

The MEANING approach for automatic acquisition of sense examples.

Montse Cuadros, Jordi Atserias, Mauro Castillo

TALP Research Center, UPC
Jordi Girona, 1-3 Edifici Omega 305
08034 Barcelona
Spain,
{cuadros,batalla,castillo@lsi.upc.edu

German Rigau

IXA Group, EHU
Euskal Herriko Unibersitatea
08034 Donostia
Spain,
{rigau}@si.ehu.edu

1 ExRetriever

ExRetriever characterises each sense of a word as a specific query. This is automatically done by using a particular query construction strategy, which is defined *a priori* by an expert. Each different strategy can take into account the information related to words and available into a lexical knowledge base in order to automatically generate the set of queries.

The current version of ExRetriever is able to use different lexical databases (through the MCR) and different corpora (SemCor, BNC, the Web, etc.) through a common API.

In order to easily implement different query construction strategies, ExRetriever has been powered with a declarative language. This language allows the manual definition of complex query construction strategies and it is briefly described in the following section.

2 Experiments

Within the framework of the MEANING project, both direct and indirect evaluation experiments of the ExRetriever performance have been designed.

Using ExRetriever on SemCor we can perform detailed micro-analysis on the data available. That is, we can easily perform many adjustments for building queries and filtering appropriately those unwanted examples, balancing the trade-off between coverage and precision.

Each one of such experiments consists of applying a particular query construction strategy to a set of 73 English words from Senseval-2 *lexical sample* task. The resulting specific queries (one for each sense word) automatically generated by applying each strategy have been tested against Semcor. Due to the small size of Semcor, specific queries are likely to produce poor recall. However, Semcor is the unique sense tagged resource providing large quantities of examples for all-words.

Six different query construction strategies have been tested, some of them inspired in those used in the literature, Lea(Leacock *et al.* 98), Mol1-3 (Mihalcea & Moldovan 99), Mea1-2 (Fernández *et al.* 04).

3 Results

Q	P	R	F1	NSen	WSC*
Lea1	98,84	3,66	7,06	23254	23
Mol1	99,35	4,72	9,01	3241	10
Mol3	8,12	26,11	12,39	7611	47
Mea1	9,28	24,38	13,44	9490	54
Mea2	10,69	26,28	15,20	17171	58

Table 1: Overall figures

Table 1 shows the overall figures for each query when applied to the total 73 words of the test set. WSC* (Cuadros *et al.* 04) is a measure of the coverage of the senses of the word.

References

- (Cuadros *et al.* 04) M. Cuadros, M. Castillo, G. Rigau, and J. Atserias. Automatic Acquisition of Sense Examples using ExRetriever. In *Iberamia'04*, pages 97–104, 2004.
- (Fernández *et al.* 04) J. Fernández, M. Castillo, G. Rigau, J. Atserias, and J. Turmo. Automatic Acquisition of Sense Examples using ExRetriever. In *LREC'04*, pages 25–28, 2004.
- (Leacock *et al.* 98) C. Leacock, M. Chodorow, and G. Miller. Using Corpus Statistics and WordNet Relations for Sense Identification. *Computational Linguistics*, 24(1):147–166, 1998.
- (Mihalcea & Moldovan 99) R. Mihalcea and I. Moldovan. An Automatic Method for Generating Sense Tagged Corpora. In *Proceedings of the 16th National Conference on Artificial Intelligence*. AAAI Press, 1999.