An Empirical Study for the Automatic Acquisition of Topic Signatures

Montse Cuadros
IXA Group
Univ. of the Basque Country
P.Manuel Irdiazabal, 1
20018 Donostia
mcuadros001@ikasle.ehu.es

German Rigau
IXA Group
Univ. of the Basque Country
P.Manuel Irdiazabal, 1
20018 Donostia
rigau@si.ehu.es

Lluís Padró
TALP Research Center
Technical University of Catalonia
C/Jordi Girona 1-3, Omega S107
08034 Barcelona
padro@lsi.upc.es

Abstract

The main goal of this work is to compare different methods for building Topic Signatures, which are vectors of weighted words acquired from large corpora. We used two different software tools, ExRetriever [Cuadros et al., 2004] and Infomap [Dorow and Widdows, 2003], for acquiring Topic Signatures from corpus. Using these tools, we retrieve sense examples from large text collections. We also include in the comparison the Topic Signatures acquired previously by [Agirre and de la Calle, 2004] from the web. The three systems construct queries for each word sense using WordNet. ExRetriever and Infomap acquire the sense examples from the British National Corpus. The quality of the acquired Topic Signatures is indirectly evaluated on the Word Sense Disambiguation English Lexical Task of Senseval-2.

1 Introduction

Even now, building large and rich knowledge bases takes a great deal of expensive manual effort; this has severely hampered Natural Language Processing (NLP) application development. For example, dozens of person-years have been invested in the development of wordnets for various languages, but the data in these resources is still not sufficiently rich to support advanced concept-based NLP applications directly. Applications will not scale up to working in open domains without more detailed and rich general-purpose and also domain-specific linguistic knowledge build by automatic means.

Topic Signatures (TS) are word vectors related to a particular topic. Topic Signatures are built by retrieving context words of a target topic from large text collections. In our case, we consider word senses as topics. In particular, our task consist on A) acquiring the best possible corpus examples for a particular word sense, and then, B) building the TS by deriving from the selected corpora the context words that best represents the word sense.

TS have been used in a variety of ways, such as in Summarization Tasks [Lin and Hovy, 2000], ontology population [Alfonseca et al., 2004] where they compare different weighting measures to create TS and approximate the link distance between synsets in WordNet [Fellbaum, 1998], or word sense disambiguation (WSD) [Agirre et al., 2000] and [Agirre et al., 2001]. [Agirre and de Lacalle, 2003] shows that the best method to clustering wordnet nominal senses in comparison to other methods proposed in the same work is by using TS. [Agirre and de la Calle, 2004] provide the Topic Signatures[1] for all nominal senses of WordNet [Fellbaum, 1998] using the web as a corpus.

Obviously, part of the success for building high quality TS resides on acquiring high quality sense examples (part A of the acquisition process described above).

Furthermore, some recent work is focusing on reducing the acquisition cost of sense examples to be used by supervised WSD.

In fact, [Leacock et al., 1998; Mihalcea and Moldovan, 1999] and [Agirre and Martinez, 2000] automatically generate arbitrarily large sense corpora to be used by supervised WSD training. This work also uses the knowledge contained in WordNet to formulate search engine queries over large text collections or the Web.

The work of Leacock et al. [Leacock et al., 1998] using AutoTrain retrieves examples of the "closest" word sense relatives first. The quality of the sense examples was evaluated indirectly comparing the results of a WSD system for 14 nouns when trained on sense corpora acquired from monosemous relatives and on manually tagged training materials. The result of this experiment was that some words could be automatically tagged with nearly human rates of success, but there were other words for which automatic tagging was not worthy.

Mihalcea and Moldovan [Mihalcea and Moldovan, 1999] try to overcome these limitations (1) by using the word definitions provided by glosses and (2) by using the Web as a very large corpus. In this case, they use Altavista search engine to create complex queries using boolean operators for increasing the quality of the information retrieved. Their approach was tested on 20 polysemous words giving an accuracy of 91%. Using this method for these words, they obtained thirty times more examples than appearing in SemCor.

Agirre and Martinez [Agirre and Martinez, 2000] implemented the previously described method of Mihalcea and Moldovan to obtain training data for 13 words, and tested on examples from SemCor. Only a few words get better results than random and for a particular word the error rate reached 100%.

[1] http://ixa.si.ehu.es/Ixa/resources/sensecorpus

Petr Sojka, Key-Sun Choi, Christiane Fellbaum, Piek Vossen (Eds.): GWC 2006, Proceedings, pp. 51–59.

Agirre and Martínez suggest that one possible explanation of this apparent disagreement could be that the acquired examples, being correct on themselves, provide systematically misleading features (for instance, as suggested by [Leacock et al., 1998] when using a large set of local closed-class and part-of-speech features).

