Studying the role of Qualia Relations for Word Sense Disambiguation

Antonio Toral **Monica Monachini**
Istituto di Linguistica Computazionale
Consiglio Nazionale delle Ricerche
Pisa, Italy
firstname.lastname@ilc.cnr.it

Aitor Soroa **German Rigau**
Informatika Fakultatea
University of the Basque Country
Donostia, Spain
{a.soroa,german.rigau}@ehu.es

Abstract

This paper studies the importance of qualia relations for Word Sense Disambiguation (WSD). We use a graph-based WSD algorithm over the Italian WordNet and evaluate it when adding different kinds of qualia relations (agentive, constitutive, formal and telic) taken from PAROLE-SIMPLE-CLIPS (PSC), a Language Resource based on the Generative Lexicon theory. Some qualia relations, specially telic, appear to have a positive impact on the results despite their low coverage in PSC. Therefore we propose to extract further such relations from the web by applying multi-level patterns following the so-called Kybot model, as by doing so it is expected to improve the WSD performance.

1 Introduction

The enormous amount of written language data, mainly due to the continuous growth of the internet, poses a requirement, that of automatic means of processing in order to allow users to perform tasks such as Question Answering. The Natural Language Processing research area plays a privileged role as it provides the necessary infrastructure to automatically derive analysis of the several linguistic levels (morphologic, syntactic, semantic, etc.).

Semantic processing, usually built on top of morpho-syntactic analysis, is a must to attain language understanding. A key task in this respect is Word Sense Disambiguation (WSD), whose aim is to identify which sense of a word is used in a given context when the word has a number of distinct senses. For example, consider the sentence "I like the bass in that song" and the polysemous word bass, with different meanings such as (taken from WordNet):

- the lowest part in polyphonic music
- the lean flesh of a saltwater fish of the family Serranidae

It is obvious then that in order to understand the sentence, the automatic analysis should decide to which of the senses of bass the sentence refers to.

The Generative Lexicon (GL) is a linguistics theory which sees the sense as a complex bundle of orthogonal dimensions that express the multidimensionality of word meaning. The most important component for representing the lexical semantics of a word sense is the qualia structure which consists of four qualia roles:

- Formal. Makes it possible to identify an entity.
- Constitutive. Expresses the constitution of an entity.
- Agentive. Provides information about the origin of an entity.
- Telic. Specifies the function of an entity.

This paper studies the importance of qualia relations for WSD. The rest of the paper is structured as follows. Section 2 describes the WSD system we have used and how it was set up. Next, section 3 reports on the results obtained by this system when exploiting different types of relations. Subsequently, we propose the extraction of telic relations in section 4. Finally, we derive conclusions and propose future work lines.

2 Graph-based WSD

Our WSD system is based on the application of UKB[1] to the Italian WordNet (IWN). UKB is a knowledge-based unsupervised WSD system

[1] http://ixa2.si.ehu.es/ukb/

which exploits the structure of an underlying Language Resource (LR) and finds the most relevant concepts given an input context (Agirre and Soroa, 2009). UKB starts by taking the LKB as a graph of concepts $G = (V, E)$ with a set of vertices V derived from LKB senses and a set of edges E representing relations between senses. Giving an input context, UKB applies the so called *Personalized PageRank* over it to obtain the most representative senses for the context.

PageRank (Brin and Page, 1998) is a method for scoring the vertices V of a graph according to each node's structural importance. The algorithm can be viewed as random walk process that postulate the existence of a particle that randomly traverses the graph, but at any time may jump to a new vertex with a given *teleport probability*. After PageRank calculation, the final weight of node i represents the proportion of time that a random particle spends visiting node i after a sufficiently long time. In standard PageRank, the teleport vector is chosen uniformly, whereas for Personalized PageRank it is chosen from a nonuniform distribution of nodes, specified by a *teleport vector* (Haveliwala, 2002).

