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

Ekaterina Ovchinnikova[1], Laure Vieu[2,3], Alessandro Oltramari[2], Stefano Borgo[2], Theodore Alexandrov[4]

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

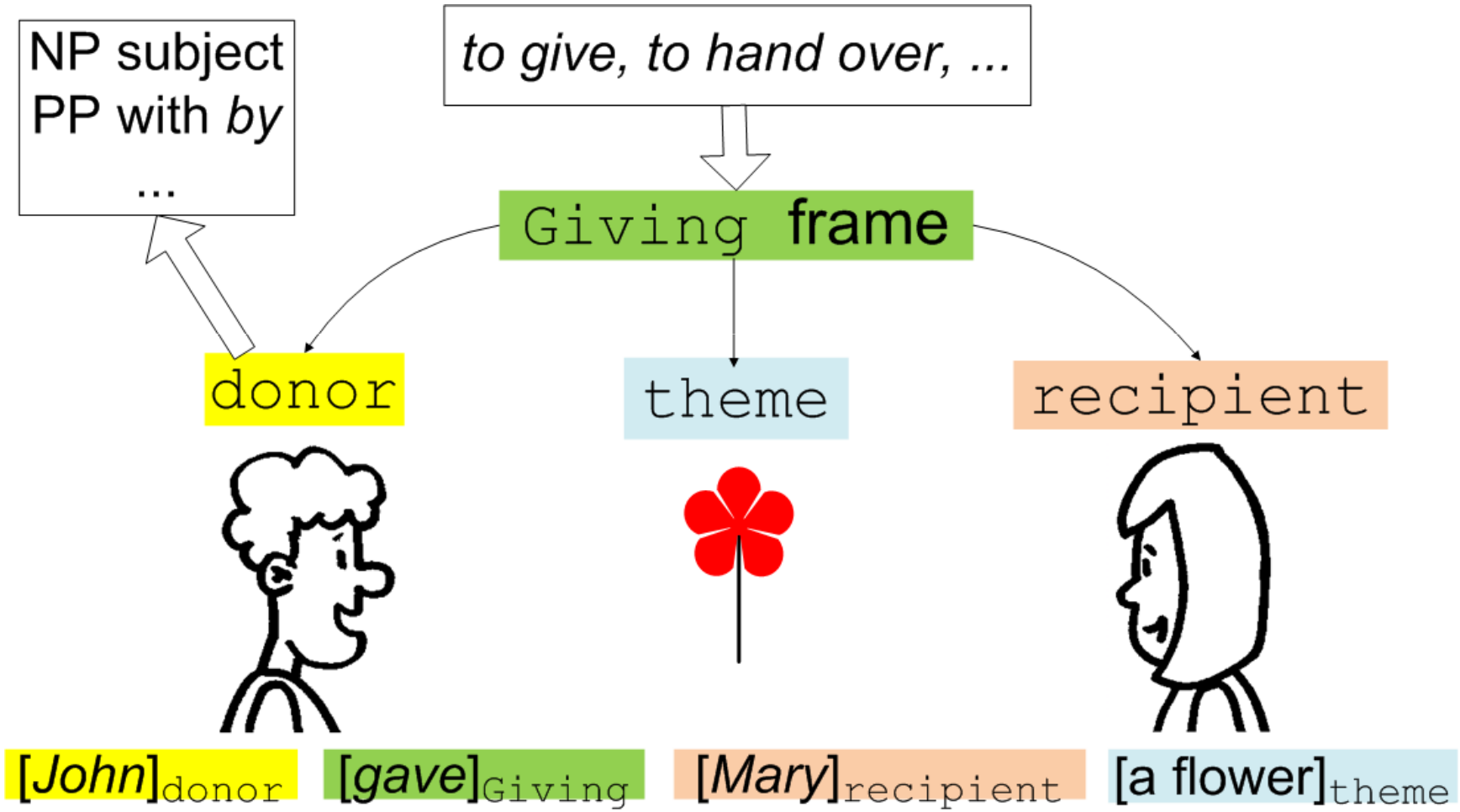
Outline

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

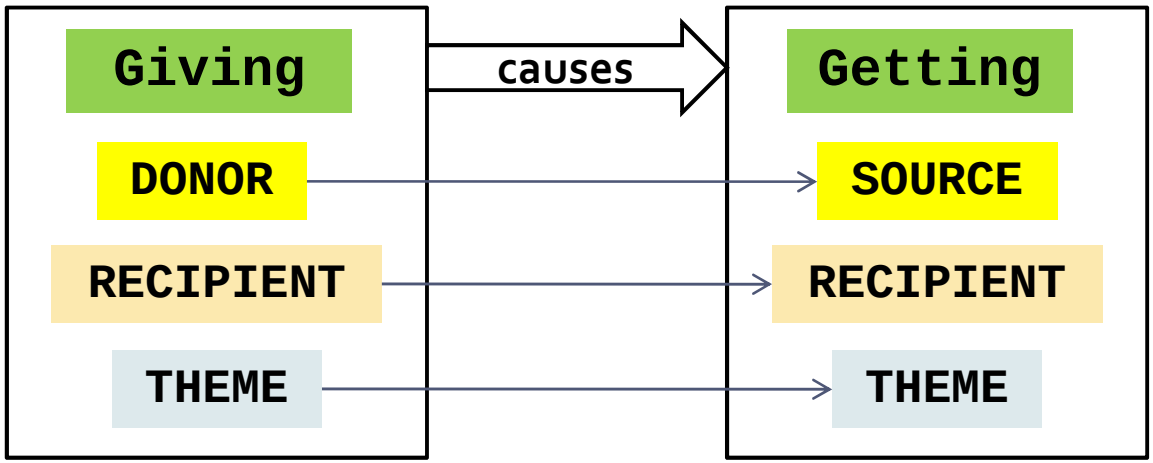
FrameNet for reasoning



FrameNet for reasoning

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

(b) [*Mary*]$_{\texttt{RECIPIENT}}$ [*got*]$_{\texttt{Getting}}$ [*a flower*]$_{\texttt{THEME}}$ [*from John*]$_{\texttt{SOURCE}}$



Frame relations

1. Inheritance: 441
   - `Vehicle – Arfitact, 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 analized 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) …*X* [*survived*]$_{\texttt{Surviving}}$ *Sars*…

