KnowNet: Building a Large Net of Knowledge from the Web

Montse Cuadros

cuadros@lsi.upc.edu
Universitat Politècnica de Catalunya
NLP Group - TALP

Seminari UPC
23 de gener de 2009

Outline

Preliminaries I

- Large-scale Semantic Processing dealing with concepts (senses) rather than words
- Two complementary and interdependent problems:
  - Acquisition bottleneck
    - Autonomous large-scale knowledge acquisition systems
  - Ambiguity
    - Highly accurate and robust semantic systems

Preliminaries II

Acquire Large-scale Semantic Resources for NLP

- Manually-> Hard and expensive task.
    - Small coverage.
    - Dozens of persons developing those resources.
    - for example, WordNet 250.000 relations.
- Automatically-> Difficult but cheap.
    - Wide coverage.
    - Major Challenge of NLP.
    - for example, Multilingual Central Repository [Atserias et al., 2004] 1.600.000 relations

Motivation: Improve the Knowledge Bases I

- Given the noun party, which has 5 senses in WordNet 1.6, the first 2 senses:
  - *Sense 1*:<party, political party> (an organization to gain political power;)
  - Sense 2: <party> (an occasion on which people can assemble for social interaction and entertainment;)
- The number of semantic relations existing actually:

	sense 1	sense 2
Manually	29	7
spSemCor	20	7
spBNC	226	181
XWN	111	15
MCR	386	210

Motivation: Improve the Knowledge Bases II

Our main goal is:

- Increase the connectivity between the word-senses of the Knowledge Resources.
  - Acquire Large-scale Topic Signatures (TS [Agirre et al., 2000]) for each WordNet sense.
    - For each noun-sense:

party#n#2 an occasion on which people can assemble for social interaction and entertainment;

- We construct a query:

party#n#2: (house_party or record_hop or bunfight or photo_op or barn_dance or cocktail_party or ceremonial_occasion or photo_opportunity or tea_party or birthday_party or ceilidh)

Motivation: Improve the Knowledge Bases III

- Find a set of examples in a corpus

2 –> And it was, I mean, right, things like House **Party** and all that lot white man Malcolm X, they 're all black films.
2 –> These two ladies decided to form a social club, and organised a tea **party** which took place at the Ladies Park Club, where the Spurs Club was launched.
2 –> He was dissatisfied, as always, with his previous work, and he had detected flaws in The Cocktail **Party** which he wished to remove from the new play.

Motivation: Improve the Knowledge Bases IV

- Calculate the set of related words to a topic using an specific metric or Topic Signatures:

birthday 1.8875; party 1.8677; tea 0.9478; cocktail 0.9162; house 0.4577; dance 0.4030; ceilidh 0.3996; barn 0.3797; photo 0.3067; teaparty 0.2759; night 0.2732;

- Disambiguate establish relation between synset-word

party#n#2	cocktail#n	
	terrace#n	? 1
	baby#n	
	flagstaff#n	
	potter#n	
	birth#n	

Motivation: Improve the Knowledge Bases V

- Characterize discovered relations
  Which is the relation between party#n#2 and terrace#n#1??
  Using ontological properties appearing in MCR [Alvez et al.,2008] [1].
  - *the relation is* **LOCATION***.*

[1] http://adimen.si.ehu.es/drupal/WordNet2TCO

Topic Signatures

Topic Signatures(TS):

- Word vectors related to a particular topic (synset)
- Built by retrieving context words of a target word

Utilization

- Summarization Tasks [Lin & Hovy, 2000]
- Ontology population [Alfonseca et al., 2004]
- Word sense disambiguation [Agirre et al., 2000], [Agirre et al., 2001b]
- Knowledge Adquisition

There are publicly available Topic Signatures for all WordNet 1.6 [Fellbaum 98] nominal senses
[Agirre & Lopez de Lacalle, 2004]
http://ixa.si.ehu.es/Ixa/resources/sensecorpus.

Example of TS

Related to the noun church, sense 1.

service	0.776187	anglican	0.651298
church	0.776186	services	0.651127
clergy	0.718070	tower	0.651071
hymns	0.695500	st	0.650787
peterś	0.695215	congregational	0.648595
episcopal	0.689341	congregation	0.647037
presbyterian	0.685548	priest	0.644656
cathedral	0.685220	memorial	0.644652
churches	0.683878	charters	0.642540
royal	0.673297	worship	0.637472
parish	0.671534	bishop	0.634107
pastoral	0.670789	volunteer	0.629541
maryś	0.666601	...	

Acquisition of TS

Methods to automatically acquire Large-scale Knowledge Resources (Topic Signatures), using several parameters:

- Corpus: British Nacional Corpus (BNC), SEMCOR vs the WEB.
- Measures: TFIDF, MI and AR.
- Knowledge Resources: WordNet.
- Query Construction: QueryW vs QueryA.
    - **QueryA:** Union set of synonyms, hyponyms and hyperonyms.
    - **QueryW:** Union set of synonyms, direct hyponyms, hypernyms, indirect hyponyms (distance 2 and 3) and siblings. Words are Monosemous Total.