UKB concentrates the initial probability mass of the teleport vector in the words coocurring with the target word in a context, causing all random jumps on the walk to return to these context word and thus assigning a higher rank to the senses linked to these words. Moreover, the high rank of the words will spread through the links in the graph and make all the nodes in its vicinity also receive a high rank. Given a target word, it suffices to check which is the relative ranking of its senses, and the WSD system would output the one ranking highest.

The first step consisted on converting IWN to the graph input format required by UKB, which leaded to a graph made up of 49,263 vertices and 64,258 edges. Subsequently, we can increase the number of connections in the graph with more relations by exploiting the mapping between IWN and PAROLE-SIMPLE-CLIPS (PSC) (Ruimy et al., 2008). IWN (Alonge et al., 1999) is an Italian LR created in the framework of the EuroWordNet project. PSC (Ruimy et al., 2002) is also an Italian LR, this one is based on the GL theory. A core element of this lexicon is the qualia structure, which enables to express different or orthogonal aspects of word sense. Within PSC, the qualia structure was extended by assigning subtypes to each of the qualia roles (e.g. "Usedfor" and "Usedby" are subtypes of the telic role).

We proceed as follows; for each type of qualia relation we extract all the wordsense pairs that instantiate this relation in PSC. For each wordsense in each pair we obtain the corresponding IWN synset from the mapping (and thus the corresponding vertex in the graph). Finally, we add to the graph an edge between the two vertices that correspond to the wordsense pair. It can be the case that a PSC wordsense does not have a corresponding IWN synset or that it has more than one. In the first scenario, no edge is added, while in the second more than one edge are added.

3 Evaluation

3.1 Data and Measures

The evaluation data is an Italian corpus annotated with IWN senses for the Senseval-3 and EVALITA WSD tasks (Ulivieri et al., 2004). This corpus was extracted from the Italian Syntactic Semantic Treebank, which comprises a selection of newspaper articles concerning several topics: politics, economy, sports, etc, and contains 5,000 words. The annotation was restricted to nouns (2,583), verbs (1,858), adjectives (748), multiword expressions (97) and a set of general proper nouns (163).

Results were evaluated by taking into account the following standard measures: Precision (P) and Recall (R). The evaluation was carried our by using the automatic scorer used for the Senseval-2 and Senseval-3 exercises. We also include the confidence intervals for each precision score, these have been computed by using bootstrap resampling with 95% confidence.

3.2 Results

We have evaluated UKB with different sets of relations. The first configuration ("IWN") includes the relations of IWN. All the rest include, apart from the relations of IWN, different sets of relations from PSC: "IWN+PSC" includes all the relations of PSC, "IWN+PSC q" includes only the qualia relations of PSC. In order to gain a clearer insight at the role of qualia relations, we consider adding one type at a time: "IWN+PSC f" includes formal relations, "IWN+PSC a" agentive, "IWN+PSC c" consitutive and "IWN+PSC t" telic. Table 1 shows the graph properties (number of vertices and edges) for each of these configurations.

Table 1: Graph properties of the different 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

Table 2 shows the impact that the addition of relations from PSC has on the score obtained by UKB.

Table 2: Results adding relations from PSC

set	P	R	conf. 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]
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]

The first conclusion that arises is that the addition of relations from a mapped resource introduces noise: by adding all the relations from PSC ("IWN + PSC") we obtain slightly worse results than IWN although the number of edges is substantially higher (77,580 vs. 64,258). The addition of only qualia relations ("IWN + PSC q") also has a negative impact, but it is worth mentioning that the scores obtained in this case are higher than those obtained when adding all the relations, even if the amount of edges is smaller (76,716 vs. 77,580). Adding one type of qualia relation has positive effects for the formal ("IWN + PSC f") and specially for the telic relation ("IWN + PSC t"). Agentive ("IWN + PSC a") and constitutive ("IWN + PSC c") relations seem to have a negative impact though. However, we can not draw any definitive conclusion because all the confidence intervals overlap, and therefore the differences might not be significant.

Due to the fact that telic and formal relations improve the results, and given that the overwhelming majority of wordsenses in both IWN and PSC are already connected by formal relations[2], while there are not so many telic relations, we hypothesise that adding more telic relations might provide a significant improvement.

