme

recipient

FrameNet for reasoning

(a) \([John]_{\text{DONOR}} \ [gave]_{\text{Giving}} \ [Mary]_{\text{RECIPIENT}} \ [a \ flower]_{\text{THEME}}\)

(b) \([Mary]_{\text{RECIPIENT}} \ [got]_{\text{Getting}} \ [a \ flower]_{\text{THEME}} \ [from \ John]_{\text{SOURCE}}\)
Frame relations

1. Inheritance: 441
   - Vehicle – Artifact, Motion_directional – Motion

2. Precedence: 55
   - Being_awake – Fall_asleep

3. Perspective: 43
   - Buy, Sell – Goods_transfer

4. Causation: 49
   - Giving – Getting

5. Subframe: 87
   - Trial, Sentencing – Criminal_process

6. Using: 426
   - Recovery – Medical_conditions

7. See_also: 669
   - Scrunity – Seeking
Research goals

1. Axiomatizing frame relations
2. Finding missing frame relations
3. Cleaning up frame relations
4. Applying frame relations to NL reasoning
Outline

1. FrameNet for Reasoning
2. Proposed Methodology
3. Conceptual Problems
4. Data-Driven Analysis
5. Ontological Analysis
6. Case Study
7. Conclusion
Proposed improvement methodology

1. Conceptual problems in FrameNet: Frame-Annotated Corpus for Textual Entailment (FATE)
2. Clustering frames
3. Ontological analysis of frames and frame relations
   - axiomatizing frame relations
   - constraints on frame relations
4. Evaluation: enriched, axiomatized and cleaned up frame relations in RTE
Outline

1. FrameNet for Reasoning
2. Proposed Methodology
3. Conceptual Problems
4. Data-Driven Analysis
5. Ontological Analysis
6. Case Study
7. Conclusion
Frame-Annotated Corpus for Textual Entailment

FATE [Burchardt & Pennacchiotti, 2008]

- 800 $T$-$H$ entailment pairs annotated with FrameNet frames and roles

- we have analyzed cases when $T$ was known to entail $H$ (400 pairs) applying a frame matching strategy
FATE analysis results

- **170 pairs**: matching is possible
- **131 pairs**: this approach does not work
  - annotation disagreements
  - different conceptualizations of $T$ and $H$
- **99 pairs**: the same facts in $T$ and $H$ are represented by different frames which are related semantically and could be mapped on each other with the help of reasoning
  - FrameNet enables inferences only for 17 pairs
Discovered problems

1. missing relations
   (t) \( \ldots X [\text{survived}]_{\text{Surviving}} \text{Sars}\ldots \)
   (h) \( \ldots X [\text{recovered}]_{\text{Recovery}} \text{from Sars}\ldots \)

2. problems in the relational structure
   \( \ldots [\text{parts}]_{\text{Part\_whole}} \text{[of Aceh province]}_{\text{WHOLE}} \ldots \)
   \text{Part\_whole} \rightarrow \text{Part\_piece}, \text{WHOLE} \rightarrow \text{SUBSTANCE}

3. missing axiomatization of the relations
   (t) \( \ldots X [\text{recovered}]_{\text{Recovery}} \text{from Sars}\ldots \)
   (h) \( \ldots X [\text{was ill}]_{\text{Medical\_conditions}} \ldots \)
   \text{Recovery uses Medical\_conditions}
Outline

1. FrameNet for Reasoning
2. Proposed Methodology
3. Conceptual Problems
4. Data-Driven Analysis
5. Ontological Analysis
6. Case Study
7. Conclusion
Clustering frames

For every two frames $f_1$ and $f_2$ we apply similarity measures based on \cite{Pennacchiotti2009}:

1. overlapping frame elements in $f_1$ and $f_2$
2. co-occurrence of lexemes evoking $f_1$ and $f_2$ in corpora (pmi)
Clustering results

1. Clusters based on overlapping frame elements
   - 228 clusters in total
   - 1497 relations not contained in FrameNet
   - 73 clusters from 100 random contain semantically related frames (2 experts, agreement 0.85)

2. Clusters based on co-occurrence of lexemes evoking frames
   - 113 clusters in total
   - 1149 relations not contained in FrameNet
   - 65 clusters from 100 random contain semantically related frames (2 experts, agreement 0.85)
Outline

1. FrameNet for Reasoning
2. Proposed Methodology
3. Conceptual Problems
4. Data-Driven Analysis
5. **Ontological Analysis**
6. Case Study
7. Conclusion
Frames and situations

*What do frames describe?*

- Frames abstract from natural language expressions (predicates with their arguments)
- Natural language expressions describe situations
- Frames can be seen as abstractions from situations
Types of situations

From which types of situations do frames abstract?

- categories from the DOLCE ontology [Masolo et al., 2002] for describing types of situations

Types of situations:

1. "Event" situation
   - e.g. Motion (John is running in the park)

2. "Object" situation
   - e.g. People (A man)

3. "Quality" situation
   - e.g. Color (This rose is red)

4. "Relation" situation
   - e.g. Part_whole (This park is a part of the town)
Situations and time

1. Situations having temporal qualities
   - *John is running in the park, a clerk, This rose is red, John is next to Mary*
   - can participate in temporal relations (*precedence, temporal inclusion* etc.)