Systems I

TSWEB: Topic Signatures [Agirre & Lopez de Lacalle, 2004].

- **size:** all nominal senses of WordNet 1.6
- **corpus:** WEB (1000 snipets per concept)
- **measure:** TFIDF
- **query:** queryW

EXRET: Topic Signatures using ExRetriever [Cuadros et al., 2004].

- **size:** Senseval-3 20 nouns senses
- **corpus:** BNC (100 milion words)
- **measure:** TFIDF, MI, AR
- **query:** queryW and queryA

Systems II

TSSEM: Topic Signatures using SemCor [Cuadros et al., 2007].

- **size:** all SemCor tagged words by pos, lemma and sense (WN 1.6)
- **corpus:** SEMCOR (192,639)
- **measure:** TFIDF
- **query:** Building a set of words-senses with an specified weight for each synset with TFIDF using all word-sense coocuring in sentences with contain an specific synset.

Systems III

Example of the first word-senses we calculate from party#n#1 from TSSEM (SemCor)

political_party#n#1	2.3219
party#n#1	2.3219
election#n#1	1.0926
nominee#n#1	0.4780
candidate#n#1	0.4780
campaigner#n#1	0.4780

What's KnowNet?[2]

- Extensible, large and accurate knowledge base.
- Derived by semantically disambiguating the TS acquired from the web [Agirre & Lopez de Lacalle, 2004].
  - Knowledge-based WSD algorithm (SSI) to assign the most apropiate sense.
  - Knowlege Base (WN+XWN) [Fellbaum 98], [Mihalcea & Moldovan, 2001]

[2] http://adimen.si.ehu.es

SSI-Dijkstra description

Based on Structural Semantic Interconnections algorithm (SSI) [Navigli & Velardi, 2005].

- A knowledge-based iterative approach to Word Sense Disambiguation.
  - Initialization step (All monosemous are included in I (interpreted words)), polysemous in P (pending words))
  - Iterative loop, while no more pending words...
    - Selects a word W from P
    - Selects the sense S of W closer to I (disambiguate W using I as context)
    - Adds S to I and removes W from P

SSI-Dijkstra graph

- Proximity of one synset to the rest of synsets of I:
  - Build a very large connected graph with 99,635 nodes (synsets) and 636,077 edges. (Boost C++ Library[3])
  - Containing direct relations between synsets gathered from WordNet and eXtended WordNet.
  - Compute Dijkstra algorithm. The Dijkstra algorithm is a greedy algorithm for computing the shortest path distance between one node and the rest of nodes of a graph.

[3] http://www.boost.org

SSI-Dijkstra vs. original SSI

- SSI-Dijkstra uses the available knowledge from WN and XWN.
- SSI-Dijkstra always provides the minimum distance between a synset and the set of already interpreted words.
- Original SSI uses an in-house knowledge base.
- Original SSI not always provide a path distance because it depends on a grammar.

Example of Use (party#n#1)

Query

party#n alinement#n federalist#n whig#n greenback#n nazi#a republican#n constitutional#n

Example of Use (party#n#1)

Query

party#n alinement#n federalist#n whig#n greenback#n nazi#a republican#n constitutional#n

Monosemous

I

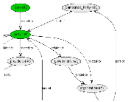
party#n#1
tammany#n#1
alinement#n#1
greenback#n#1
constitutional#n#1

Polisemous

P

federalist#n (2)
whig#n (3)
nazi#a (2)
republican#n (3)

WN+XWN



Dijkstra Distance

party#n#1
tammany#n#1
alinement#n#1
greenback#n#1
constitutional#n#1

federalist#n#1

Example of Use (party#n#1)

Query

party#n alinement#n federalist#n whig#n greenback#n nazi#a republican#n constitutional#n

Monosemous

I

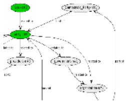
- party#n#1
- tammany#n#1
- alinement#n#1
- greenback#n#1
- constitutional#n#1

Polisemous

P

- federalist#n (2)
- whig#n (3)
- nazi#a (2)
- republican#n (3)

WN+XWN



Dijkstra Distance

party#n#1	d_1	federalist#n#1
tammany#n#1	d_2	
alinement#n#1	d_3	
greenback#n#1	d_4	
constitutional#n#1	d_5	

federalist#n#1

a member of a former political party in the United States that favored a strong centralized federal government

Example of Use (party#n#1)

Query

party#n alinement#n federalist#n whig#n greenback#n nazi#a republican#n constitutional#n

Monosemous

I

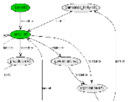
party#n#1
tammany#n#1
alinement#n#1
greenback#n#1
constitutional#n#1

Polisemous

P

federalist#n (2)
whig#n (3)
nazi#a (2)
republican#n (3)

