Two Web-based approaches for Noun Sense Disambiguation

Paolo Rosso[1], Manuel Montes-y-Gómez[2,1], Davide Buscaldi[3],
Aarón Pancardo-Rodrǵuez[2], and Luis Villaseñor Pineda[2]

[1] Dpto. de Sistemas Informáticos y Computación (DSIC),
Universidad Politécnica de Valencia, Spain
{prosso,mmontes}@dsic.upv.es
[2] Lab. de Tecnologías del Lenguaje
Instituto Nacional de Astrofísica, Optica y Electrónica, Mexico
{mmontesg,aaron_cyberman,villsen}@inaoep.mx
[3] Dipartimento di Informatica e Scienze dell'Informazione (DISI),
Università di Genova, Italy
buscaldi@disi.unige.it

Abstract. The problem of the resolution of the lexical ambiguity, which is commonly referred as Word Sense Disambiguation (WSD), seems to be stuck because of the knowledge acquisition bottleneck. It is worthwhile to investigate the possibility of using the Web as a lexical resource. This paper presents two attempts to use the Web not for extracting training samples but for helping during the WSD task. These two approaches investigate the effectiveness of using the redundancy of the Web to disambiguate nouns in context as well as using modifier adjectives as supporting evidence. Preliminary results show that the direct use of Web statistics allows only for the adjective-noun pairs approach to obtain a better precision than the baseline, even if with a low recall. The evaluation was carried out with different search engines, and the obtained results were almost identical. Finally, the Web was more effective than the WordNet Domains when integrated rather than stand-alone.

1 Introduction

The problem of the resolution of the lexical ambiguity that appears when a given word in a context has several different meanings is commonly referred as Word Sense Disambiguation. The state of the art of WSD [14] shows that the supervised paradigm is the most efficient. However, due to the lack of big sense tagged corpora (and the difficulty of manually creating them), the unsupervised paradigm tries to avoid, or at least to reduce, the knowledge acquisition problem the supervised methods have to deal with. In fact, unsupervised methods do not need any learning process and they use only a lexical resouce (e.g. WordNet) to carry out the word sense disambiguation task [1] [15] [16].

In order to tackle the problem of the knowledge acquisition bottleneck, it seems to make sense to investigate the possibility to use the Web as an extra

lexical resource for WSD. The majority of methods which use the Web try to automatically generate sense tagged corpora [12] [2] [8] [17]. In this paper, we describe our first attempt to use the Web not for extracting training samples but for helping during the word sense disambiguation process. Our work was initially inspired by [13] in which noun-verb relationships are looked for in the Web. The Web as corpus for linguistic research [20] was alredy used with success in many Natural Language Processing areas: question answering [4], question classification [18], machine translation [9], anaphora resolution [6], PP attachment ambiguity [19], and ontology learning [3].

The two approaches we present are based on the idea of redundancy of information in the Web [5]. The first approach exploits the redundancy of the Web to calculate the probabilities associated to each sense of a certain noun according to its context. The second approach tries intuitively to disambiguate a noun on the basis of the adjective that modifies it. Because of the redundancy of the Web, the probability of finding an adjective-noun pair or a noun together with its noun-context increases. The two approaches are based on the hypothesis that a document has a thematic cohesion which is reflected in a semantic relantionship between its words. Therefore, the noun to disambiguate and its noun-context (in the second approach the noun and the adjective which goes before it) have a certain semantic relationship which should become apparent in the Web in a high co-occurrence of the noun-context with the synonyms and hypernyms of the noun to disambiguate (in the second approach, the adjective with the synonyms, the hypernyms and also the hyponyms of the noun).

In the following two sections we describe the two unsupervised context nouns and the adjective-noun pairs Web-based approaches. In the fourth section we present the results of the preliminary experiments, whereas in the fifth section we compare them when different search engines are used. In the last section we discuss the results we obtained when frequency of terms in SemCor was also taken into account. Finally, the conclusions are drawn and the further work is planned.