2. Non-temporal situations
   - *A man, The war lasted four years, Einstein's birth preceded my birth*
   - cannot participate in temporal relation
Causation: $f_1$ is causative of $f_2$

\[ \forall s_1 (f_1(s_1) \rightarrow \exists s_2 (f_2(s_2) \land \text{causes}(s_1, s_2))) \]

\[ \forall s_1 s_2 (\text{causes}(s_1, s_2) \rightarrow \neg \text{starts\_before}(s_2, s_1))) \]
Subframe: \( f_1 \) is subframe of \( f_2 \)

1. **Subframe of “Events”**
\[
\forall s_1 s_2 (\text{sub}\_\text{ev}(s_1, s_2) \rightarrow (\text{strict}\_\text{temp}\_\text{inc}(s_2, s_1) \land \text{spatially}\_\text{inc}(s_2, s_1)))
\]
   - part presupposes whole
     \[
     \forall s_1 (f_1(s_1) \rightarrow \exists s_2 (f_2(s_2) \land \text{sub}\_\text{ev}(s_1, s_2)))
     \]
   - whole presupposes part
     \[
     \forall s_2 (f_2(s_2) \rightarrow \exists s_1 (f_1(s_1) \land \text{sub}\_\text{ev}(s_1, s_2)))
     \]

2. **Subframe of “Objects”**
   - part presupposes whole
     \[
     \forall s_1 \text{en}_1 (f_1(s_1) \land \text{FE}_1(s_1, \text{en}_1) \rightarrow \exists s_2 \text{en}_2 (f_2(s_2) \land \text{FE}_2(s_2, \text{en}_2) \land \text{part}\_\text{of}(\text{en}_1, \text{en}_2)))
     \]
   - whole presupposes part
     \[
     \forall s_2 \text{en}_2 (f_2(s_2) \land \text{FE}_2(s_2, \text{en}_2) \rightarrow \exists s_1 \text{en}_1 (f_1(s_1) \land \text{FE}_1(s_1, \text{en}_1) \land \text{part}\_\text{of}(\text{en}_1, \text{en}_2)))
     \]
Using and See_also

- the most frequent relations in FN
- sometimes can be represented in terms of other axiomatized relations
- otherwise
  \[ \forall s_1 (f_1(s_1) \rightarrow \exists s_2 (f_2(s_2) \land \text{depends}(s_1, s_2))) \]
- often represent typical rather than necessary dependence (e.g. Medical_professionals–Cure)
Mapping frame elements

If $f_1$ is related to $f_2$ with a relation in FN then

$$\forall s_1 \ s_2 ( (f_1(s_1) \land f_2(s_2)) \rightarrow \ (rel(s_1, s_2) \leftrightarrow \forall x (FE_1(x, s_1) \leftrightarrow FE_2(x, s_2))),$$

where $FE_1$ in $f_1$ is mapped to $FE_2$ in $f_2$. 
Example

\[ \forall s_1 (\text{Giving}(s_1) \rightarrow \exists s_2 (\text{Getting}(s_2) \land \text{causes}(s_1, s_2))) \]

\[ \forall s_1 s_2 ((\text{Giving}(s_1) \land \text{Getting}(s_2)) \rightarrow
(\text{causes}(s_1, s_2) \leftrightarrow \forall x (\text{DONOR}(x, s_1) \leftrightarrow \text{SOURCE}(x, s_2))) \]
Cleaning up constraints

Given frames $f_1$ and $f_2$ connected with a relation $r$

1. define the types of situations that instantiate $f_1$ and $f_2$
2. if $r$ is a temporal relation, make sure that both $f_1$ and $f_2$ refer to "temporal" situations
3. define whether $r$ has a typical or a necessary character
4. check whether the frame relation axioms apply to all instantiations of $f_1$ and $f_2$
Case Study: „medical cluster“
Enriched and cleaned up „medical“ cluster
„Medical“ cluster in RTE

- **39** \(T-H\) entailment pairs (**18** true entailments) annotated in FATE with „medical“ frames
- TE computed by the Nutcracker system [Bos & Markert, 2006]

<table>
<thead>
<tr>
<th></th>
<th>NFA</th>
<th>FA</th>
<th>FA&amp;A</th>
<th>FA&amp;CA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correct proofs</td>
<td>1</td>
<td>4</td>
<td>7</td>
<td>10</td>
</tr>
<tr>
<td>Wrong proofs</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Overall accuracy</td>
<td>0.56</td>
<td>0.5</td>
<td>0.61</td>
<td>0.78</td>
</tr>
</tbody>
</table>

**Problems:**
- Incompleteness of the FATE annotation: **8**
- Nutcracker processing errors: **5**
- Lack of general non-definitional knowledge: **7**
Outline

1. FrameNet for Reasoning
2. Proposed Methodology
3. Conceptual Problems in FrameNet
4. Data-Driven Analysis of FrameNet
5. Ontological Analysis of FrameNet
6. Case Study
7. Conclusion
Conclusion

1. Presented
   i. Conceptual problems in FN
   ii. Methodology for improvement
      • data-driven analysis
      • ontological analysis
   iii. Case study

2. Lessons learned
   i. Many useful relations can be acquired automatically
   ii. Axiomatization helps
   iii. RTE is still a difficult task
      • difficult to create an appropriate annotation
      • difficult to provide necessary knowledge
      • difficult to find a proof
Ongoing and future work

1. Automatic relation extraction
   - automatic mapping of frame roles
   - detecting type of the relation

2. Ontological analysis
   - applying OntoClean to FN hierarchy of frames and roles

3. FrameNet in RTE
   - applying frame relations to a full RTE set
   - using frame similarity measures for weighting axioms
   - using probabilistic reasoning
Thank you!
Any questions?