WN+XWN



Dijkstra Distance

party#n#1
tammany#n#1
alinement#n#1
greenback#n#1
constitutional#n#1

d_1 d_2 d_3 d_4 d_5

federalist#n#2

an advocate of federalism

Example of Use (party#n#1)

Query

party#n alinement#n federalist#n whig#n greenback#n nazi#a republican#n constitutional#n

Monosemous	Polisemous
I	P
party#n#1	federalist#n (2)
tammany#n#1	whig#n (3)
alinement#n#1	nazi#a (2)
greenback#n#1	republican#n (3)
constitutional#n#1	

federalist#n#1 ?? federalist#n#2

Example of Use (party#n#1)

Query

party#n alinement#n federalist#n whig#n greenback#n nazi#a republican#n constitutional#n

Monosemous

I

party#n#1
tammany#n#1
alinement#n#1
greenback#n#1
constitutional#n#1

Polisemous

P

federalist#n (2)
whig#n (3)
nazi#a (2)
republican#n (3)

federalist#n#1 ?? federalist#n#2

Example of Use (party#n#1)

Query

party#n alinement#n federalist#n whig#n greenback#n nazi#a republican#n constitutional#n

Monosemous

I

party#n#1
tammany#n#1
alinement#n#1
greenback#n#1
constitutional#n#1
federalist#n#1

Polisemous

P

whig#n (3)
nazi#a (2)
republican#n (3)

Example of Use (party#n#1)

Query

party#n alinement#n federalist#n whig#n greenback#n nazi#a republican#n constitutional#n

Final Disambiguation

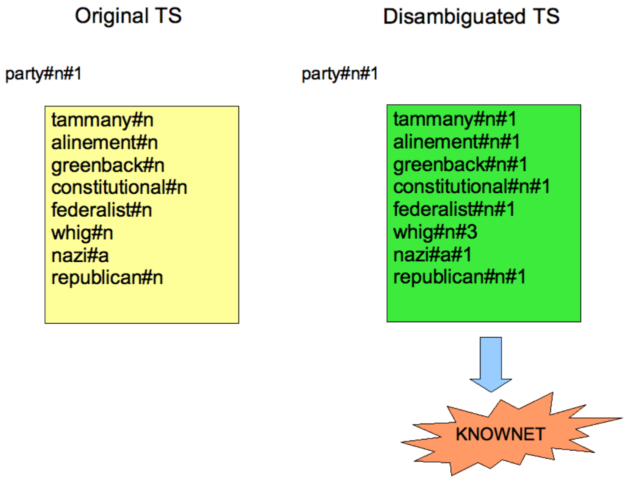
I

party#n#1
tammany#n#1
alinement#n#1
greenback#n#1
constitutional#n#1
federalist#n#1
whig#n#3
nazi#a#1
republican#n#1

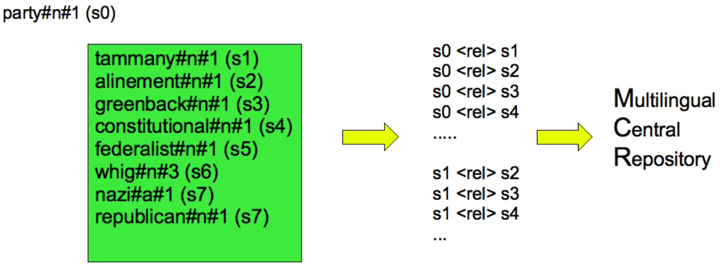
KnowNet Construction

Original TS	Disambiguated TS
party#n#1	party#n#1
tammany#n	tammany#n#1
alinement#n	alinement#n#1
greenback#n	greenback#n#1
constitutional#n	constitutional#n#1
federalist#n	federalist#n#1
whig#n	whig#n#3
nazi#a	nazi#a#1
republican#n	republican#n#1

KnowNet Construction



KnowNet Construction



Few Examples I

00003095-n 04197156-n

00003095-n cell#n#2

the basic structural and functional unit of all organisms; cells may exist as independent units of life (as in monads) or may form colonies or tissues as in higher plants and animals

04197156-n blood#n#1

the fluid (red in vertebrates) that is pumped by the heart

Few Examples II

00003095-n 10785027-n

00003095-n cell#n#2

the basic structural and functional unit of all organisms; cells may exist as independent units of life (as in monads) or may form colonies or tissues as in higher plants and animals

10785027-n antibody#n#1

any of a large variety of immunoglobulins normally present in the body or produced in response to an antigen which it neutralizes, thus producing an immune response

Few Examples III

00013700-n 04955371-n

00013700-n motivation#n#1 motive#n#1 need#n#3

the psychological feature that arouses an organism to action; the reason for the action

04955371-n lesson#n#3 moral#n#1

the significance of a story or event

Few Examples IV

00013700-n 05085310-n

00013700-n motivation#n#1 motive#n#1 need#n#3

the psychological feature that arouses an organism to action; the reason for the action