2 Description of the Noun-Context Approach

Given a noun w, with $|w|$ senses and within a context C of nouns, the function $F(w_k, C)$ indicates the thematic cohesion between the sense w_k of the noun and its context. The estimation of $F(w_k, C)$ is based on the Web. It is carried out by considering the n synonyms $\{s_{ik}, 0 < i \leq n\}$ of the k-th sense w_k of the noun, and the m words in the direct hypernym synset $\{h_{jk}, 0 < j \leq m\}$. The occurrence of these elements within the given context is computed by the fuction $f_S(x, y)$. This function returns the number of pages containing the pattern x AND y, obtained by searching the web with a search engine S. Besides, we name $f_S(x)$ the function returning the number of Web pages containing the string x according to the search engine S. If we assume that $F(w_k, C) \approx P_{Web}(w_k|C)$, i.e., that the thematic cohesion between the k-th sense w_k of the noun and its context is proportional to the probability of finding an occurence of the noun w with sense

w_k in a Web page containing the nouns from the context C, then $F(w_k, C)$ can be calculated using one of the two following formulae:

$$F_1(w_k, C) = \frac{1}{n+m} \left(\sum_{i=1}^{n} P(s_{ik}|C) + \sum_{j=1}^{m} P(h_{jk}|C) \right) \qquad (1)$$

$$F_2(w_k, C) = \arg \max_{\substack{0 < i \leq n, \\ 0 < j \leq m}} (P(s_{ik}|C), P(h_{jk}|C)) \qquad (2)$$

where $P(s_{ik}|C) = f_S(C, s_{ik})/f_S(C)$, and $P(h_{jk}|C) = f_S(C, h_{jk})/f_S(C)$.

Similarly, if we presume that $F(w_k, C) \approx P_{Web}(C|w_k)$, that is, that the thematic cohesion between the k-th sense w_k of the noun and its context is proportional to the probability of finding all the nouns of the context in a Web page containing an occurrence of the noun w with the k-th sense w_k, then the thematic cohesion can be computed using one of the two formulae below:

$$F_3(w_k, C) = \frac{1}{n+m} \left(\sum_{i=1}^{n} P(C|s_{ik}) + \sum_{j=1}^{m} C|P(h_{jk}) \right) \qquad (3)$$

$$F_4(w_k, C) = \arg \max_{\substack{0 < i \leq n, \\ 0 < j \leq m}} (P(C|s_{ik}), P(C|h_{jk})) \qquad (4)$$

where $P(C|s_{ik}) = f_S(C, s_{ik})/f_S(s_{ik})$, and $P(C|h_{jk}) = f_S(C, h_{jk})/f_S(h_{jk})$.

The formulae F_1 and F_3 are based on the *average* weights of the probabilities, and they presume that all the synonyms and hypernyms of a current sense must be related to its context in order to distinguish a thematic cohesion. On the contrary, the formulae F_2 and F_4 are based on the *maximum* of the probabilities (i.e., it is enough to find one frequent synonym-hypernym of the current sense w_k of the noun in a given context in order to establish the thematic cohesion between them).

The algorithm to disambiguate a noun using its noun-context is basically divided into the following steps:

1. Select the set of nouns around the noun to disambiguate w using the sentence as window size (let us named this set C);
2. for each sense w_k of w, and for every synonym s_{ik} and direct hypernym h_{jk} of w_k, compute $f_S(C, s_{ik})$, $f_S(C, h_{jk})$, $f_S(s_{ik})$, and $f_S(h_{jk})$;
3. assign to each sense w_k a weight depending on a function $F(w_k, C)$ which combines the results obtained in the step before;
4. select the sense having the resulting higher weight.

