Robust – Word Sense Disambiguation exercise

UBC: **Eneko Agirre**, Oier Lopez de Lacalle, Arantxa Otegi, German Rigau

UVA & Irion: Piek Vossen

UH: Thomas Mandl

Introduction

- Robust: emphasize difficult topics using non-linear combination of topic results (GMAP)
- This year also automatic word sense annotation:
  - English documents and topics (English WordNet)
  - Spanish topics (Spanish WordNet - closely linked to the English WordNet)
- Participants explore how the word senses (plus the semantic information in wordnets) can be used in IR and CLIR
- See also QA-WSD exercise, which uses same set of documents

Documents

- News collection: LA Times 94, Glasgow Herald 95
- Sense information added to all content words
  - Lemma
  - Part of speech
  - Weight of each sense in WordNet 1.6
- XML with DTD provided
- Two leading WSD systems:
  - National University of Singapore
  - University of the Basque Country
- Significant effort (100Mword corpus)
  - Special thanks to Hwee Tou Ng and colleagues from NUS and Oier Lopez de Lacalle from UBC

Documents: example XML

```
<DOC>
    <DOCNO>GH950102-000000</DOCNO>
    <DOCID>GH950102-000000</DOCID>

    <HEADLINE>
        <TERM ID="GH950102-000000-1" LEMA="alien" POS="JJ">
            <WF>Alien</WF>
            <SYNSET SCORE="0.6" CODE="01295935-a"/>
            <SYNSET SCORE="0.4" CODE="00984080-a"/>
        </TERM>

        <TERM ID="GH950102-000000-2" LEMA="treatment" POS="NN">
            <WF>treatment</WF>
            <SYNSET SCORE="0.82790418008605" CODE="00735486-n"/>
            <SYNSET SCORE="0" CODE="03857483-n"/>
            <SYNSET SCORE="0.172095881991395" CODE="00430183-n"/>
            <SYNSET SCORE="0" CODE="05340429-n"/>
        </TERM>
```

Topics

We used existing CLEF topics in English and Spanish:

- 2001; 41-90; LA 94
- 2002; 91-140; LA 94
- 2004; 201-250; GH 95
- 2003; 141-200; LA 94, GH 95
- 2005; 251-300; LA 94, GH 95
- 2006; 301-350; LA 94, GH 95

First three as training (plus relevance judg.)

Last three for testing

Topics: WSD

- English topics were disambiguated by both NUS and UBC systems
- Spanish topics: no large-scale WSD system available, so we used the first-sense heuristic
  - Word sense codes are shared between Spanish and English wordnets
- Sense information added to all content words
  - Lemma
  - Part of speech
  - Weight of each sense in WordNet 1.6
- XML with DTD provided

Topics: WSD example

```
<top>
    <num>10.2452/141-WSD-AH</num>

    <EN-title>
        <TERM ID="10.2452/141-WSD-AH-1" LEMA="letter" POS="NNP">
            <WF>Letter</WF>
            <SYNSET SCORE="0" CODE="05115901-n"/>
            <SYNSET SCORE="0" CODE="05362432-n"/>
            <SYNSET SCORE="0" CODE="05029514-n"/>
            <SYNSET SCORE="1" CODE="04968965-n"/>
        </TERM>

        <TERM ID="10.2452/141-WSD-AH-2" LEMA="bomb" POS="NNP">
            <WF>Bomb</WF>
            <SYNSET SCORE="0.888888888888889" CODE="02310834-n"/>
            <SYNSET SCORE="0" CODE="05484679-n"/>
            <SYNSET SCORE="0.111111111111111" CODE="02311368-n"/>
        </TERM>
```

Evaluation

- Reused relevance assessments from previous years
- Relevance assessment for training topics were provided alongside the training topics
- MAP and GMAP
- Participants had to send at least one run which did not use WSD and one run which used WSD

Participation

- 8 official participants, plus two late ones:
  - Martínez et al. (Univ. of Jaen)
  - Navarro et al. (Univ. of Alicante)
- 45 monolingual runs
- 18 bilingual runs

	Organization	Country	Mono	Bilingual
geneva	University of Geneva	Switzerland	x	x
inaoe	INAOE	Mexico	x	
ixa	Univ. Basque Country	Spain	x	x
know-center	Knowledge Relationship Discovery	Austria	x	
ucm	Univ. Complutense de Madrid	Spain	x	
ufrgs	Univ. Fed. do Rio Grande do Sul	Brazil	x	x
uniba	Universita' di Bari	Italy	x	x
unine	U.Neuchatel-Informatics	Switzerland	x	
	Univ. of Alicante	Spain	x	
	Univ. of Jaen	Spain	x	

Monolingual results

	Rank	Participant	MAP	GMAP	MAP	GMAP
	1st	unine	**45.14**	21.17		
	2nd	geneva	39.17	16.53		
Non-WSD	3rd	ucm	38.34	15.28		
	4th	ixa	38.10	15.72		
	5th	ufrgs	33.94	13.96		
	1st	unine	44.98	**21.54**	*-0.16*	**0.37**
	2nd	ucm	39.57	16.17	**1.23**	**0.89**
WSD	3rd	ixa	38.99	15.52	**0.89**	*-0.20*
	4th	geneva	38.13	16.25	*-1.04*	*-0.28*
	5th	ufrgs	34.64	14.17	**0.70**	**0.21**

- MAP: non-WSD best, 3 participants improve it
- GMAP: WSD best, 3 participants improve it

Monolingual: using WSD

- UNINE: synset indexes, combine with results from other indexes
  - Improvement in GMAP
- UCM: query expansion using structured queries
  - Improvement in MAP and GMAP
- IXA: expand to all synonyms of all senses in topics, best sense in documents
  - Improvement in MAP
- GENEVA: synset indexes, expanding to synonyms and hypernyms
  - No improvement, except for some topics
- UFRGS: only use lemmas (plus multiwords)
  - Improvement in MAP and GMAP

Monolingual: using WSD

- UNIBA: combine synset indexes (best sense)
  - Improvements in MAP
- Univ. of Alicante: expand to all synonyms of best sense
  - Improvement on train / decrease on test
- Univ. of Jaen: combine synset indexes (best sense)
  - No improvement, except for some topics

Bilingual results

	Rank	Participant	MAP	GMAP	MAP	GMAP
	1st	ufrgs	**36.38**	**13.00**		
Non-WSD	2nd	geneva	30.36	10.96		
	3rd	ixa	19.57	1.62		
	4th	uniba	2.56	0.04		
	1st	ixa	23.56	1.71	**3.99**	**0.09**
WSD	2nd	ufrgs	21.77	5.14	*-14.61*	*-7.86*
	3rd	geneva	9.70	0.37	*-20.66*	*-10.59*
	4th	uniba	7.23	0.16	**4.67**	**0.12**

- MAP and GMAP: best results for non-WSD
- Only IXA and UNIBA improve using WSD, but very low GMAP.

Bilingual: using WSD

- IXA: wordnets as the sole sources for translation
  - Improvement in MAP
- UNIGE: translation of topic for baseline
  - No improvement
- UFRGS: association rules from parallel corpora, plus use of lemmas (no WSD)
  - No improvement
- UNIBA: wordnets as the sole sources for translation
  - Improvement in both MAP and GMAP

Conclusions and future

- Novel dataset with WSD of documents
- Successful participation
  - 8+2 participants
  - Some positive results with top scoring systems
  - Room for improvement and for new techniques
- Analysis
  - Correlation with polysemy and difficult topics underway
  - Manual analysis of topics which get improvement with WSD
- New proposal for 2009

Robust – Word Sense Disambiguation exercise

Thank you!