05085310-n belief#n#2 dogma#n#1 tenet#n#1

a religious doctrine that is proclaimed as true without proof

Few Examples V

00014045-n 05349662-n

00014045-n feeling#n#1

the psychological feature of experiencing affective and emotional states

05349662-n expression#n#4 locution#n#1 saying#n#1

a word or phrase that particular people use in particular situations

Few Examples VI

00014045-n 06661163-n

00014045-n feeling#n#1

the psychological feature of experiencing affective and emotional states

06661163-n impulse#n#1 urge#n#1

an instinctive motive

Few Examples VII

00262937-n 00870386-v

00262937-n operation#n#4

a planned activity involving many people performing various actions

00870386-v puncture#v#1

pierce with a pointed object; make a hole into

Overlapping and Size

Setting a window of 5, 10, 15 since 20 words of the Topic Signature

KB	WN+XWN	#relations	#synsets
KN-5	3,1%	231163	39,864
KN-10	5,0%	689610	45,817
KN-15	6,9%	1378286	48,521
KN-20	8,5%	2358927	50,789

Table: Size and percentage of overlapping relations between KnowNet versions and WN+XWN

Comparing number of relations

Source	**#relations**
Princeton WN3.0	235,402
Selectional Preferences from SemCor	203,546
eXtended WN	550,922
Co-occurring relations from SemCor	932,008
New KnowNet-5	231,163
New KnowNet-10	689,610
New KnowNet-15	1,378,286
New KnowNet-20	2,358,927

Table: Number of synset relations

Evaluating Knowledge Resources

Main Goals:

- Establishment of a relative quality of KnowNet in a neutral environment.
- Compare the performance of KnowNet with other Knowledge Resources.

Framework: Senseval-3/SemEval-07

- Senseval covers a wide number of tasks and languages.
- We focus our Evaluation on a WSD task, the Lexical Sample Task for English and Spanish:
  - Determines the sense of a tagged word in a sentence
  - Uses the context words of the sentence for the Disambiguation Algorithms.
- We expect better resources will obtain higher accuracy figures on the task

Evaluation Procedure

- For each noun-test example, between the Test and all the Topic Signatures related to the possible senses of the noun:
  - A simple word overlapping (by occurrence or weight) is defined for each sense example as:
    - **occurrence** evaluation measure: counts the words occurring from the TS.

The synset having higher overlapping for a chosen method is selected for a particular test example.

Ev. Process: Test Example of Party sense 1 in Senseval-3

<instance id="party.n.bnc.00008131" docsrc="BNC"> <context> Up to the **late** 1960s , catholic nationalists were **split** between two **main political** groupings . There was the Nationalist Party , a weak **organization** for which **local** priests had to **provide** some **kind** of legitimation . As a <head>party</head>, it really only exercised a modicum of **power** in relation to the Stormont administration . Then there were the republican **parties** who focused their attention on Westminster **elections** . The disorganized **nature** of catholic nationalist **politics** was only turned round with the emergence of the **civil rights movement** of 1968 and the subsequent forming of the SDLP in 1970. </context> </instance>

Studied Resources I

Wide range of large-scale Knowledge Resources integrated into **Multilingual Central Repository** (MCR):

- WordNet (WN) [Fellbaum 98]
  - tree#n#1–hyponym–>teak#n#2
- eXtended WordNet [Mihalcea & Moldovan, 2001]
  - teak#n#2–gloss–>wood#n#1
- large collections of semantic preferences:
  - Acquired from SemCor [Agirre & Martinez, 2001, Agirre & Martinez, 2002].
  - Acquired from the BNC [McCarthy, 2001]
    - read#v#1–tobj–>book#n#1

Studied Resources II

Large-Scale Automatically Acquired **Topic Signatures**:

- TSWEB Acquired from the web [Agirre & Lopez de Lacalle, 2004].
  - querying google for each wordNet 1.6 synset.
  - building a set of words with an specified weight for each synset with TFIDF using all the snippets obtained.
- TSSEM Acquired from Semcor [Cuadros et al., 2007].
  - constructed using all Semcor tagged words by pos, lemma and sense (WN 1.6)
  - Building a set of words-senses with an specified weight for each synset with TFIDF using all word-sense coocuring in sentences with contain an specific synset.

English Baselines

- **RANDOM**: A random sense **lower-bound**.
- **WordNet MFS (WN-MFS)**: The first sense in WordNet.
- **TRAIN-MFS**: The most frequent sense in the training corpus.
- **Train Topic Signatures (TRAIN)**: Using the training corpus to directly build a Topic Signature **upper-bound** of our evaluation framework.