3 Description of the Adjective-Noun Pairs Approach

The disambiguation of a noun w with $|w|$ senses is carried out by taking into account the adjective a referring to the noun itself, the n synonyms $\{s_{ik}, 0 \leq i < n\}$ of the k-th sense (w_k) of the noun, and the m words in the direct hypernym synset of w_k, $\{h_{jk}, 0 \leq j < m\}$ (or also in one of its hyponym synsets). We name $f_S(x, y)$ the function returning the number of pages containing the pair "x y", obtained by searching the Web with a search engine S, where x and y are strings. Moreover, we name $f_S(x)$ the function returning the number of pages containing the string x, by using a search engine S. The weights obtained for the k-th sense (w_k) of the noun to disambiguate w depend on a formula $F(w_k, a)$. The quite comprehensive set of weighting methodologies for combining Web counts used during the experiments is described below. This study give an interesting picture of what are the various possibilities for performing the disambiguation of nouns using modifier adjectives as supporting evidence.

- F_I: This formula is based on the *average* of weights and it takes into account the probabilities of each synonym s_{ik} and each hypernym h_{jk} of having the same sense of w_k. The probability, $p(x|w_k)$, was calculated over the SemCor corpus: even if probabilities vary with domain, in this approximation we assumed that they are the same over the SemCor and the Web.

$$F_I(w_k, a) = 1/2 * \left(\frac{\sum_{i=0}^{n} f_S(a, s_{ik}) * p(s_{ik}|w_k)}{n} + \frac{\sum_{j=0}^{m} f_S(a, h_{jk}) * p(h_{jk}|w_k)}{m} \right)$$
(5)

  The motivation of taking into account this probability is that some words can appear in the Web with a different sense than the appropriate one, e.g. *air* as synonym of *melody* is rare, with a probability $p("air"|6598312) = 0.0022$, where 6598312 is the WordNet 2.0 offset corresponding to the synset {*tune, melody, strain, air, line, melodic line, melodic phrase*}.

- F_{II}: This formula derives directly from F_I and it takes into account also the hyponyms of the sense w_k of the noun to disambiguate. The hyponyms can be seen as "use cases" of the sense of the word to disambiguate. The hyponym weights are computed in exactly the same way of the synonyms and hypernyms in the formula above, where 1/2 is replaced by 1/3.

- F_{III}: This formula calculates the *maximum* of weights instead of the average. It also takes into account the probabilities of synonyms, hypernyms and hyponyms.

$$F_{III}(w_k, a) = \max_{\substack{0 \leq i < n, \\ 0 \leq j < m}} (f_S(a, s_{ik}) * p(s_{ik}|w_k), f_S(a, h_{jk}) * p(h_{jk}|w_k))$$
(6)

- F_{IV}: This formula is based on the *Mean Mutual Information*[21], or *Relative Entropy, similarity measure*. The formula measures how much information

is in the dependency of two successive words. It has been adapted to take into consideration information obtained both by synonyms and hypernyms:

$$F_{IV}(w_k, a) = \sum_{i=0}^{n} f_S(a, s_{ik}) \log \frac{f_S(a, s_{ik})}{f_S(a) * f_S(s_{ik})} + \sum_{j=0}^{m} f_S(a, h_{jk}) \log \frac{f_S(a, h_{jk})}{f_S(a) * f_S(h_{jk})}$$

(7)

The algorithm to disambiguate a noun using an adjective as modifier is basically divided into the following steps:

1. Select the adjective a immediately before the noun w;
2. for each sense w_k of w, and for every synonym s_{ik} and direct hypernym h_{jk} of w_k, compute $f_S(a, s_{ik})$ and $f_S(a, h_{jk})$ (in some formulae we used also the direct hyponyms of the noun);
3. assign to each sense w_k a weight depending on a formula F which combines the results obtained in the step before;
4. select the sense having the resulting higher weight.

For example, consider the following sentence, extracted from the Senseval-3 All-Words task corpus: *A faint crease appeared between the man's eyebrows.* Suppose we are disambiguating the word *crease*, having three senses, according to WordNet 2.0: $crease_1 : \{fold, crease, plication, flexure, crimp, bend\}$, $crease_2 : \{wrinkle, furrow, crease, crinkle, seam, line\}$ and $crease_3 : \{kris, creese, crease\}$. The direct hyperonyms are, for each sense: $h_1 = \{angular\ shape,\ angularity\}$, $h_2 = \{depression,\ impression,\ imprint\}$ and $h_3 = \{dagger,\ sticker\}$. Then we search the web for the following pairs: (*faint fold*), (*faint plication*), (*faint flexure*), (*faint crimp*), (*faint bend*), (*faint angular shape*), (*faint angularity*) for the first sense, (*faint wrinkle*), (*faint furrow*), etc. for the second sense and so on.