This work presents a comparison of three different techniques for building Topic Signatures.

The first one, ExRetriever [Cuadros et al., 2004; Cuadros et al., 2005], retrieve sense contexts using queries which consist of a set of literal words. Although these systems have been improved with several enhancements such as term weighting, authority linking, and ad-hoc heuristics to improve their performance, these lexical matching methods can be inaccurate because the queries are based on words instead of concepts. However, there are many ways to characterize a given concept. In this case, the corpus used is the British National Corpus (BNC).

The second technique uses Latent Semantic Indexing (LSI) [Dorow and Widdows, 2003]. LSI tries to overcome the problems of lexical matching by using statistically derived conceptual indexes instead of literal words for retrieval. This technique assumes that there is some underlying latent semantic structure in the data. In this case, the corpus used is also the British National Corpus (BNC).

The third technique, which is very similar to the first one, correspond to the work described in [Agirre and de la Calle, 2004]. In this case, instead of a large text collection such as the BNC, the method uses the web to retrieve sense examples.

Our main goal with this study, as mentioned before, is to compare the performances and quality of these methods for the automatic TS acquisition. In order to perform this comparison, we evaluated the TS acquired by the three systems in a specific task, the English-Lexical Sample task of Senseval-2.

This paper is organised as follows: In section 2, we explain in detail the software tools we use for the task, providing a brief explanation of Latent Semantic Indexing (LSI). In section 3, we explain the steps followed to construct the Topic Signatures, in section 4 we explain the Agreement and Kappa measure. In section 5, the results of the indirect evaluation we carried out. Finally, in section 6, some concluding remarks and future work are provided.

2 Tools

2.1 ExRetriever

ExRetriever is a flexible tool to perform sense queries on large corpora [Cuadros et al., 2004]. This tool 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 priory*. Each different strategy can take into account the information related to each particular word sense and available into a lexical knowledge base in order to automatically generate the set of queries. The lexical knowledge mainly used are the relatives, the synonyms, hyponyms and the words of the definitions. In order to easily implement different query construction strategies, ExRetriever was 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.

2.1.1 The Query Language

ExRetriever query language consist on the following three component types: logical operators, functions and constants.

- **Operators** are the usual boolean operators **and** , **or** and **not** .
- **Functions** Currently implemented,
  - **Glos** used to obtain the words appearing in the gloss.
  - **rel** used to obtain the different relations in the lexical knowledge base
  - **nrel** similar to *rel*, but establishing the maximum polysemy of the returned senses.
- **Constants** can be divided in:
  - **noempty** a parameter for the **Glos** function, used to remove all stopwords from a gloss.
  - **senses** particular senses (e.g. church#n#2)
  - **relations** particular relationships used as parameters to "rel" and "nrel" (e.g. *hypo*).

2.1.2 Example for chair

In this section we explain, using an example, the construction of a query accordingly to a particular query construction strategy. We apply the query strategy **Meaning1**, { Glos(or,and,noempty) **or** or(nrel(1,syns)) **or** or(nrel(1,hypo))} to the third sense of *chair*. Table 1 provides a brief description of word *chair* in WN1.6.

The first function *Glos(or,and,noempty)* returns a logical formula which is the target word (i.e. *chair*) and the union set with *or* of the non *noempty* words of the *gloss*. Applied to chair#n#3: (*chair* AND (*officer* **or** *presides* **or** *meetings* **or** *organization*)). The second function, *or(nrel(1,syns))* returns the union set with *or* of the monosemous synonyms. Applied to chair#n#3: (*chairman* **or** *chairwoman* **or** *chairperson*). Finally, *or(nrel(1,hypo))* returns the union set of the monosemous hyponyms. Applied to chair#n#3: **or** (*vice chairman*). Table 5 shows the resulting queries for all the sense of the word *chair* (noun).

The queries for chair#n for the query Construction Meaning1 strategy, are the following:

- chair#n#1: (*chair* **and** (*seat* **or** *person* **or** *support* **or** *back*)) **or** (*barber chair* **or** *chaise longue* **or** *folding chair* **or** *highchair* **or** *feeding chair* **or** *ladder-back chair* **or** *lawn chair* **or** *garden chair* **or** *rocking chair* **or** *straight chair*
  or *side chair* **or** *swivel chair* **or** *tablet-armed chair* **or** *wheelchair*)
- chair#n#2: (*chair* **and** (*position* **or** *professor*)) **or** (*professorship*)

Table 1: Senses of the noun *chair* in WordNet 1.6

sense	gloss	hypo	syn
n#1	*a* seat *for one* person , *with a* support *for the* back	*armchair (2)* barber_chair ...	
n#2	*the* position *of* professor		professorship
n#3	*the* officer *who* presides *at the* meetings *of an* organization	vice_chairman	*president (6)* chairman chairwoman chairperson
n#4	*an* instrument *of* death *by* electrocution *that* resembles *a* chair		electric_chair death_chair hot_seat

