

The upv-unige-CIAOSENSO WSD System

Davide Buscaldi^{1,2}, Paolo Rosso², Francesco Masulli³

¹DISI, Università di Genova, Italy

²DSIC, Universidad Politécnica de Valencia, Spain

³INFM-Genova and Dip. di Informatica, Università di Pisa, Italy

²{dbuscaldi, proso}@dsic.upv.es

³masulli@disi.unige.it

Abstract

The CIAOSENSO WSD system is based on Conceptual Density, WordNet Domains and frequencies of WordNet senses. This paper describes the *upv-unige-CIAOSENSO* WSD system, we participated in the english all-word task with, and its versions used for the english lexical sample and the WordNet gloss disambiguation tasks. In the last an additional goal was to check if the disambiguation of glosses, that has been performed during our tests on the SemCor corpus, was done properly or not.

Introduction

The CIAOSENSO WSD system is an unsupervised system based on *Conceptual Density* (Agirre and Rigau, 1995), frequencies of WordNet senses, and WordNet Domains (Magnini and Cavagli`a, 2000).

Conceptual Density (*CD*) is a measure of the correlation among the sense of a given word and its context. The foundation of this measure is the *Conceptual Distance*, defined as the length of the shortest path which connects two concepts in a hierarchical semantic net. The starting point for our work was the *CD* formula of Agirre and Rigau (Agirre and Rigau, 1995), which compares areas of subhierarchies. The noun sense disambiguation in the CIAOSENSO WSD system is performed by means of a formula combining Conceptual Density with WordNet sense frequency (Rosso et al., 2003).

WordNet Domains is an extension of WordNet 1.6, developed at ITC-irst¹, where each synset has been annotated with at least one domain label, selected from a set of about two hundred labels hierarchically organized (Magnini and Cavagli`a, 2000). Since the lexical resource used by the upv-unige-CIAOSENSO WSD system is WordNet 2.0 (WN2.0), it has been necessary to map the synsets of WordNet Domains from version 1.6 to the version 2.0. This has been done in a fully automated way, by using the WordNet mappings for nouns and

verbs, and by checking the similarity of synset terms and glosses for adjectives and adverbs. Some domains have also been assigned by hand in some cases, when necessary.

1 Noun Sense Disambiguation

In our upv-unige-CIAOSENSO WSD system the noun sense disambiguation is carried out by means of the formula presented in (Rosso et al., 2003), which gave good results for the disambiguation of nouns over the SemCor corpus (precision 0.815). This formula has been derived from the original Conceptual Density formula described in (Agirre and Rigau, 1995):

$$CD(c, m) = \frac{\sum_{i=0}^{m-1} nhyp^i}{\sum_{i=0}^{h-1} nhyp^i} \quad (1)$$

where c is the synset at the top of subhierarchy, m the number of word senses falling within a subhierarchy, h the height of the subhierarchy, and $nhyp$ the averaged number of hyponyms for each node (synset) in the subhierarchy. The numerator expresses the expected area for a subhierarchy containing m marks (word senses), while the divisor is the actual area.

Due to the fact that the averaged number of hyponyms for each node in WN2.0 is greater than in WN1.4 (the version which was used originally by Agirre and Rigau), we decided to consider only the *relevant* part of the subhierarchy determined by the synset paths (from c to an ending node) of the senses of both the word to be disambiguated and its context. The base formula is based on the M number of relevant synsets, corresponding to the *marks* m in Formula 1 ($|M|=|m|$, but we determine the subhierarchies before adding such marks instead of vice versa like in (Agirre and Rigau, 1995)), divided by the total number nh of synsets of the subhierarchy.

$$baseCD(M, nh) = M/nh \quad (2)$$

The original formula and the above one do not take into account sense frequency. It is possible that both

¹Istituto per la Ricerca Scientifica e Tecnologica, Trento, Italy

formulas select subhierarchies with a low frequency related sense. In some cases this would be a wrong election. This pushed us to modify the CD formula by including also the information about frequency that comes from WN:

$$CD(M, nh, f) = M^\alpha (baseCD)^{\log f} \quad (3)$$

where M is the number of relevant synsets, α is a constant (the best results were obtained over the SemCor corpus with α near to 0.10), and f is an integer representing the frequency of the subhierarchy-related sense in WN (1 means the most frequent, 2 the second most frequent, etc.). This means that the first sense of the word (i.e., the most frequent) gets at least a density of 1 and one of the less frequent senses will be chosen only if it will exceed the density of the first sense. The M^α factor was introduced to give more weight to the subhierarchies with a greater number of relevant synsets, when the same density is obtained among many subhierarchies.