Finally, we compare the results obtained with UKB (configuration "IWN + PSC t") to those of other systems (see table 3), JIGSAW (Basile et al., 2007), the only participant of the WSD task at EVALITA 2007, and a baseline calculated on the basis of the "first-sense-heuristics", i.e. taking as point of reference the functioning of a system which always chooses the first sense available. The motivation for setting up this baseline is that in IWN the first sense of each lemma usually corresponds to the commonest one.

Table 3: Comparison to other systems

system	P	R	conf. interval
UKB	62.0	50.8	[59.9, 64.3]
JIGSAW	56	41.4	[53.8, 58.2]
first-sense	66.9	66.9	[65.1, 68.7]

In this case the confidence intervals do not overlap, which allow us to state that UKB performs significantly better than JIGSAW but worse than the first-sense baseline.

4 Extraction of Telic Relations to Improve WSD

We propose to carry out a pattern-based extraction on the web to increment the number of telic relations in PSC (table 4 shows the types of telic relations present in this LR). It has several phases; on the first we search in the web all the wordsense pairs that instantiate a given telic relation. On the second, the sentences gathered are linguistically analysed (Part-of-speech tagging, chunking, WSD, Named Entity Recognition). Subsequently, these texts are analysed in order to generalise Kybot patterns that express the input relation. Finally, these patterns are applied to a corpus in order to extract further word pairs that instantiate the relation, and these pairs are imported into PSC.

Kybots (Knowledge-yielding robots) (Vossen et al., 2008) are data miners that can be defined using constraints among relations at a generic level. The Kybots are applied here to detect relations in running text by using expression rules, these represent

[2] The is-a formal relation is required as it is the relation that forms the hierarchy

Table 4: Telic relations in PSC

relation	number
Indirect telic	834
Is the ability of	171
Is the activity of	908
Is the habit of	95
Object of the activity	1,367
Purpose	210
Telic	165
Used against	44
Used as	376
Used by	68
Used for	3,293

general morpho-syntactic and semantic conditions on sequences of terms, and relevant pieces of linguistically analysed text. Moreover, these rules can encode semantic conditions expressed by resources connected to WordNet such as Base Concepts, the Top Concept Ontology and others such as the Suggested Merged Ontology (SUMO).

Let us clarify the procedure proposed by presenting examples of wordsense pairs that instantiate relations, sentences that are gathered by searching them in the web, and Kybot patterns that can be generalised for them for the telic relations "Used against":

- (insetticida, insetto) l'insetticida elimina l'insetto istantaneamante *[the insecticide eliminates the insect instantly]*
- (antidolorifico, dolore) l'antidolorifico togliendo il dolore *[the painkiller by removing the pain]*
- (antigelo, congelamento) l'antigelo evita il congelamento dell'acqua *[the antifreeze prevents freezing of water]*
- (spermicida, spermatozoo) gli spermicidi neutralizzano gli spermatozoidi *[spermicides neutralise the sperm]*

```
terms:
(term(@pos="DA*"))?
$t1=term(@pos="N*")
term(@lemma="togliere" | "neutralizzare"
  | "eliminare" | "evitare")
term(@pos="DA*")?
$t2=term(@pos="N*")
fact:
"source_relation" = $t1(@lemma)
"target_relation" = $t2(@lemma)
```

This kybot detects sequences in which an article (optional) is followed by a noun, this by a verb which a given lemma, followed on its turn by another article (optional) and finally a second noun. The kybot would extract a telic relation in which the source is the first noun and the target the second. Appendix A presents the formal XML syntax of this Kybot. We present further examples for the relation "Used for":