- chair#n#3: (*chair* **and** (*officer* **or** *presides* **or** *meetings* **or** *organization*)) **or** (*chairman* **or** *chairwoman* **or** *chairperson*) **or** (*vice chairman*)
- chair#n#4: (*chair* **and** (*instrument* **or** *death* **or** *electrocution* **or** *resembles*)) **or** (*electric chair* **or** *death chair* **or** *hot seat*)

ExRetriever uses these queries to obtain sense examples (sentences) automatically from a large text collection. The current implementation of ExRetriever accesses directly the content of the Multilingual Central Repository (MCR) [Atserias et al., 2004] of the MEANING project which includes several WordNet versions. The system also uses SWISH-E[2] to index large collections of text such as SemCor [Miller et al., 1993] or BNC. SWISH-E is a fast, powerful, flexible, free, and easy to use system for indexing collections of Web pages or other files. ExRetriever has been designed to be easily ported to other lexical knowledge bases and corpora, including the possibility to query search engines such as Google. In the figure 1, we can see more in detail the chair example explained in this section.

2.2 Infomap

The Infomap NLP Software package[3] uses a variant of Latent Semantic Indexing (LSI) on free-text corpora to learn vectors representing the meanings of words in a reduced vector-space known as Word-Space [Dorow and Widdows, 2003].

The Infomap software performs two basic functions: building models by learning them from a free-text corpus using certain learning parameters specified by the user, and searching an existing model to find the words or documents that best match a query according to that model.

The system can perform information retrieval and word-word semantic similarity computations using the resulting model.

Novel features include a negation operator which uses orthogonal projection, as in quantum logic. The software has been used for lexical acquisition, disambiguation, relation extraction and document retrieval.

[2] `http://swish-e.org`

[3] `http://infomap-nlp.sourceforge.net/`

2.2.1 Latent Semantic Indexing

Latent Semantic Indexing (LSI) allows to extract and represent the contextual meaning of words by statistical computations applied to a large corpus of text [Schütze, 1998]. The underlying idea is that when reducing the dimensionality of the original word-space, similar words are projected closer to each other in the reduced space while dissimilar words are projected to distant locations. The reduced space is obtained using linear algebra methods, in particular, the Singular Value Decomposition (SVD). Part of the motivation for using SVD for word vectors is the success of LSI in information retrieval.

The singular-value decomposition (SVD) technique [Susan T. Dumais and Littman, 1996] is closely related to eigenvector decomposition and factor analysis. For information retrieval we begin with a large term-document matrix, in much the same way as vector or Boolean methods. This term-document matrix is decomposed into a set of k, orthogonal factors from which the original matric can be approximated by linear combination; this analysis reveals the "latent" structure in the matrix that is obscured by noise or by variability in word usage (synonymy and polysemy).

Latent Semantic Indexing maps the contextual relationships between words in terms of common usage across a collection of documents. LSI enables to understand how words relate to each other through the creation of a similarity measure, which reveals whether a given word or document is similarly used compared with another word or document.

3 Strategies for acquiring Topic Signatures

In order to evaluate the performance of both approaches, we designed a preliminary set of strategies for acquiring the Topic Signatures from a given corpus.

3.1 Acquisition Process

The acquisition process consist of the following steps:

1. Devise a particular strategy for query construction and apply the query construction schema to all the senses of a word.
2. Perform the sense queries on the corpus.
3. Collect the sense corpus.
4. Obtain a Topic Signature for each sense.