4   Preliminary Experimental Results

The preliminary noun sense disambiguation experiments were carried out over 215 nouns of the Senseval-3[4] corpus for the English all words task [22]. Web counts were collected through the MSN Search[5]. In the first approach, frequencies are used to calculate the cohesion of the different noun senses with respect to its noun-context. Table 1 shows the poor results obtained for the noun-context approach, always below the Most Frequently Sense (MFS) baseline (0.689 over the all Senseval-3 English all-words task corpus). The two leftmost columns show the overall precision obtained by the four formulae when just the best answer was accepted, whereas the following two columns illustrate the precision when the first two answers were accepted. Finally, the two rightmost columns indicate that also for a fuzzy WSD in which $\lceil n/2 \rceil$ answers were accepted results were pretty poor. The aim of this part of the experiments was to understand if this approach could be useful at least for putting aside the not likely senses.

---

[4] http://www.senseval.org
[5] http://search.msn.com

F	P	F	P	F	P
F_1	0.181	F_1	0.362	F_1	0.548
F_2	0.190	F_2	0.358	F_2	0.609
F_3	0.209	F_3	0.400	F_3	0.637
F_4	0.218	F_4	0.469	F_4	0.679

Table 1. Noun-context approach. S: number of senses allowed in the evaluation; F: Formula; P: overall Precision; columns 1-2: best answer, columns 3-4: best two answers, columns 5-6: best $\lceil n/2 \rceil$ answers (fuzzy WSD).

In all cases, formulae F_3 and F_4 behaved better than F_1 and F_2. This result is quite important because the hypothesis of thematic cohesion in the Web described in the second section shows that normalising with respect to $f_S(s_{ik})$ and $f_S(h_{ik})$ makes the calculation of probabilities less sensitive to very rare or general synonyms and hypernyms. Finally, we observed that also the formulae F_2 and F_4 which take into account the maximum probability between the context and one of the synonyms or the hypernyms of the noun to disambiguate, are also less sensitive to very rare or general synonyms and hypernyms (the calculation of the maximum seems to be less sensitive than the average).

The result analysis allowed us to understand that one of the reasons of the poor performance of the method could be that whereas a window of a sentence is used to select the set of nouns of the context, the pattern x AND y may be contained in the context of a document (the Web page returned by the search engine).

The approach failed especially for the disambiguation of highly polysemic nouns (with more than five senses on average), and in case of right answer, generally, the probability of the right sense was much higher than those of the other senses. Therefore, Web-based approaches like this could be more effective if integrated with other WSD methods rather than stand-alones.

F	P	R	C	P_{na}
MFS	0.689	0.689	100%	0.623
F_I	0.627	0.271	43.3%	0.318
F_{II}	0.661	0.286	43.3%	0.392
F_{III}	0.660	0.278	42.0%	0.373
F_{IV}	0.579	0.239	41.2%	0.179

Table 2. Adjective-noun pairs approach. F: formula; P: overall Precision; R: overall Recall; C: overall Coverage; P_{na}: Precision over the disambiguated nouns (i.e., nouns with an adjective before: "adjective noun").

With respect to the adjective-noun pairs approach, Table 2 shows the results of the preliminary experiments we carried out using the MSN search engine and

the formulae described in the third section. For each formula we obtained worse results than the Most Frequently Sense baseline. An interesting result is that the hyponym information seems to be helpful when using the Web to perform WSD. For instance, the formula F_{II} obtained a better performance than the related formula F_I which used only synonyms and direct hypernyms.