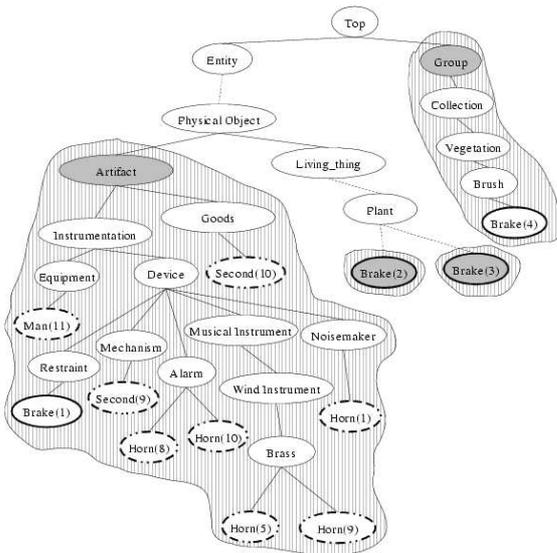


Figure 1: Subhierarchies resulting from the disambiguation of *brake* with the context words $\{horn, man, second\}$. Example extracted from the Senseval-3 english-all-words test corpus. ($M_1 = 11$, $nh_1 = 21$, $M_2 = M_3 = nh_2 = nh_3 = 1$, $M_4 = 1$, $nh_4 = 5$, where M_i and nh_i indicates, respectively, the M and nh values for the i -th sense)

In figure 1 are shown the resulting WordNet subhierarchies from the disambiguation of *brake* with the context words $\{horn, man, second\}$ from the sentence: “Brakes howled and a horn blared furiously, but the man would have been hit if Phil hadn’t called out to him a second before”, extracted from the all-words test corpus. The areas of subhierarchies are drawn with a dashed background, the root of subhierarchies are the darker nodes, while

the nodes corresponding to the synsets of the word to disambiguate and those of the context words are drawn with a thicker border. Four subhierarchies have been identified, one for each sense of *brake*. The senses of the context words falling out of these subhierarchies are not taken into account. The resulting CDs are, for each subhierarchy, respectively: $11^{0.10} * (11/21)^{\log 1} = 1.27$, 1, 1 and $1^{0.10} * (1/5)^{\log 4} = 0.07$, therefore the first one is selected and sense 1 is assigned to *brake*.

In the upv-unige-CIAOSENSE WSD system, additional weights (*Mutual Domain Weights, MDWs*) are added to the densities of the subhierarchies corresponding to those senses having the same domain of context nouns’ senses. Each weight is proportional to the frequency of such senses, and is calculated as $MDW(f, i) = 1/f * 1/i$, where f is an integer representing the frequency of the sense of the word to be disambiguated and i gives the same information for the context word. E.g. if the word to be disambiguated is *doctor*, the domains for senses 1 and 4 are, respectively, *Medicine* and *School*. Therefore, if one of the context words is *university*, having the third sense labeled with the domain *School*, the resulting weight for *doctor(4)* and *university(3)* is $1/4 * 1/3$.

Those weights are not considered in the upv-unige-CIAOSENSE2 system, which has been used only for the all-words task.

We included some adjustment factors based on context hyponyms, in order to assign an higher conceptual density to the related subhierarchy in which a context noun is an hyponym of a sense of the noun to be disambiguated (the hyponymy relation reflects a certain correlation between the two lexemes). We refer to this technique as to the *Specific Context Correction (SCC)*. The idea is to select as the winning subhierarchy the one where one or more senses of the context nouns fall beneath the synset of the noun to be disambiguated.