KnowNet Results: Senseval-3 in English

KB	P	R	F1	Av. Size
TRAIN	*65.2*	*65.1*	*65.1*	450
TRAIN-MFS	*54.5*	*54.5*	*54.5*	-
WN-MFS	*53.0*	*53.0*	*53.0*	-
TSSEM	52.5	52.4	52.4	103
SEMCOR-MFS	*49.0*	*49.1*	*49.0*	
MCR^2	45.1	45.1	45.1	26,429
WN+XWN+KN-20	44.8	44.8	44.8	671
MCR	45.3	43.7	44.5	129
KnowNet-20	44.1	44.1	44.1	610
KnowNet-15	43.9	43.9	43.9	339
spSemCor	43.1	38.7	40.8	56
KnowNet-10	40.1	40.0	40.0	154
$(\mathrm{WN+XWN})^2$	38.5	38.0	38.3	5,730
WN+XWN	40.0	34.2	36.8	74
TSWEB	36.1	35.9	36.0	1,721
XWN	38.8	32.5	35.4	69
KnowNet-5	35.0	35.0	35.0	44
WN^3	35.0	34.7	34.8	503
WN^4	33.2	33.1	33.2	2,346
WN^2	33.1	27.5	30.0	105
spBNC	36.3	25.4	29.9	128
WN	44.9	18.4	26.1	14
RANDOM	*19.1*	*19.1*	*19.1*	-

Table: P, R and F1 fine-grained results for the resources evaluated at Senseval-3, English Lexical Sample Task.

KnowNet Results: SemEval-07

KB	P	R	F1	Av. Size
TRAIN	*87.6*	*87.6*	*87.6*	450
TRAIN-MFS	*81.2*	*79.6*	*80.4*	
WN-MFS	*66.2*	*59.9*	*62.9*	
WN+XWN+KN-20	53.0	53.0	53.0	627
$(WN+XWN)^2$	54.9	51.1	52.9	5,153
TSWEB	54.8	47.8	51.0	700
KnowNet-20	49.5	46.1	47.7	561
KnowNet-15	47.0	43.5	45.2	308
XWN	50.1	39.8	44.4	96
KnowNet-10	44.0	39.8	41.8	139
WN+XWN	45.4	36.8	40.7	101
SEMCOR-MFS	*42.4*	*38.4*	*40.3*	
MCR	40.2	35.5	37.7	149
TSSEM	35.1	32.7	33.9	428
KnowNet-5	35.5	26.5	30.3	41
MCR^2	32.4	29.5	30.9	24,896
WN^3	29.3	26.3	27.7	584
RANDOM	*27.4*	*27.4*	*27.4*	
WN^2	25.9	27.4	26.6	72
spSemCor	31.4	23.0	26.5	51
WN^4	26.1	23.9	24.9	2,710
WN	36.8	16.1	22.4	13
spBNC	24.4	18.1	20.8	290

Table: P, R and F1 fine-grained results for the resources evaluated at SemEval-2007, English Lexical Sample Task.

Conclusions for English

- Regarding the baselines, all knowledge resources surpass RANDOM, but none achieves neither WN-MFS, TRAIN-MFS nor TRAIN.
- Different versions of KnowNet seem to be more robust and stable across corpora changes.
- The knowledge they contain outperform any other resource when is empirically evaluated in Senseval-3 and SemEval-07.
- KN-20 $>$ KN-15 $>$ KN-10 $>$ KN-5

Spanish Baselines

- **RANDOM**: A random sense **lower-bound**.
- **Minidir MFS (Minidir-MFS)**: Most Frequent Sense in MiniDir (Minidir-MFS is equal to TRAIN-MFS).
- **Train Topic Signatures (TRAIN)**: Using the training corpus to directly build a Topic Signature **upper-bound** of our evaluation framework.

KnowNet Results: Senseval-3 in Spanish

KB	P	R	F1	Av. S
TRAIN	*81.8*	*68.0*	*74.3*	*450*
MiniDir-MFS	*67.1*	*52.7*	*59.2*	
KnowNet-15	54.7	48.9	**51.6**	176
KnowNet-20	51.8	**49.6**	50.7	319
KnowNet-10	53.5	43.1	47.7	81
MCR	46.1	41.1	43.5	66
WN^2	56.0	29.0	42.5	51
$(\text{WN+XWN})^2$	41.3	41.2	41.3	1,892
KnowNet-5	58.5	26.9	36.8	22
TSSEM	33.6	33.2	33.4	208
XWN	42.6	27.1	33.1	24
WN	**65.5**	13.6	22.5	8
RANDOM	*21.3*	*21.3*	*21.3*	

Table: P, R and F1 fine-grained results for the resources evaluated individually on Spanish.

Conclusions for Spanish

- Regarding the baselines, all knowledge resources surpass RANDOM, but none achieves neither Minidir-MFS (equal to TRAIN-MFS) nor TRAIN.
- MCR (F1 of 43.5) almost doubles WN (F1 of 22.5).
- KnowNet-5 performs better than WN, XWN and the TS acquired from SemCor.
- From KnowNet-10 all KnowNet versions perform better than any other knowledge resource on Spanish derived by manual or automatic means (including the MCR).
- KN-15 $>$ KN-20 $>$ KN-10 $>$ KN-5

Overall Conclusions

- Different versions of KnowNet seem to be more robust and stable across corpora changes and across languages.
- The initial results obtained for the different versions of KnowNet seem to be a major step towards the autonomous acquisition of knowledge from raw corpora.
- They are several times larger than the available knowledge resources which encode relations between synsets.
- The knowledge they contain outperform any other resource when is empirically evaluated in Senseval-3 and SemEval-07.