- (cifrario, comunicare) un cifrario che gli permetteva di comunicare *[a cipher that allowed him to communicate]*
- (cifrario, comunicare) Giulio Cesare usava un semplice cifrario per comunicare con i suoi generali *[Julius Caesar used a simple cipher to communicate with his generals]*
- (cifrario, comunicare) si usi ancora il cifrario per comunicare *[still uses the cipher to communicate]*
- (thermos, contenere) il thermos permette di contenere cibi caldi per molte ore *[the thermos allows to keep food warm for several hours]*
- (teca, contenere) una teca puo' contenere piu' cose *[a theca can hold more things]*
- (radiolina, ascoltare) una radiolina che mi consentiva di ascoltare la radio *[a radio that allowed me to listen to the radio]*
- (lente, osservare) questa lente gravitazionale ci permette di osservare questo processo *[this gravitational lens allows us to observe this process]*
- (lente, correggere) una lente torica possa correggere questa loro condizione *[toric lenses can correct their condition]*
- (basilico, aromatizzare) il basilico puo' essere usato per aromatizzare l'olio *[basil can be used to flavour the oil]*

```
terms:
$t1=term(@pos="N*")
(term(@pos="P*"))?
term(@lemma="permettere" | "potere"
  | "consentire")
$t2=term(@pos="V*")
fact:
"source_relation" = $t1(@lemma)
"target_relation" = $t2(@lemma)
```

```
terms:
$t1=term(@lemma="usare")
(term(@lemma="ancora"))?
(term(@pos="D*")
term(@pos="A*")
term(@pos="N*")
)*
term(@lemma="per")
$t2=term(@pos="V*")
fact:
"source_relation" = $t1(@lemma)
"target_relation" = $t2(@lemma)
```

5 Related work

This section presents work related to the two research topics tackled in this paper: WSD and relation extraction.

WSD is recognised as a key task of semantic processing within the Natural Language Processing research community. Research in this area has been promoted by the organisation of evaluation tasks during the last decade, Senseval/Semeval[3] campaigns being the most known example. There are also other regional campaigns that focus on the evaluation for single languages, an example is the WSD task at the 2007 edition of the Italian evaluation campaign EVALITA[4].

Pattern-based relation extraction initiates with (Hearst, 1992). Several similar works have been carried out in the last years, including among many others (Cimiano et al., 2004; Banko et al., 2007). There is also one piece of research, (Cimiano and Wenderoth, 2007), that faces the extraction of the specific type of qualia relations. Compared to these proposals, ours presents two novel aspects; on one hand we deal with multi-level patterns (in which elements can pertain to different language levels; elements can include words, lemmas, pos tags, chunks, wordsenses, ontology nodes, entity types, etc.), on the other we tackle the extraction of fine-grained qualia relation types (see table 4).

6 Conclusions

This paper has studied the impact that qualia relations have on WSD. In order to do this, we have set up a graph-based WSD system for Italian by converting IWN to a graph. We have determined the impact of different types of qualia relations by extracting them (one type at a time) from a GL-based lexicon, and evaluating the resulting system over a manually annotated corpus. From the results, it arises that telic and formal relations have a positive impact on the results, while the addition of agentive and constitutive relations worsened the performance of the WSD algorithm.

Furthermore, we have proposed to apply Kybots (multi-level linguistic patterns) in order to extract more telic relations. Future work then will consist on the design of Kybot patterns for several fine-grained types of telic relations, the extraction of such relations and the evaluation of their impact on the WSD algorithm.

Acknowledgements

This work has been partially funded by the EU Commission under the project KYOTO (ICT-2007-211423).

References

Eneko Agirre and Aitor Soroa. 2009. Personalizing pagerank for word sense disambiguation. In *Proceedings of the 12th conference of the European chapter of the Association for Computational Linguistics (EACL-2009)*.

A. Alonge, F. Bertagna, N. Calzolari, and A. Roventini. 1999. The Italian Wordnet, EuroWordNet Deliverable D032D033 part B5. Technical report.

Michele Banko, Michael J. Cafarella, Stephen Soderland, Matthew Broadhead, and Oren Etzioni. 2007. Open information extraction from the web. In Manuela M. Veloso and Manuela M. Veloso, editors, *IJCAI*, pages 2670–2676.