   (h) …*X* [*recovered*]$_{\texttt{Recovery}}$ *from Sars*…

2. problems in the relational structure

   …[*parts*]$_{\texttt{Part\_whole}}$ [*of Aceh province*]$_{\texttt{WHOLE}}$…

   Part_whole → **Part_piece**, **WHOLE** → **SUBSTANCE**

3. missing axiomatization of the relations

   (t) …*X* [*recovered*]$_{\texttt{Recovery}}$ *from Sars…*

   (h) …*X* [*was ill*]$_{\texttt{Medical\_conditions}}$…

   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

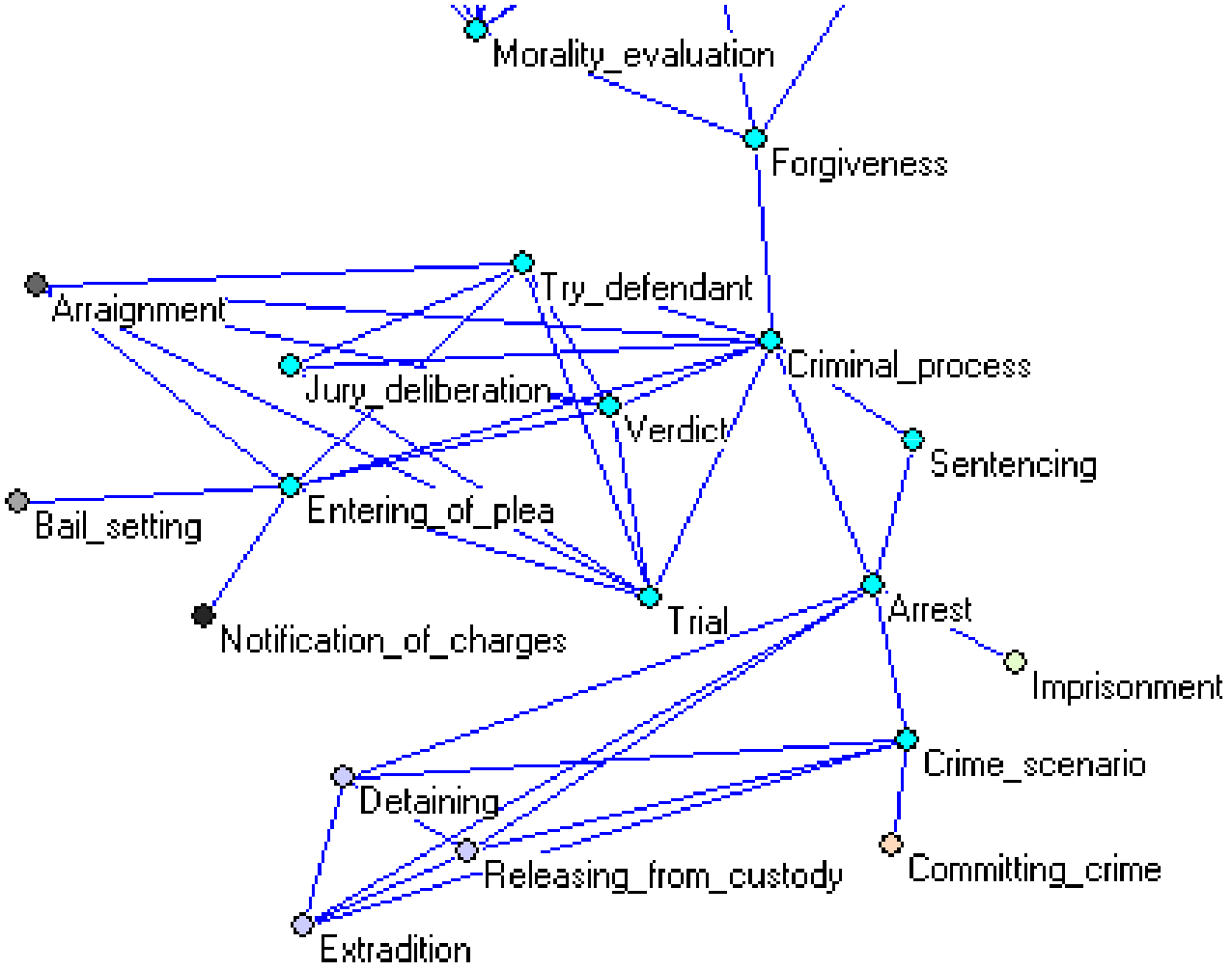
For every two frames *f1* and *f2* we apply similarity measures based on [*Pennacchiotti & Wirdth, 2009*] :

1. overlapping frame elements in *f1* and *f2*
2. co-occurrence of lexemes evoking *f1* and *f2* 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-occurence 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)

Frame clusters: visualization (Pajek tool)



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

$$\forall s_1 \, (f_1(s_1) \rightarrow \exists \, s_2 \, (f_2(s_2) \wedge \mathit{causes}(s_1, s_2)))$$

$$\forall s_1 \, s_2 \, (\mathit{causes}(s_1, s_2) \rightarrow \neg \, \mathit{starts\_before}(s_2, \, s_1)))$$

Subframe: f_1 *is subframe of* f_2

1. Subframe of "Events"