An idea connected to the previous one is to give more weight to those subhierarchies placed in deeper positions. We named this technique as *Cluster Depth Correction (CDC)* (we use improperly the word “cluster” here to refer to the *relevant* part of a subhierarchy). When a subhierarchy is below a certain averaged depth (which was determined in an empirical way to be approximately 4) and, therefore, its sense of the noun to be disambiguated is more specific, the conceptual density of Formula 3 is augmented proportionally to the number of the contained relevant synsets:

$$CD * (depth(sh) - avgdepth + 1)^\beta \quad (4)$$

where $depth(sh)$ returns the depth of the current subhierarchy (sh) with respect to the top of the WordNet hierarchy; $avgdepth$ is the averaged depth of all subhierarchies in SemCor; its value, as said before, was empirically determined to be equal to 4; and β is a constant (the best results were obtained, over SemCor, with $\beta = 0.70$).

These depth corrections have been used only in the upv-unige-CIAOSENSO-eaw and upv-unige-CIAOSENSO-ls systems for the english all-words task and english lexical sample tasks. We found that they are more useful when a large context is available, and this is not the case of the gloss disambiguation task, where the context is very small. Moreover, in the upv-unige-CIAOSENSO2 system we aimed to achieve the best precision, and these corrections usually allow to improve recall but not precision.

2 Adjectives, Verbs and Adverbs Sense Disambiguation

The disambiguation of words of POS categories other than noun does not take into account the Conceptual Density. This has been done for the following reasons: first of all, it could not be used for adjectives and adverbs, since in WordNet there is not a hierarchy for those POS categories. With regard to verbs, the hierarchy is too shallow to be used efficiently. Moreover, our system performs the disambiguation one sentence at a time, and this results in having in most cases only one verb for each sentence (with the consequence that no density can be computed).

The sense disambiguation of an adjective is performed only on the basis of the domain weights and the context, constituted by the *Closest Noun (CN)*, i.e., the noun the adjective is referring to (e.g. in “family of musical instruments” the CN of *musical* is *instruments*). Given one of its senses, we extract the synsets obtained by the *antonymy*, *similar_to*, *pertainymy* and *attribute* relationships. For each of them, we calculate the MDW with respect to the senses of the context noun. The weight assigned to the adjective sense is the average between these MDWs. The selected sense is the one having the maximum average weight.

In order to achieve the maximum coverage, the *Factotum* domain has been also taken into account to calculate the MDWs between adjective senses and context noun senses. However, due to the fact that in many cases this domain does not provide a useful information, the weights resulting from a *Factotum* domain are reduced by a 0.1 factor. E.g. suppose to disambiguate the adject-

ive *academic* referring to the noun *credit*. Both *academic(1)* and *credit(6)* belong to the domain *School*. Furthermore, the *Factotum* domain contains the senses 1 4 and 7 of *credit*, and senses 2 and 3 of *academic*. The extra synsets obtained by means of the WN relationships are: *academia(1):Sociology*, pertainym of sense 1; *theoretical(3):Factotum* and *applied(2):Factotum*, similar and antonym of sense 2; *scholarly(1):Factotum* and *unscholarly(1):Factotum*, similar and antonym of sense 3. Since there are no senses of *credit* in the *Sociology* domain, *academia(1)* is not taken into account. Therefore, the resulting weights for *academic* are:

$$1 * 1/6 = 0.16 \text{ for sense 1;}$$

$$0.1 * (1/2 + 1/2 * 1/4 + 1/2 * 1/7 + [1/2 * 1/3 + 1/2 * 1/2])/5 \simeq 0.02 \text{ for sense 2;}$$

$$0.1 * (1/3 + 1/3 * 1/4 + 1/3 * 1/7 + [1/3 * 1 + 1/3 * 1])/5 \simeq 0.02 \text{ for sense 3.}$$

The weights resulting from the extra synsets are represented within square brackets. Since the maximum weight is obtained for the first sense, this is the sense assigned to *academic*.