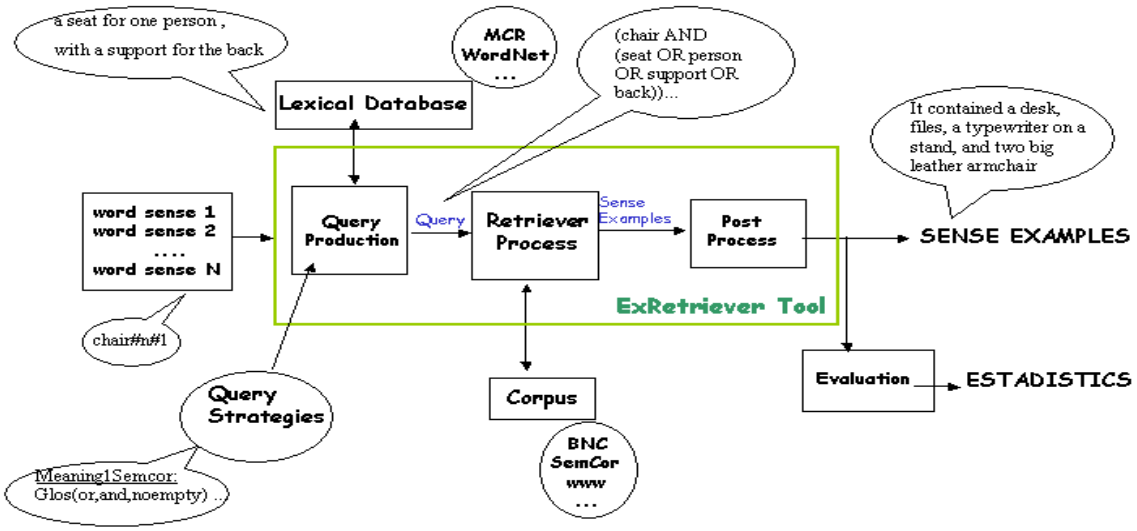

Figure 1: ExRetriever's general schema

3.2 Query construction strategies

We have designed a few preliminary set of query construction strategies based on synonymy, hyponymy and hypernymy relationship of WordNet inspired by the work of [Leacock et al., 1998].

- A) Monosemous strategy: (OR monosemous-words) the union set of all the synonym, hyponym and hyperonym monosemous words of a WordNet sense.
- B) Polysemous strategy: (OR polysemous-words) the union set of all the synonym, hyponym and hyperonym polisemous words of a WordNet sense.
- C) Monosemous and Polysemous strategy: (word AND (OR polysemous-words)) OR* (OR monosemous-words) the union set of all synonym, hyponym and hyperonym monosemous and polisemous words of a WordNet sense in such a way. OR* stands for a particular OR boolean function to express that there is at least one monosemous word or the word and one polysemous word.

We remove those words (monosemous or polysemous) appearing in more than one sense query, trying to construct the sense queries in such a way, that there is no overlapping words in different sense queries of the same word.

3.3 Construction of the Topic Signatures using ExRetriever

These queries have been applied to locate particular sentences of the BNC using ExRetriever. In that way, we are able to retrieve a set of examples for each word sense. In all cases, we remove all stop words from the corpus. Afterwards, we calculate the Mutual Information for each word in the sense corpus with respect to their synset using formula (1).

$$MI(w,s) = log\frac{P(w \wedge s)}{P(w)P(s)} \quad (1)$$

Given a word w and a word sense s, $P(w \wedge s)$ represents the probability of appearing w in the corpus acquired for s sense. $P(w)$ is the probability of occurring w in the BNC corpus, and $P(s)$ is the probability of a document (sentence) to belong to the s sense.

As an example, we will show the full process of obtaining a Topic Signature.

For example, a query of type C for the word *church#n* is constructed using ExRetriever as follows:

In Table 2, there are the three senses of *church#n* at WordNet 1.7.

ExRetriever builds three different queries:

- sense 1: ((*church* **and** (*christianity* **or** *protestant* **or** *religion*)) **or** *christian_church* **or** *catholic_church* **or** *coptic_church*)
- sense 2: ((*church* **and** (*abbey* **or** *basilica* **or** *cathedral*)) **or** *church_building* **or** *kirk* **or** *place_of_worship* **or** *house_of_prayer* **or** *house_of_god*)
- sense 3: ((*church* **and** (*service*)) **or** *church_service* **or** *religious_service* **or** *divine_service*)

Once we construct each sense query, we use ExRetriever to gather all matching sentence examples from the BNC corpus. Afterwards, we calculate the Mutual Information of all the words appearing in the corpora obtained related to a particular word focusing on the wordnet sense.

After this process, we obtain for each word sense, a word vector with weights (Topic Signature). Table presents a part of the TS related to senses 1, 2, and 3 of *church#n* using the strategy A).

Table 2: Sense of *church* noun in wordNet 1.7

sense	**syns**	**hypo**	**hype**
n#1	church Christian_church Christianity	Catholic_Church Coptic_Church Protestant_Church Protestant	religion faith
n#2	church church_building	abbey basilica cathedral kirk	place_of_worship house_of_prayer house_of_God house_of_worship
n#3	church_service church		service religious_service divine_service

Table 3: Example of a Topic Signature obtained with ExRetriever for sense 1, 2 and 3 of church

churches	0.393508
unique	0.374089
traditions	0.374089
today	0.374089
symbol	0.374089
strength	0.374089
step	0.374089
sources	0.374089
significant	0.374089
severe	0.374089
services	0.374089
secure	0.374089
scheme	0.374089
representing	0.374089
remains	0.374089
regard	0.374089
recognized	0.374089
radical	0.374089
door	0.374089
provision	0.374089
promise	0.374089
proceeded	0.374089
priests	0.374089
...	