5 Comparison of Search Engines

The results of the noun-context approach obtained using the MSN search engine were compared with those obtained when AltaVista[6] and Google[7] were used. Table 3 shows the results of this comparison. It is interesting to notice that the precision obtained with the three different search engines are almost identical (a difference of 0.03 on average).

F	P_{MSN}	P_{AV}	P_G	F	P_{MSN}	P_{AV}	P_G	F	P_{MSN}	P_{AV}	P_G
F_1	0.181	0.186	0.227	F_1	0.362	0.353	0.348	F_1	0.548	0.534	0.595
F_2	0.190	0.186	0.279	F_2	0.358	0.339	0.395	F_2	0.609	0.576	0.595
F_3	0.209	0.215	0.237	F_3	0.400	0.386	0.432	F_3	0.637	0.613	0.641
F_4	0.218	0.215	0.251	F_4	0.469	0.427	0.451	F_4	0.679	0.623	0.646

Table 3. Noun-context approach: comparison of search engines. S: number of senses allowed in the evaluation; F: formula; P_{MSN}: overall Precision with MSN; P_{AV}: overall Precision with Altavista; P_G: overall Precision with Google; columns 1-4: best answer, columns 5-8: best two answers, columns 9-12: best $\lceil n/2 \rceil$ answers (fuzzy WSD).

With respect to the adjective-noun pairs approach, the experimental results of the fourth section, in which with the MSN search engine was used, were compared to those obtained when AltaVista. In this comparison, we decided also to use the Lucene search engine[8], substituting the Web with the TREC-8 collection[9], in order to evaluate the effectiveness of the Web with respect to a large document collection. We calculated the precision, recall and coverage over the Senseval-3 AWT corpus, for every search engine, using the formulae described in the third section. The results obtained evidentiate that MSN and AltaVista are equivalent (even if we obtained slight differences in some experiments, on average results are almost the same for both engines). We decided to use MSN to carry out the experiments due to a lower response time for the queries (e.g. the duration of the most demanding experiment needed 237 minutes with AltaVista whereas 171 minutes with MSN). A remarkable difference between the Web-based search engines and the Lucene is the 6% drop in coverage obtained when using the

---

[6] http://www.altavista.com
[7] http://www.google.com
[8] http://jakarta.apache.org/lucene
[9] http://trec.nist.gov

TREC-8 collection instead of the Web, confirming that the huge redundancy of data in the web allows to disambiguate a greater number of words. Moreover, the precision obtained with Lucene is only 1% higher than the precision obtained when using the web. We expected a higher precision, due to the lower quality of the information in the Web. However the improvement is not so evident to justify the use of a large text collection instead of the Web. Figure 1 shows the obtained results.
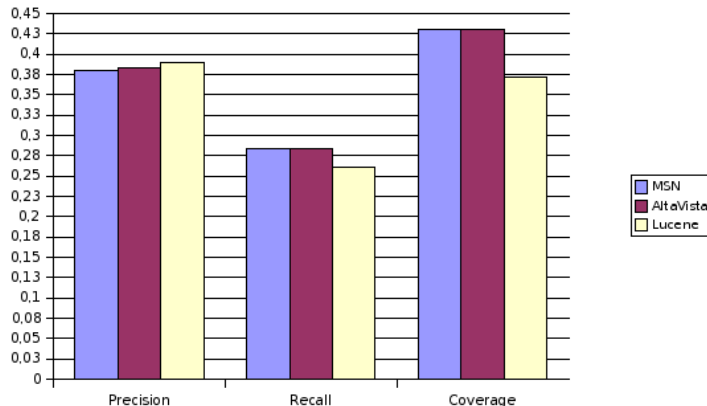
Fig. 1. Adjective-noun pairs approach: comparison of search engines; averaged Precision (polysemic words only), Recall and Coverage over the used formulae.