$\forall s_1\, s_2(sub\_ev(s_1,s_2) \rightarrow (strict\_temp\_inc(s_2,s_1) \wedge spatially\_inc(s_2,s_1)))$

- part presupposes whole

$\forall s_1\, (f_1(s_1) \rightarrow \exists\, s_2\, (f_2(s_2) \wedge sub\_ev(s_1,s_2)))$

- whole presupposes part

$\forall s_2\, (f_2(s_2) \rightarrow \exists\, s_1\, (f_1(s_1) \wedge sub\_ev(s_1,s_2)))$

2. Subframe of "Objects"

- part presupposes whole

$\forall s_1 en_1 (f_1(s_1) \wedge FE_1(s_1, en_1) \rightarrow \exists\, s_2 en_2\, (f_2(s_2) \wedge FE_2(s_2, en_2) \wedge part\_of(en_1,en_2)))$

- whole presupposes part

$\forall s_2 en_2 (f_2(s_2) \wedge FE_2(s_2, en_2) \rightarrow \exists\, s_1 en_1\, (f_1(s_1) \wedge FE_1(s_1, en_1) \wedge part\_of(en_1,en_2)))$

Using and See_also

- the most frequent relations in FN
- sometimes can be represented in terms of other axiomatized relations
- otherwise

$$\forall s_1 (f_1(s_1) \rightarrow \exists s_2 (f_2(s_2) \wedge \mathit{depends}(s_1, s_2)))$$

- 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

$$\forall s_1\, s_2\, ((f_1(s_1) \wedge f_2(s_2)) \rightarrow (\mathit{rel}(s_1,s_2) \leftrightarrow \forall\, x\, (FE_1(x,s_1) \leftrightarrow FE_2(x,s_2))),$$

where FE_1 in f_1 is mapped to FE_2 in f_2 .