Future Work

- Label the new relations.
- Create new links and relations for the unknown words/concepts (e.g Gatwick#n#0).
- Evaluate KnowNet in another languages like italian, ...

Moltes gràcies!

Álvez, J., J. Atserias, J. Carrera, S. Climent, A. Oliver, and G. Rigau.
2008.
Consistent annotation of eurowordnet with the top concept ontology.
In *Proceedings of Fourth International WordNet Conference (GWC'08).*

Agirre, E., Ansa, O., Arregi, X., Arriola, J., Díaz de Ilarraza, A., Pociello, E., & Uria, L. (2002).
Methodological issues in the building of the basque wordnet: quantitative and qualitative analysis.
Proceedings of the first International Conference of Global WordNet Association.
Mysore, India.

Agirre, E., Ansa, O., Martinez, D., & Hovy, E. (2000).

Enriching very large ontologies with topic signatures.
Proceedings of ECAI'00 workshop on Ontology Learning.
Berlin, Germany.

- Agirre, E., Ansa, O., Martinez, D., & Hovy, E. (2001b).
  Enriching wordnet concepts with topic signatures.
  Proceedings of the NAACL workshop on WordNet and Other lexical Resources: Applications, Extensions and Customizations.
  Pittsburg.

- Agirre, E., & Lopez de Lacalle, O. (2004).
  Publicity available topic signatures for all wordnet nominal senses.
  LREC'04 (pp. 97–104).

- Agirre, E., & Lopez de Lacalle, O. (2003).
  Clustering wordnet word senses.

International Conference Recent Advances in Natural Language Processing, RANLP'03.
Borovets, Bulgary.

Agirre, E., & Martinez, D. (2000).
Exploring automatic word sense disambiguation with decision lists and the web.
Proceedings of the COLING workshop on Semantic Annotation and Intelligent Annotation.
Luxembourg.

Agirre, E., & Martinez, D. (2001).
Learning class-to-class selectional preferences.
Proceedings of CoNLL01.
Toulouse, France.

Agirre, E., & Martinez, D. (2002).
Integrating selectional preferences in wordnet.

Proceedings of the 1st International Conference of Global WordNet Association.
Mysore, India.

- Alfonseca, E., Agirre, E., & de Lacalle, O. L. (2004).
Approximating hierachy-based similarity for wordnet nominal synsets using topic signatures.
Proceedings of the Second International Global WordNet Conference (GWC'04). Panel on figurative language.
Brno, Czech Republic.
ISBN 80-210-3302-9.

- Alfonseca, E., & Manandhar, S. (2002a).
Distinguishing concepts and instances in wordnet.
Proceedings of the first International Conference of Global WordNet Association.
Mysore, India.

- Alfonseca, E., & Manandhar, S. (2002b).
  An unsupervised method for general named entity recognition and automated concept discovery.
  Proceedings of the first International Conference of Global WordNet Association.
  Mysore, India.

- A.Philpot, E.Hovy, & P.Pantel (2005).
  The omega ontology.
  In IJCNLP Workshop on Ontologies and Lexical Resources, OntoLex-05, 2005.

- Artale, A., Magnini, B., & Strapparava, C. (1997).
  Lexical discrimination with the italian version of wordnet.
  Proceedings of ACL Workshop Automatic Information Extraction and Building of Lexical Semantic Resources.
  Madrid. Spain.

Atserias, J., Climent, S., Farreres, X., Rigau, G., & Rodríguez, H. (1997).
Combining multiple methods for the automatic construction of multilingual wordnets.
proceedings of International Conference on Recent Advances in Natural Language Processing (RANLP'97).
Tzigov Chark, Bulgaria.

Atserias, J., Villarejo, L., Rigau, G., Agirre, E., Carroll, J., Magnini, B., & Vossen, P. (2004).
The meaning multilingual central repository.
Proceedings of the Second International Global WordNet Conference (GWC'04).
Brno, Czech Republic.
ISBN 80-210-3302-9.

Navigli, R. and P. Velardi.

2005.
Structural semantic interconnections: a knowledge-based approach to word sense disambiguation.
IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), 27(7):1063–1074.

Bentivogli, L., Pianta, E., & Girardi, C. (2002).
Multiwordnet: developing an aligned multilingual database.
First International Conference on Global WordNet. Mysore, India.

Buitelaar, P., & Sacaleanu, B. (2001).
Ranking and selecting synsets by domain relevance.
NAACL Workshop 'WordNet and Other Lexical Resources: Applications, Extensions and Customizations' (NAACL'2001)..