churchbuilding	0.84563
chucked	0.84563
chosen	0.84563
choke	0.84563
choir	0.84563
chip	0.84563
chilli	0.84563
cherishes	0.84563
chapter	0.84563
chapels	0.84563
chances	0.84563
chancellor	0.84563
champion	0.84563
chambers	0.84563
certificate	0.84563
cerebral	0.84563
cave	0.84563
cautiously	0.84563
caught	0.84563
cashing	0.84563
cash	0.84563
carted	0.84563
carpenter	0.84563
...	

service	2.71512
sermon	2.51729
participants	2.51729
husband	2.51729
burial	2.51729
context	2.41193
witness	2.22961
visible	2.22961
surprised	2.22961
sport	2.22961
sponsor	2.22961
soul	2.22961
smoke	2.22961
shocked	2.22961
royal	2.22961
restrictions	2.22961
relationships	2.22961
radio	2.22961
quickly	2.22961
provided	2.22961
pleasant	2.22961
nominal	2.22961
myth	2.22961
...	

3.4 Construction of the Topic Signatures using Infomap

Infomap only allows AND and ANDNOT operator and does not consider the OR operator. For this reason, the queries have been modified slightly. We use the same words that we used when querying with ExRetriever but we remove all the operators (by default Infomap uses the AND operator).

After building a model with the corpus, the *associate* command of Infomap returns a list of the words or documents best matching the query, in descending order of relevance. Using this option provided by Infomap, once we have the queries, we can get a list of weighted words that we consider the Topic Signature of the query. Table 4 presents the resulting words for sense 1, 2 and 3 of *church#n* using the strategy C) with higher relevance. We have tryied several vector sizes and finally we have used 200 words as a word size vector.

3.5 Using Topic Signatures acquired from the web

Topic Signatures acquired from the web, were constructed using monosemous synonyms or hyponyms to construct the queries. [Agirre and de la Calle, 2004] retrieve the occurrences of the monosemous relatives from Google (up to 1.000 per query), retrieving the context words from he snippets returned. In order to build the Topic Signatures, they follow this method:

- Organize the retrieved examples from the web in collections, one collection per word sense.
- Foreach collection extract the words and their frequencies, and compare them with the data in the collections pertaining to other word senses using statistic, shown in Figure 2.
- The words with distinctive frequency for one of the collections are collected in a list, which constitutes the topic signature for the respective word sense.

Table 4: Example of a Topic Signature obtained with Infomap for sense 1,2,3 of church

christian	0.931791
salvation	0.891434
christians	0.889939
spiritual	0.876109
jesus	0.873607
religion	0.873230
christ	0.872533
worship	0.870086
faith	0.865143
gospel	0.863204
christianity	0.861794
spirit	0.850388
believers	0.846123
scripture	0.840664
god's	0.838087
holy	0.837218
testament	0.832517
heaven	0.828530
church	0.827378
sacred	0.823867
eternal	0.815834
prophets	0.815299
communion	0.806467
teachings	0.804546
god	0.800715
...	

cathedral	0.945974
church	0.901716
abbey	0.900342
chapel	0.874647
priory	0.865647
st	0.857241
trinity	0.855919
paul's	0.838272
peter's	0.816121
parish	0.811324
baptist	0.804895
mary's	0.794569
patron	0.789416
tower	0.786666
congregational	0.780261
castle	0.773482
shrine	0.773119
methodist	0.761350
dei	0.758000
chiswick	0.756093
saint	0.754287
hampstead	0.753816
presbyterian	0.746527
canterbury	0.744466
congregation	0.742333
...	

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

$$tf \cdot idf = \frac{tf_t}{\max_t f_t} \times \log \frac{N}{df_t} \quad (2)$$

4 Indirect evaluation on Word Sense Disambiguation

In order to measure the quality of the acquired TS by these three different approaches, we performed an indirect evaluation by using the acquired Topic Signatures (TS) for a Word Sense Disambiguation (WSD) task. In particular, the Senseval-2 English Lexical Sample task. We used this evaluation framework instead of the the one provided by Senseval-3 because in this case, the verbal part was not directly annotated using WordNet senses. We have calculated for both methods all the Topic Signatures related to all senses of all the Senseval-2 Lexical Sample word-set, which contain words of three different part-of-speech (nouns, verbs and adjectives).

The TS are applied to all the examples of the test set of the Senseval-2 using a simple word overlapping (or weighting) counting. That is, the program calculates the total number of overlapping words between the Topic Signature and the test example. The **occurrence** evaluation measure simply counts the amount of overlapped words and the **weights** evaluation measure counts the weight of the overlapped words. The sense having higher counting (or weighting) is selected for that particular test example. In Table 5, we can see an example of the evaluation test corresponding to sense 3 of *church#n*. As we can see, in bold there are some words that appear in the Topic Signatures for sense 3 obtained using Infomap showed in Table 4, where there are also a part of the Topic Signature for the other three senses.