6 Experimental Results with Frequency Correction

A second tranche of experiments was carried out just for the adjective-noun pairs approach which obtained better results. In this second attempt, we decided to investigate the possibility of including a *frequency correction* (fc) factor in each of the formulae described in the third section. This frequency factor indicates whether the resulting weight for a sense w_k was multiplied by $p(w|w_k)$, that is, the probability for the word w of having sense w_k in the SemCor corpus, or not. Moreover, a new formula based on a different similarity measure was studied during these experiments. This formula resembles the *Similarity Theorem* [10], and the problem due to large denominators is reduced thanks to the use of logarithms.

$$F_V(w_k, a) = \max_{\substack{0 \le i < n, \\ 0 \le j < m}} \left(f_S(a, s_{ik}) * \frac{\log f_S(a, s_{ik})}{\log f_S(s_{ik})}, f_S(a, h_{jk}) * \frac{\log f_S(a, h_{jk})}{\log f_S(h_{jk})} \right) \quad (8)$$

Table 4 shows that the frequency-corrected formulae outperform those without the frequency factor. Moreover, an average 4% gain is obtained in recall when

frequency is taken into account. In one case (with F_V) we obtained better results than the MFS baseline. We believe that it could depend on the fact that this formula has the advantage of not taking into account only the relevance of the adjective but also the number of co-occurrences of the pair adjective-noun. We also compared the adjective-noun pairs approach with one based on Conceptual Density (CD) [7] and WordNet Domains (WND) [11]. The Web-based disambiguation provided better results, especially over the words not disambiguated by the standard CD method (16.6% with respect to the CD+WND formula). Therefore, the Web was more effective than the WordNet Domains. Moreover, the Web-based approach was more effective when integrated with another system rather than stand-alone.

F	fc	P	R	C	P_{na}	P_{nd}
MFS		**0.689**	0.689	100%	**0.623**	**0.629**
CD		0.734	0.518	70.5%	0.625	0.000
$CD + WND$		0.653	0.584	89.3%	0.583	0.500
F_I	no	0.627	0.271	43.3%	0.318	0.328
F_I	yes	0.718	0.311	43.3%	0.511	0.478
$F_I I$	no	0.661	0.286	43.3%	0.392	0.367
$F_I I$	yes	0.759	0.329	43.3%	0.596	0.507
$F_I II$	no	0.660	0.278	42.0%	0.373	0.333
$F_I II$	yes	0.755	0.326	43.1%	0.586	0.500
F_{IV}	no	0.579	0.239	41.2%	0.179	0.152
F_{IV}	yes	0.720	0.259	42.1%	0.532	0.565
F_V	yes	**0.777**	0.337	43.3%	**0.634**	**0.666**

Table 4. Adjective-noun pairs approach with and without frequency factor. F: Formula; fc: frequency correction; P: overall Precision; R: overall Recall; C: overall Coverage; P_{na}: Precision over the disambiguated nouns (i.e., nouns with an adjective before: "adjective noun"); P_{nd}: Precision over the nouns not disambiguated by the CD method.

We investigated the importance of polysemy of the adjective in the disambiguation of a noun. We calculated the averaged polysemy of the adjective when the referred word was disambiguated correctly and when the word was assigned the wrong sense. The results showed in Table 5 demonstrate that the less polysemic is the adjective, the higher will be the probability of selecting the right sense. However, frequency-corrected formulae tend to be less subject to the polysemy of the adjective, obtaining values for the polysemy of the adjective closer to the values obtained with the MFS heuristics.

Finally, we investigated also the possibility of using the same counter-intuitive approach to disambiguate an adjective based on the noun which goes after (i.e., using the Web to look for $f_S(a_{ik}, w)$, where a_{ik} is the i-th synonym of the k-th sense of adjective a). In fact, traditionally, this is done the other way around and evidence for the sense of an adjective is found by looking at the noun it modifies. In order to do so, we used an equivalent formula to F_V (the formula

F	fc	$Right$	$Wrong$
MFS		4.26	4.3
F_I	no	3.65	4.55
F_I	yes	4.18	4.40
F_{II}	no	3.28	4.86
F_{II}	yes	4.17	4.40
F_{III}	yes	4.17	4.41
F_{IV}	no	3.9	4.36
F_V	yes	4.87	5.54

Table 5. Polysemy of adjectives. F: Formula; fc: frequency correction; $Right$: average polisemy of adjectives for correctly disambiguated nouns; $Wrong$: average polisemy of adjectives for incorrectly disambiguated nouns.

we obtained the best results for the disambiguation of nouns). Unfortunately, in this first attempt we obtained quite poor results (21.3% precision).

