

GL2009

Sept 17-19, 2009 — Pisa, Italy

**Studying the role of Qualia Relations
for Word Sense Disambiguation**

Antonio Toral, Monica Monachini
{firstname.lastname}@ilc.cnr.it
Istituto di Linguistica Computazionale, Pisa

Aitor Soroa, German Rigau
{a.soroa,german.rigau} [at] ehu.es
University of the Basque Country

Outline

- Intro
- Graph-based Word Sense Disambiguation
- Evaluation
- Proposal: Extract new Qualia Relations
- Conclusions

Intro

- Word Sense Disambiguation (WSD)
 - Disambiguate polysemous words
- e.g. “I like the **bass** in that song”
 - Senses:
 - “The lowest part in polyphonic music”
 - “The lean flesh of a saltwater fish of the family Serranidae”
 - ...

Intro (II)

- Generative Lexicon (GL)
 - Ling theory: sense as bundle of orthogonal dimensions → express multidimensionality of word meaning
 - Qualia Structure: represent lexical semantics of sense, 4 roles:
 - Formal → taxonomy
 - Constitutive → physical properties
 - Agentive → origin
 - Telic → function or purpose

Intro (III)

- What is the role of Qualia Relations in WSD?
- Study the impact of each Qualia Role in graph-based WSD approach
- Enrich LRs automatically

Graph-based WSD

- UKB: knowledge-based WSD system, exploits structure or pre-existing LR
- Represents LR as graph
 - Vertices → senses
 - Edges → relations between senses
- Given a text → applies Personalised PageRank → obtains most representative senses

Graph-based WSD (II)

- Apply UKB over ItalWordNet (IWN)
 - Graph: 49k vertices, 64k edges
- PAROLE-SIMPLE-CLIPS (PSC)
 - Italian GL-based Lexicon → Qualia Relations, ...
- Mapping IWN-PSC
 - Extend graph with Qualia Relations
 - e.g. telic “scalpel *used_for* engrave”

Evaluation

•Evaluation Framework

- Data from Evalita-2007 all-words task
 - 5k word Corpus annotated with IWN senses
 - 2.5k nouns, 1.8k verbs, 0.7k adj
 - Measures: precision, recall
- Confidence intervals for precision → check statistical significance

Evaluation (II)

Configurations

set	rels	vertices	edges
IWN	128,517	49,263	64,258
IWN+PSC	166,884	49,361	77,580
IWN+PSC q	163,222	49,360	76,716
IWN+PSC f	148,640	49,357	70,466
IWN+PSC a	132,801	49,278	65,942
IWN+PSC c	134,964	49,297	66,734
IWN+PSC t	133,808	49,295	66,391

Evaluation (III)

Results

set	P	R	interval
IWN	61.8	50.6	[59.8,64.2]
IWN+PSC	61.3	50.3	[59.3,63.5]
IWN+PSC q	61.5	50.4	[59.2,63.4]
IWN+PSC f	62.0	50.8	[59.2,64.0]
IWN+PSC a	61.4	50.3	[59.1,63.4]
IWN+PSC c	61.2	50.1	[59.1,63.6]
IWN+PSC t	62.0	51.7	[59.9,64.3]

Evaluation (IV)

Comparison to other systems

system	P	R	interval
UKB	62.0	51.7	[59.9,64.3]
JIGSAW	56.0	41.4	[53.8,58.2]
first sense	66.9	66.9	[65.1,68.7]

Evaluation Lessons

- Addition of formal and telic relations improves* the results
- Same improvement but telic rels added (5k) much less than formal (20k)
- → Gather further telic relations to obtain conclusive results

Extract more relations

- Proposal: Pattern-based extraction of telic relations
 - Search in web word pairs that instantiate a relation in SIMPLE
 - Analyse sentences gathered (PoS, chunking, NER, ontology tags, ...)
 - Generalise patterns that express the relation
 - Extract further word pairs with patterns
 - Add to SIMPLE and apply to WSD

Kybots

- **Knowledge-Yielding Robots**
 - Data miners. Defined using constraints from different linguistic levels (KAF)
 - Used in KYOTO to extract facts
- Use Kybots to extract relations
 - Expression rules: morpho-syntactic and semantic conditions on sequences of text
 - Output: source and target terms

Example

(radio, listen)

a radio that allowed me to listen to ...

(lens, observe)

this lens allows to observe ...

Example

(radio, listen)

a radio that allowed me to listen to ...

N*

PR*

allow

PP* to

V*

(lens, observe)

this lens allows to observe ...

Example

(radio, listen)

a radio that allowed me to listen to ...

N*

PR*

allow

PP* to

V*

(lens, observe)

this lens allows to observe ...

N*

allow

to

V*

Example

(radio, listen)

a radio that allowed me to listen to ...

N*

PR*

allow

PP* to

V*

(lens, observe)

this lens allows to observe ...

N*

allow

to

V*

Generalised pattern: “**N*** (PR*)? allow (PP*)? to **V***”

```
<Kybot id="used-for">
```

```
<variables>
```

```
<var name="X" type="term" pos="N*"/>
```

```
<var name="Y" type="term" pos="PR*"/>
```

```
<var name="Z" type="term" lemma="allow" pos="V*"/>
```

```
<var name="U" type="term" pos="PP*"/>
```

```
<var name="V" type="term" token="to"/>
```

```
<var name="W" type="term" pos="V*"/>
```

```
</variables>
```

```
<relations>
```

```
<root span="X"/>
```

```
<rel span="Y" pivot="X" direction="following" dist = "1" opt="yes"/>
```

```
<rel span="Z" pivot="Y" direction="following" dist = "1"/>
```

```
<rel span="U" pivot="Z" direction="following" dist = "1" opt="yes"/>
```

```
<rel span="V" pivot="U" direction="following" dist = "1"/>
```

```
<rel span="W" pivot="V" direction="following" dist = "1"/>
```

```
</relations>
```

```
<output>
```

```
<fact id="used-for-rel">
```

```
<source value="$X/@lemma"/>
```

```
<target value="$W/@lemma"/>
```

```
</fact>
```

```
</output>
```

```
</Kybot>
```

Example (III)

What can we extract with the pattern?

Five **jobs** that allow you to **work** at home

a **scholarship** that allows him to **achieve** a doctorate

Conclusions

- Studied impact of different qualia roles in WSD
 - Telic and formal increase* results
- Proposal: Apply multi-level ling patterns (with Kybots) → extract more telic rels → get conclusive results
 - Side effect: enrichment of PSC, ItalWordNet

End

Thank you for listening!
Questions, suggestions?

Antonio Toral, Monica Monachini
{firstname.lastname}[at]ilc.cnr.it
Istituto di Linguistica Computazionale, Pisa

Aitor Soroa, German Rigau
{a.soroa,german.rigau} [at] ehu.es
University of the Basque Country