In Table 6 appears a summary of the results of the indirect evaluation of Infomap and ExRetriever. This table presents the results for each type of query construction strategy (either A, B or C), each system (either Infomap or ExRetriever), and with several levels of sense granularity (either fine or coarse). In this table, P stands for Precision, R for Recall and F1 for F1 measure.

The best figures are obtained by using the Infomap method with occurrences, which is not surprising due to the LSI effect (39.1 precision and recall for fine grained granularity).

As expected, regarding the query construction strategy, in general it seems that strategy A (Monosemous strategy), is better than C (Monosemous and Polysemous strategy) and B (Polysemous strategy), which is the one with the lowest results. We also obtain similar figures with respect occurrences vs. weights methods: using Infomap we obtain slightly better figures for occurrences while when using ExRetriever the best results appear for weights.

In Table 7, we present the results per POS of the queries for each system. We can see that the best query for each POS always rely on A (monosemous strategy), the only difference is that sometimes the best result uses the occurrence or the weight measure method.

In Table 7, we also include the results of the evaluation of the publicly available Topic Signatures acquired from the web. As there are only available the TS for the nom-

Table 5: Test example number 40039, for the church#n#3

In developing measuring tools for the local **church** we are concerned with quality control as much as quantity performance, to use commercial language. Responsible leaders want to know how people are growing in their understanding of the Christian faith, whether relationships are deepening and extending throughout the **church-fellowship**, and to what extent the Christian presence is evident in the community outside. Such information cannot be gathered with such precision as numerical data, but it is essential that each area be investigated to ensure that there is a balance between **worship**, fellowship, learning, evangelism and **service**. Healthy organic growth is proportionate, with each area and function developing in relation to the other. Quality of <*head*> **church** <*head*> life can be measured in the following three ways

Table 6: Overall results of the systems using Senseval-2 with respect fine-grained and coarse-grained senses

		fine			**coarse**		
Method	Query	P	R	F1	P	R	F1
Infomap	A	**39.1**	**39.1**	**39.1**	**51.0**	**51.0**	**51.0**
occurrences	B	37.8	33.2	35.3	50.0	43.8	46.7
	C	37.8	33.2	33.2	50.0	43.8	46.7
Infomap	A	**39.1**	**39.1**	**39.1**	**50.7**	**50.7**	**50.7**
weights	B	38.4	32.8	35.4	49.9	42.7	46.02
	C	38.4	32.8	35.38	49.9	42.7	46.02
ExRetriever	A	**28.5**	**27.1**	**27.8**	**42.3**	**40.3**	**41.3**
occurrences	B	24.1	17.2	20.0	35.4	25.3	29.5
	C	21.7	21.3	21.5	36.6	36.0	36.3
ExRetriever	A	**28.9**	**27.2**	**28.02**	**41.9**	**39.3**	**40.6**
weights	B	22.6	15.9	18.67	33.0	23.2	27.3
	C	25.1	24.6	24.85	36.9	36.1	36.5

inal senses of WordNet, there are only the results for this part-of-speech. In this case, there is a considerable difference between the evaluation using occurrences and weights. It seems that the weighting schema based on the Topic Signatures acquired from the web is not as significant as the appearance of certain words in the Topic Signature.

[Alfonseca et al., 2004] analyze different weighting measures in order to acquire TS: x^2, two versions of tf-idf[1], mutual information (MI) and t-score. The results show that the best rate is for the monosemous relatives construction and when the MI is used as a weighting measure.

4.1 Agreement and Kappa measures

In order to see the different behaviour of the three different methods when acquiring TS, we have calculated the Kappa statistic and the agreement between ExRetriever, Infomap and the Topic Signatures acquired from the web.

We have calculate the Agreement and the Kappa values like in [L.Marquez et al., 2004], where they use this measures to compare the systems presented in Senseval-3 at the Catalan Lexical Sample Task.

The kappa statistic is used to measure interannotation agreement. It determines how strongly two annotators agree by comparing the probability of the two agreeing by chance with the observed agreement. If the observed agreement is significantly greater than that expected by chance, then it is safe to say that the two annotators agree in their judgments.

Mathematically, the formula of the kappa statistic appears in equation 3, where K is the kappa value, $p(A)$ is the probability of the actual outcome and $p(E)$ is the probability of the expected outcome as predicted by chance.

$$K = \log \frac{P(A) - p(E)}{1 - p(E)} \quad (3)$$

In Table 8, there are the results of the kappa measure and the agreement between all the systems for each part-of-speech and overall. In the upper diagonal there is the agreement between the systems and in the lower diagonal there is the kappa statistic between them.