7 Conclusions and Further Work

The paper explores the disambiguation of nouns using Noun-context or modifier adjectives as supporting evidence, and using Web counts collected through different search engines. The comparison we made across the different search engines should be useful, and we consider interesting that the Web counts (and consequently disambiguation results) obtained with different search engines are almost identical (despite the fact that the search engines considered in the experiments could cover different sections of the Web).

The main aim of the paper is to bring a contribution in terms of various weighting methodologies for Web counts, and in terms of insights into methodologies that work best for the purpose of word sense disambiguation. A noun-context approach was first investigated. In this approach the hypothesis of thematic cohesion in a document is made and the sense is chosen as a statistics of the co-occurrence in the Web of the context and the synonyms and hypernyms of the noun to disambiguate. Although the system obtained very poor results (only a precision of 28%, if only one sense is accepted, and of 68% in case of fuzzy WSD) it could be a promising approach if more contextual information rather nouns is taken into account to increase the precision (especially for highly polysemic nouns). As further work we plan to carry out some experiments including hyponyms instead of hypernyms (we realised that hypernyms do not characterise very well the meaning of a particular noun sense with respect to the other senses of the same noun, whereas hyponyms usually do better when dealing with corpora). Moreover, it could be also interesting to take into account the probability in SemCor for each noun (as we did for the adjective-noun pairs approach).

The approach based on adjective-noun pairs obtained instead a better precision than the baseline, even if with a low recall. We believe that this depends on

that the majority of pairs is still ambiguous. That is, the adjective is not enough to understand the meaning of the noun and a bigger context should be taken into account. Our study over the importance of the polisemy of the adjective in the disambiguation seems to confirm our intuition. For example, the pair *cold fire* is ambiguous, since it can be assigned both the sense corresponding to *cold passion* or the sense corresponding to *cold fire.*

We detected some problems in the use of WordNet synonyms and hypernyms, since they are composed of multi-word expressions rarely found in the Web. Our further investigation directions will be to investigate the use of another ontology to overcome the multi-word expression issue, as well as to use shallow parsers in order to determine an unambiguous context for the word to disambiguate. It could be more intuitive to disambiguate a noun based on a syntactically related verb than on a modifier adjective. The aim is to understand whether it is most likely obtaining significantly better results using such noun-verb relationships.

We conclude remarking that preliminary results showed that it should be better using the Web as lexical resource for WSD if integrated with existing systems rather than using it stand-alone. Moreover, when we integrated the Web, instead of the WordNet Domains, with a unsupevised method based on conceptual density we obtained better results.

As further work for both approaches we plan to take into consideration the complete direct surroundings of each concept: hyponyms, part-of, derived, or even words from the corresponding gloss. Moreover, with respect to the comparison across search engines, it would also be interesting to make a study across results provided by the same search engine, at different points in time. This could make an interesting result that could validate (invalidate) claims made by some researchers that Web counts are not very stable over time.

Acknowledgments

References

1. Agirre, E., Rigau, G.: A Proposal for Word Sense Disambiguation using Conceptual Distance. In: Proc. of the Int. Conf. on Recent Advances in NLP. RANLP'95. (1995)
2. Agirre, E., Martinez, D.: Exploring Automatic Word Sense Disambiguation with Decision Lists and the Web. In: Proc. of the the COLING-2000. (2000)
3. Agirre, E., Olatz, A., Hovy, Martinez, D.: Enriching Very Large Ontologies using the WWW. ECAI 2000, Workshop on Ontology Learning. Berlin. (2000)
4. Brill, E., Lin, J., Banko, M., Dumais, S., Ng, A.: Data-intensive Question Answering. In: Proc. of the Tenth Text REtrieval Conference TREC-2001. (2001).