Example

$$\forall s_1\,(\textbf{Giving}(s_1) \rightarrow \exists\, s_2\,(\textbf{Getting}(s_2) \wedge \mathit{causes}(s_1,s_2)))$$

$$\forall s_1\, s_2\,((\textbf{Giving}(s_1) \wedge \textbf{Getting}(s_2)) \rightarrow$$
$$(\mathit{causes}(s_1,s_2) \leftrightarrow \forall\, x\,(\textbf{DONOR}(x,s_1) \leftrightarrow \textbf{SOURCE}(x,s_2)))$$

Cleaning up constraints

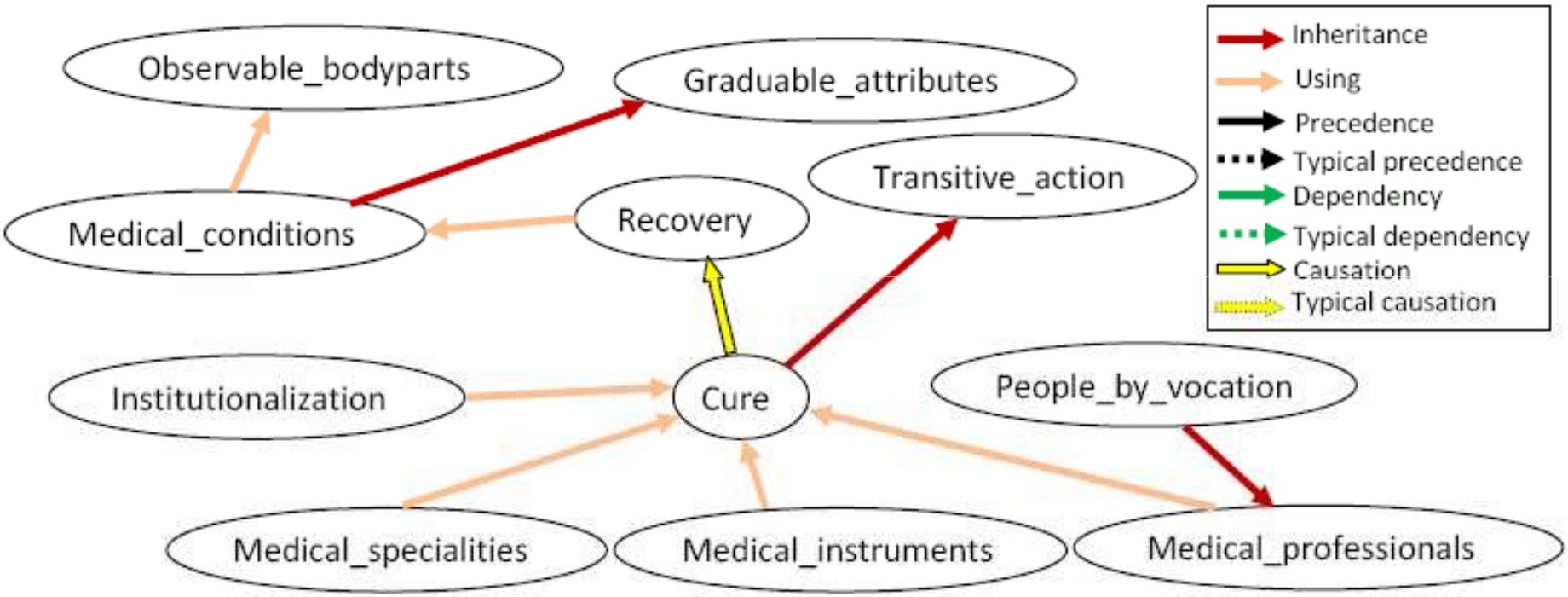
Given frames f_1 and f_2 connected with a relation r