It is considered that Kappa values lower than 0.4 represent poor agreement, values between 0.4 and 0.75 fair to good agreement, and values higher than 0.75 excellent agreement.

The table shows that there is a fair agreement between Infomap and ExRetriever if we take in consideration all part-of-speech. Regarding verbs, ExRetriever and Infomap provide poor agreement and kappa measures. However, for adjectives the agreement and kappa measure are good for both systems. Regading nouns, the results are nearly fair for those systems, while it is poor for ExRetriever and web based TS. Between web based TS and infomap the figures show a fair agreement and kappa measures.

This indicates that the results obtained for nouns from Infomap using LSA and BNC and web based TS are quite similar, while the most disimilar results are obtained when comparing ExRetriever and the web based TS (using both a similar method on different corpora).

4.2 Comparision with other SENSEVAL-2 systems

In Table 10, we present the official results of the Senseval-2 of those systems declared to be unsupervised. When comparing with those systems, Infomap would score second while

Table 7: F1 related to each POS with fine-grained Senseval Evaluation

Method	Query	Noun	Verb	Adj
Infomap	A	40.1	**32.2**	**53.3**
occurrences	B	34.26	29.47	51.29
	C	34.26	29.47	51.29
Infomap	A	**40.6**	31.7	53
weights	B	34.93	29.19	50.77
	C	34.93	29.19	50.77
ExRetriever	A	27.8	**28**	**27.03**
occurrences	C	25.3	17.1	22.79
ExRetriever	A	**34.6**	23.25	23.64
weights	C	32.45	18.2	23.39
Web TS occurrences		**37.2**	-	-
Web TS weights		29.2	-	-

Table 8: Agreement and Kappa results for ExRetriever, Infomap and web TS

	overall			**nouns**			**adj**			**verbs**		
system	Inf.	webTS	ExRetr.	Inf.	webTS	ExRet.	Inf.	webTS	ExRet.	Inf.	webTS	ExRetr.
Inf.	-	-	49.03	-	45.95	39.97	-	-	69.60	-	-	21.35
webTS	-	-	-	0.42	-	26.74	-	-	-	-	-	-
ExRet.	0.47	-	-	0.38	0.25	-	0.67	-	-	0.21	-	-

Table 10: Senseval-2 systems results for fine-grained and coarse-grained senses, in wining order

	fine			**coarse**		
Method	P	R	F1	P	R	F1
UNED - LS-U	40.2	40.1	40.15	51.8	51.7	51.75
Infomap - A occ	**39.1**	**39.1**	**39.1**	**51.0**	**51.0**	**51.0**
ITRI - WASPS-Workbench	58.1	31.9	41.19	66.1	36.3	46.86
CL Research - DIMAP	29.3	29.3	29.3	36.7	36.7	36.7
ExRetriever - A weights	**28.9**	**27.2**	**28.02**	**41.9**	**39.3**	**40.56**
IIT 2 (R)	24.7	24.4	24.55	34.6	34.1	34.35
IIT 1 (R)	24.3	23.9	24.1	34.1	33.6	33.85
IIT 2	23.3	23.2	23.25	32.3	32.2	32.25
IIT 1	22	22	22	32.1	32	32.05

Table 9: Senseval-2 systems results for fine-grained and coarse-grained senses, in wining order only for nouns

Method	fine	coarse
ITRI - WASPS-Workbench	53.2	64.53
UNED - LS-U	44.5	58.1
Infomap - A weights	**40.6**	**54.0**
web TS - occ	**37.2**	**50.5**
ExRetriever - A weights	**34.6**	**48.8**
CL Research - DIMAP	34.3	44.8
IIT 2 (R)	30.85	42.65
IIT 1 (R)	29.44	40.85

ExRetriever fifth getting as a ranking reference the recall of fine-grained score. However, for some authors [Agirre and Martinez, 2004], UNED-LS-U method is considered semi-supervised as the method uses some heuristics that rely on the frequency information available in Semcor. They established a filter to discard the senses that have not appeared more than 10% in the wordnet files. In that way, the sense distribution information is used to discard low-frequency senses.

In Table 9, we also include into the comparison the Topic Signatures acquired from the web, now only considering the nouns of the test set. These results show that the web based TS rates between Infomap TS and ExRetriever TS systems. As the web based TS and ExRetriever use a similar method for acquiring TS, the difference of performance between both systems could rely on the different amount of data (web vs. BNC) and the weighting schema (tf·idf vs. MI). With respect to Infomap, it seems that the LSI effect also overcome the amount of data handled by the web based TS.

We also tried three different vector sizes to evaluate the results between the systems: 200, 400 and 600. We have found out that the best results are obtained by the smallest vector for all the systems.