5. Brill, E.: Processing Natural Language Processing without Natural Language Processing. Lecture Notes in Computer Science, Vol. 2588. Springer-Verlag (2003) 360–369.

6. Bunescu, R.: Associative Anaphora Resolution:A Web-Based Approach. In: Proc. of the EACL-2003 Workshop on the Computational Treatment of Anaphora, Budapest, Hungary, April. (2003).

7. Buscaldi, D., Rosso, P., Masulli, F.: The upv-unige-CIAOSENSO WSD System. Senseval-3 Workshop, Association for Computational Linguistics (ACL-04). Barcelona, Spain (2004).

8. Gonzalo, J., Verdejo, F., Chugar, I.: The Web as a Resource for WSD. In: 1st MEANING Workshop, Spain. (2003)

9. Grefenstette, G.: The World Wide Web as a resource for example-based Machine Translation Tasks. In: Proc. of Aslib Conference on Translating and the Computer. London. (1999)

10. Lin, D.: An Information-Theoretic Definition of Similarity. In: Proc. of the 15th Int. Conf. on Machine Learning. Toronto, Canada (2003)

11. Magnini, B. and Cavaglià, G.: Integrating Subject Field Codes into WordNet. In: Proc. of LREC-2000, 2nd Int. Conf. on Language Resources and Evaluation. (2000) 1413–1418.

12. Mihalcea, R., Moldovan, D.I.: An Automatic Method for Generating Sense Tagged Corpora. In: Proc. of the 16th National Conf. on Artificial Intelligence. AAAI Press. (1999)

13. Mihalcea, R., Moldovan, D.I.: A Method for Word Sense Disambiguation of Unrestricted Text. In: Proc. of the 37th Annual Meeting of the Association for Computational Linguistics (ACL-99). Maryland, NY, U.S.A. (1999)

14. Mihalcea, R., Edmonds, P. (Eds.): Proc. of Senseval-3: The 3rd Int. Workshop on the Evaluation of Systems for the Semantic Analysis of Text. Barcelona, Spain. (2004)

15. Montoyo, A: Método basado en Marcas de Especificidad para WSD. Procesamiento del Lenguaje Natural. In: Revista n 26, Septiembre. (2000).

16. Rosso, P., Masulli, F., Buscaldi, D., Pla, F., Molina, A.: Automatic Noun Disambiguation. Lecture Notes in Computer Science, Vol. 2588. Springer-Verlag (2003) 273–276.

17. Santamaria, C., Gonzalo, J., Verdejo, F.: Automatic Association of WWW Directories to Word Senses. Computational Linguistics (2003), Vol. 3, Issue 3 - Special Issue on the Web as Corpus, 485–502.

18. Solorio, T., Pérez, M., Montes, M., Villaseñor, L., López, A.: A Language Independent Method for Question Classification. In: Proc. of the 20th Int. Conf. on Computational Linguistics (COLING-04). Geneva, Switzerland. (2004)

19. Volk, M.: Exploiting the WWW as a Corpus to Resolve PP Attachment Ambiguities. In: Proc. of Corpus Linguistics. Lancaster. (2001)

20. Volk, M.: Using the Web as Corpus for Linguistic Research. In: Catcher of the Meaning. Pajusalu, R., Hennoste, T. (Eds.). Dept. of General Linguistics 3, University of Tartu, Germany. (2002)

21. Wackerbauer, R., Witt, A., Atmanspacher, H., Kurths, J., Scheingraber, H.: A Comparative Classification of Complexity Measures. Chaos, Solitons and Fractals, 4: 133-173 (1994).

22. Zinder, B.,Palmer, M.: In: Proc. of Senseval-3: The 3rd Int. Workshop on the Evaluation of Systems for the Semantic Analysis of Text. Barcelona, Spain. (2004)