P. Basile, M. Degemmis, A. L. Gentile, P. Lops, and G. Semeraro. 2007. The jigsaw algorithm for word sense disambiguation and semantic indexing of documents. In *AI*IA '07: Proceedings of the 10th Congress of the Italian Association for Artificial Intelligence on AI*IA 2007*, pages 314–325, Berlin, Heidelberg. Springer-Verlag.

S. Brin and L. Page. 1998. The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems*, 30(1-7).

P. Cimiano and J. Wenderoth. 2007. Automatic acquisition of ranked qualia structures from the web. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 888–895.

P. Cimiano, A. Pivk, L. Schmidt-Thieme, and S. Staab. 2004. Learning taxonomic relations from heterogeneous sources.

T. H. Haveliwala. 2002. Topic-sensitive pagerank. In *WWW '02: Proceedings of the 11th international*

[3]http://www.senseval.org/
[4]http://evalita.fbk.eu/

conference on World Wide Web, pages 517–526, New York, NY, USA. ACM.

Marti A. Hearst. 1992. Automatic acquisition of hyponyms from large text corpora. In *COLING*, pages 539–545.

N. Ruimy, M. Monachini, R. Distante, E. Guazzini, S. Molino, M. Ulivieri, N. Calzolari, and A. Zampolli. 2002. Clips, a multi-level italian computational lexicon: A glimpse to data. In *Proceedings of the Third International Conference on Language Resources and Evaluation (LREC'02)*, Las Palmas de Gran Canaria, Spain.

N. Ruimy, A. Roventini, R. Marinelli, and M. Ulivieri. 2008. Linking and integrating two electronic lexicons. In N. Ide and A.C. Fang, editors, *Proceedings of the First International Conference on Global Interoperability for Language Resources (ICGL 2008)*, Hong Kong.

Marisa Ulivieri, Elisabetta Guazzini, Francesca Bertagna, and Nicoletta Calzolari. 2004. Senseval-3: The italian all-words task. In Rada Mihalcea and Phil Edmonds, editors, *Senseval-3: Third International Workshop on the Evaluation of Systems for the Semantic Analysis of Text*, Barcelona, Spain, July. Association for Computational Linguistics.

Piek Vossen, Eneko Agirre, Nicoletta Calzolari, Christiane Fellbaum, Shu kai Hsieh, Chu-Ren Huang, Hitoshi Isahara, Kyoko Kanzaki, Andrea Marchetti, Monica Monachini, Federico Neri, Remo Raffaelli, German Rigau, Maurizio Tescon, and Joop VanGent. 2008. Kyoto: a system for mining, structuring and distributing knowledge across languages and cultures. In Nicoletta Calzolari, Khalid Choukri, Bente Maegaard, Joseph Mariani, Jan Odjik, Stelios Piperidis, and Daniel Tapias, editors, *Proceedings of the Sixth International Language Resources and Evaluation (LREC'08)*, Marrakech, Morocco, may. European Language Resources Association (ELRA). http://www.lrec-conf.org/proceedings/lrec2008/.

A Kybot XML samples

```
<?xml version="1.0">
<Kybot id="used-against">
 <variables>
  <var name="X" type="term" pos="DA*"/>
  <var name="Y" type="term" pos="N*"/>
  <var name="Z" type="term"
       lemma="togliere | neutralizzare
             | eliminare | evitare"/>
  <var name="V" type="term" pos="DA*"/>
  <var name="W" type="term" pos="N*"/>
</variables>

 <relations>
  <root span="X"/>
  <rel span="Y" pivot="X"
       direction="following" dist = "1"/>
  <rel span="Z" pivot="Y"
       direction="following" dist = "1"/>
  <rel span="V" pivot="Z"
       direction="following" dist = "1"/>
  <rel span="W" pivot="V"
       direction="following" dist = "1"/>
 </relations>

 <output>
  <fact id='used-against-fact'>
    <source value='$Y/@lemma'/>
    <target value='$W/@lemma'/>
  </fact>
 </output>
</Kybot>
```