5 Conclusions

We presented some experiments using different software tools to compare the automatic acquisition of Topic Signa-

tures for word senses. Our Evaluation Framework has been the English Lexical Sample task of Senseval-2. We have focus on the Senseval-2 task because it uses the synsets of WordNet 1.7 for each part of speech, and then is more reliable to our experiments because our queries are build with WordNet 1.7.

We can observe that using Infomap, the tool developed to work with LSI vector models acquired from Corpus, we obtain promising results.

In order to improve the ExRetriever results we plan to filter out those words that seem to be very common in all senses, for example, Named Entities, Multi Words Expressions, etc. or keeping those words that have a common domain or any other semantic relation in common.

Infomap vectors seem to be more accurate for obtaining good context words of an specific word sense. Furthermore, it seems that the results could improve largely varying different system parameters such as dimensionality of the model, size of the Topic Signatures, size of the indexed corpus segments, etc.

We also plan to tune separately each part-of-speech in order to study the different ways to create queries for each of them.

It also deserves further research the study of the different weighting schemas and corpus sizes as shown by the differences between the web based Topic Signatures and the Infomap and ExRetriever systems.

As shown by the agreement and the kappa measures, the behaviour of the three kinds of Topic Signatures is quite different, allowing further improvements by simple combining the methods.

6 Acknowledgements

We want to thank the two reviewers for their valuable comments. This work have been partially suported by the European Comision (MEANING IST-2001-34460), a grant of the + NLP group and the CLASS (1/UPV 00141.226-E-15965/2004) project.

References

E. Agirre and O. Lopez de la Calle. 2004. Publicity available topic signatures for all wordnet nominal senses. In *LREC'04*, pages 97–104.

E. Agirre and O. Lopez de Lacalle. 2003. Clustering wordnet word senses. In *International Conference Recent Advances in Natural Language Processing, RANLP'03*, Borovets, Bulgary.

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

E. Agirre and D. Martinez. 2004. Unsupervised WSD based on automatically retrieved examples: The importance of bias. In *Proceedings of the EMNLP*, Barcelona.

E. Agirre, O. Ansa, D. Martinez, and E. Hovy. 2000. Enriching very large ontologies with topic signatures. In *Proceedings of ECAI'00 workshop on Ontology Learning*, Berlin, Germany.

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

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

Jordi Atserias, Luís Villarejo, German Rigau, Eneko Agirre, John Carroll, Bernardo Magnini, and Piek Vossen. 2004. The meaning multilingual central repository. In *Proceedings of the Second International Global WordNet Conference (GWC'04)*, Brno, Czech Republic, January. ISBN 80-210-3302-9.

M. Cuadros, M. Castillo, G. Rigau, and J. Atserias. 2004. Automatic Acquisition of Sense Examples using ExRetriever. In *Iberamia'04*, pages 97–104.

M. Cuadros, L. Padro, and G. Rigau. 2005. Comparing methods for automatic acquisition of topic signatures. In *Recent Advances in Natural Language Processing (RANLP). Bulgaria. Borovets. 21-23 September. 2005*, pages 181–186.

B. Dorow and D. Widdows. 2003. Discovering corpus-specific word senses. In *EACL*, Budapest.

C. Fellbaum, editor. 1998. *WordNet. An Electronic Lexical Database*. The MIT Press.

C. Leacock, M. Chodorow, and G. Miller. 1998. Using Corpus Statistics and WordNet Relations for Sense Identification. *Computational Linguistics*, 24(1):147–166.

C. Lin and E. Hovy. 2000. The automated acquisition of topic signatures for text summarization. In *Proceedings of 18th International Conference of Computational Linguistics, COLING'00*. Strasbourg, France.

L.Marquez, M.Taule, M.A.Marti, M.Garcia, F.Real, and D.Ferres. 2004. Senseval-3: The catalan lexical sample task. In *Proceedings of the Senseval-3 ACL-SIGLEX Workshop. Barcelona, Spain*.

R. Mihalcea and I. Moldovan. 1999. An Automatic Method for Generating Sense Tagged Corpora. In *Proceedings of the 16th National Conference on Artificial Intelligence*. AAAI Press.

G. Miller, C. Leacock, R. Tengi, and R. Bunker. 1993. A Semantic Concordance. In *Proceedings of the ARPA Workshop on Human Language Technology*.

H. Schütze. 1998. Automatic word sense discrimination. In *Computational Linguistics*.

Thomas K. Landauer Susan T. Dumais and Michael L. Littman. 1996. Automatic cross-linguistic information retrieval using latent semantic indexing. In *Automatic cross-linguistic information retrieval using latent semantic indexing. SIGIR96 Workshop On Cross-Linguistic Information Retrieval*.