Pittsburg, PA, USA.

Buitelaar, P., & Sacaleanu, B. (2002).
Extendings synsets with medical terms.
Proceedings of the 1st Global WordNet Association conference.
Mysore, India.

Castillo, M., Real, F., & Rigau, G. (2003).
Asignacii¿½ automi¿½ica de etiquetas de dominios en wordnet.
Proceedings of SEPLN'03.
Alcala de Henares, Spain.
ISSN 1136-5948.

Church, K. W., & Gale, W. A. (1991).

A comparison of the enhanced good–turing and deleted estimation methods for estimating probabilities of english bigrams.
Computer Speech and Language, *5*, 19–54.

Climent, S., More, Q., Rigau, G., & Atserias, J. (2005).
Towards the shallow ontologization of wordnet. a work in progress.
SEPLN 2005, Granada, Spain.

Cuadros, M., Castillo, M., Rigau, G., & Atserias, J. (2004).
Automatic Acquisition of Sense Examples using ExRetriever.
Iberamia'04 (pp. 97–104).

2005]Cuadros+'05 Cuadros, M., Padro, L., & Rigau, G. (2005).

Comparing methods for automatic acquisition of topic signatures.
RANLP'05, Borovets, Bulgaria, september 2006 (pp. 181–186).

Cuadros, M., Padro, L., & Rigau, G. (2006).
An empirical study for automatic acquisition of topic signatures.
Global Wordnet Conference GWC'06, Jeju Island, Korea, 22-26 January 2006 (pp. 51–59).

Cuadros, M., & Rigau, G. (2006).
Quality assessment of large scale knowledge resources.
To appear in EMNLP'06, Sydney, Austrlia.

Cuadros, M., G. Rigau, and M. Castillo.
2007.

Evaluating large-scale knowledge resources across languages.
In *Proceedings of RANLP*.

- Daudé, J. (2004).
  Mapping large-scale semantic networks.
  Submited to Phd. Thesis, Software Department (LSI).
  Technical University of Catalonia (UPC), Barcelona, Spain.

- 2003]daude03a Daudé, J., Padró, L., & Rigau, G. (2003).
  Making wordnet mappings robust.
  Proceedings of the 19th Congreso de la Sociedad Espaï¿½la para el Procesamiento del Lenguage Natural, SEPLN.
  Universidad Universidad de Alcalï¿½de Henares. Madrid, Spain.

- (Fellbaum 98) C. Fellbaum, editor.

WordNet. An Electronic Lexical Database.
The MIT Press, 1998.

Hamp, B., & Feldweg, H. (1997).
Germanet - a lexical-semantic net for german.
Proceedings of ACL Workshop on Automatic Information Extraction and Building of Lexical Semantic Resources.
Madrid. Spain.

Hearst, M., & Schutze, H. (1993).
Customizing a lexicon to better suit a computational task.
Proceedings of the ACL SIGLEX Workshop on Lexical Acquisition.
Stuttgart, Germany.

Kipper, K., Dang, H. T., & Palmer, M. (2000).
Class-based construction of a verb lexicon.

In Proceedings of the Seventeenth National Conference on Artifical Intelligence, AAAI'00.

- Klapaftis, I. P., & Manandhar, S. (2005).
  Google & wordnet based word sense disambiguation.
  Proceedings of the 22nd International Conference on Machine Learning (ICML05) Workshop on Learning and Extending Ontologies by using Machine Learning Methods.

  Bonn, Germany.

- Leacock, C., Chodorow, M., & Miller, G. (1998).
  Using Corpus Statistics and WordNet Relations for Sense Identification.
  Computational Linguistics, *24*, 147–166.

- Lin, C., & Hovy, E. (2000).

The automated acquisition of topic signatures for text summarization.
Proceedings of 18th International Conference of Computational Linguistics, COLING'00.
Strasbourg, France.

- Lin, D., & Pantel, P. (1994).
  Concept Discovery from Text.
  15th International Conference on Computational Linguistics, COLING'02.
  Taipei, Taiwan.

- L.Vanderwende, G.Kacmarcik, H.Suzuki, & A.Menezes (2005).
  An automatically-created lexical resource.

Proceedings of HLT/EMNLP 2005 Interactive Demonstrations, Vancouver, British Columbia, Canada, October 2005.

- Magnini, B., & Cavaglià, G. (2000).
  Integrating subject field codes into wordnet.
  In Proceedings of the Second Internatgional Conference on Language Resources and Evaluation LREC'2000.
  Athens. Greece.

- McCarthy, D. (2001).
  Lexical acquisition at the syntax-semantics interface: Diathesis aternations, subcategorization frames and selectional preferences.
  Doctoral dissertation, University of Sussex.

- Mihalcea, R., & Moldovan, D. (2001).
  extended wordnet: Progress report.

NAACL Workshop WordNet and Other Lexical Resources: Applications, Extensions and Customizations *(NAACL'2001).* (pp. 95–100).
Pittsburg, PA, USA.