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

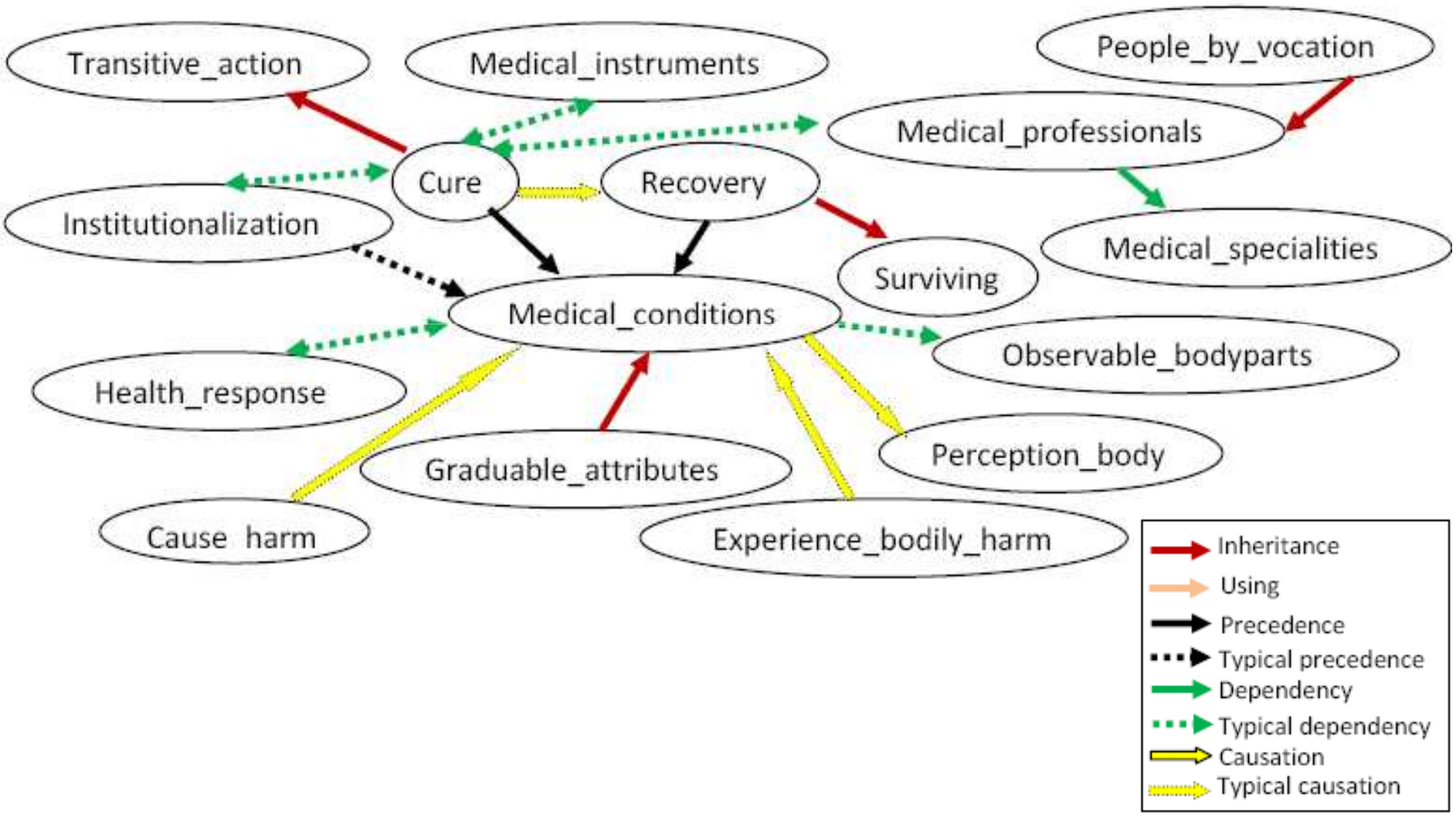
Outline

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

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*]

	NFA	FA	FA&A	FA&CA
Correct proofs	1	4	7	10
Wrong proofs	1	1	1	1
Overall accuracy	0.56	0.5	0.61	0.78

- 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 usefull relations can be aquired 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 hierachy of frames and roles

3. FrameNet in RTE
   - applying frame relations to a full RTE set
   - using frame similarity measures for weighting axioms
   - using probabalistic reasoning

Thank you!

Any questions?