Mihalcea, R., & Moldovan, I. (1999a).
An Automatic Method for Generating Sense Tagged Corpora.
Proceedings of the 16th National Conference on Artificial Intelligence.
AAAI Press.

Mihalcea, R., & Moldovan, I. (1999b).
An automatic method for generating sense tagged corpora.
Proceedings of the 16th National Conference on Artificial Intelligence.

AAAI Press.

Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D., Miller, K., & Tengi, R. (1991).
Five papers on wordnet.
Special Issue of the International Journal of Lexicography, 3, 235–312.

Miller, G. A., Leacock, C., Tengi, R., & Bunker, R. T. (1993).
A semantic concordance.
Proceedings of the ARPA Workshop on Human Language Technology.

Moldovan, D., & Girju, R. (2000).
Domain-specific knowledge acquisition and classification using wordnet.
Proceedings of FLAIRS-2000 conference.

Orlando, Fl.

Montoyo, A., Palomar, M., & Rigau, G. (2001a).
Automatic generation of a coarse grained wordnet.
Proceedings of the NAACL worshop on WordNet and Other Lexical Resources: Applications, Extensions and Customizations.
Pittsburg, USA.

Montoyo, A., Palomar, M., & Rigau, G. (2001b).
Wordnet enrichment with classification systems.
Proceedings of the NAACL worshop on WordNet and Other Lexical Resources: Applications, Extensions and Customizations.
Pittsburg, USA.

Niles, I., & Pease, A. (2001).
Towards a standard upper ontology.

Proceedings of the 2nd International Conference on Formal Ontology in Information Systems (FOIS-2001). Ogunquit, Maine.

Padró, L. (1996).
Pos tagging using relaxation labelling.

Patwardhan, S., & Pedersen, T. (2003).
The cpan wordnet::similarity package (Technical Report).
http://search.cpan.org/author/SID/WordNet-Similarity-0.03/.

Rigau, G., Atserias, J., & Agirre, E. (1997).
Combining unsupervised lexical knowledge methods for word sense disambiguation.
Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics. Joint ACL/EACL (pp. 48–55).

Madrid, Spain.

- Rigau, G., Magnini, B., Agirre, E., Vossen, P., & Carroll, J. (2002).
  Meaning: A roadmap to knowledge technologies.
  Proceedings of COLING Workshop A Roadmap for Computational Linguistics.
  Taipei, Taiwan.

- Rigau, J. D. L. P. G. (2000).
  Mapping wordnets using structural information.
  38th Annual Meeting of the Association for Computational Linguistics (ACL'2000)..
  Hong Kong.

- Stamou, S., Ntoulas, A., Hoppenbrouwers, J., Saiz-Noeda, M., & Christodoulakis, D. (2002a).

Euroterm: Extending the eurowordnet with domain-specific terminology using an expand model approach.
Proceedings of the 1st Global WordNet Association conference.
Mysore, India.

Stamou, S., Oflazer, K., Pala, K., Christoudoulakis, D., Cristea, D., Tufis, D., Koeva, S., Totkov, G., Dutoit, D., & Grigoriadou, M. (2002b).
Balkanet: A multilingual semantic network for the balkan languages.
Proceedings of the 1st Global WordNet Association conference.

Tomuro, N. (1998).
Semi-automatic induction of underspecified semantic classes.

Proceedings of the workshop on Lexical Semantics in Context: Corpus, Inference and Discourse at the 10th European Summer School in Logic, Language and Information (ESSLLI-98).
Saabruecken, Germany.

- Turcato, D., Popowich, F., Toole, J., Fass, D., Nicholson, D., & Tisher, G. (1998).
  Adapting a synonym database to specific domains.
  Proceedings of ACL'2000 workshop on Recent Advances in NLP and IR.
  Hong Kong, China.

- Vossen, P. (Ed.). (1998b).
  Eurowordnet: A multilingual database with lexical semantic networks.
  Kluwer Academic Publishers.

Vossen, P. (2001).
Extending, trimming and fusing wordnet for technical documents.
NAACL Workshop 'WordNet and Other Lexical Resources: Applications, Extensions and Customizations' (NAACL'2001)..
Pittsburg, PA, USA.

Vossen, P., Bloksma, L., Rodrï¿½uez, H., Climent, S., Roventini, A., Bertagna, F., & Alonge, A. (1997).
The eurowordnet base concepts and top-ontology (Technical Report).
Deliverable D017D034D036 EuroWordNet LE2-4003.

Pantel, P. (2005)
Inducing Ontological Co-occurrence Vectors
Proceedings of ACL'2005

Ann Arbor, USA.

Benitez, L & Cervell, S & Escudero, G & Lopez, M & Rigau, R & Taule, M. (1998)
Methods and Tools for Building the Catalan WordNet
Proceedings of ELRA Workshop on Language Resources for European Minority Languages
Granada, Spain