The sense disambiguation of a verb is done nearly in the same way, but taking into consideration only the MDWs with the verb’s senses and the context words (i.e., in the previous example, if we had to disambiguate a verb instead of an adjective, the weights within the square brackets would not have been considered). In the all-words and the gloss disambiguation tasks the two context words are the noun before and after the verb, whereas in the lexical sample task the context words are four (two before and two after the verb), without regard to their morphological category. This has been done in order to improve the recall in the latter task, whose test corpus is made up mostly by verbs, since our experiments carried out over the SemCor corpus showed that considering only the noun preceding and following the verb allows for achieving a better precision, while the recall is higher when the 4-word context is used.

The sense disambiguation of adverbs (in every task) is carried out in the same way of the disambiguation of verbs for the lexical sample task. We are still working on the disambiguation of adverbs, however, by the time we participated in SENSEVAL-3, this was the method providing the best results.

3 The English All-Words Task

We participated in this task with two systems: the *upv-unige-CIAOSENSO-eaw* system and the *upv-unige-CIAOSENSO2-eaw* system. The difference

between these systems is that in the latter the disambiguation of nouns is carried out considering only the densities of the subhierarchies obtained with the formula (3), while the first one considers the WordNet Domains weights, too. The nouns have been disambiguated in both systems with a context window of four nouns. The disambiguation of verbs, as said above, has been carried out considering the noun preceding and following the verb. Adverbs have been disambiguated with a context window of four words, while adjectives have been disambiguated with the Closest Noun, as described in the previous section.

The text, for every task we participated in, has been previously POS-tagged with the POS-tagger described in (Pla and Molina, 2001). In the tables below we show the results achieved by the upv-unige-CIAOSENSO and upv-unige-CIAOSENSO2 systems in the SENSEVAL-3. The table 1 shows the “without U” scores, which consider the missing answers as undisambiguated words and not errors (that is, how our system is intended to work). The

	CIAOSENSO	CIAOSENSO2
Precision	.581	.608
Recall	.480	.451
Coverage	84.27%	75.79%

Table 1: Results for the upv-unige-CIAOSENSO and upv-unige-CIAOSENSO2 in the english all-words task (w/o U).

baseline *MFU*, calculated by assigning to the word its most frequent (according to WordNet) sense, is .605 for both precision and recall, having a 100% coverage.

The results are roughly comparable with those obtained in our previous work over the SemCor. Considering only the polysemous words in SemCor, our tests gave a precision of .608 and a recall of .508, with a coverage of 83.55% (if monosemous words were included, the values for precision and recall would be, respectively, 0.692 and 0.602, with a coverage of 87.07%). In order to have a better understanding of the results, in the following two tables we show the precision and recall results for each morphological category, highlighting those on nouns, being the only category for which the two systems give different answers.

The behaviour of our systems is the same as we observed on the SemCor: the system relying only on Conceptual Density and frequency is more precise, even more than the most-frequent heuristic (over nouns in SemCor the precision obtained by the CIAOSENSO and CIAOSENSO2 systems was, re-

	CIAOSENSO	CIAOSENSO2	MFU
P	.665	.753	.701
R	.576	.512	.701

Table 2: Precision(P) and recall(R) results obtained by the upv-unige-CIAOSENSO and upv-unige-CIAOSENSO2, for the disambiguation of nouns, in the english all-words task (w/o U).

	Precision	Recall	MFU
Adjectives	.670	.169	.654
Verbs	.451	.340	.523
Adverbs	1.000	1.000	1.000

Table 3: Precision and Recall of the upv-unige-CIAOSENSO systems, grouped by morphological category, in the english all-words task (w/o U).

spectively, 0.737 and 0.815, with a MFU baseline of .755). Whereas the precision needs to be improved over verbs, it overtakes the baseline for nouns and adjectives.

4 The English Lexical Sample Task

The system participating in this task works in an almost identical manner of the *upv-unige-CIAOSENSO-eaw*, with the difference that verbs are disambiguated in the same way of adverbs (context of four words, the two preceding and the two following the verb). The biggest difference with the all-words task is that the training corpus has been used to change the ranking of WordNet senses for the headwords, therefore, it should be more appropriate to consider this version of the upv-unige-CIAOSENSO as a hybrid system. E.g. in the training corpus the verb *mean*, having seven senses in WordNet, appears 40 times with the WordNet sixth sense, 23 times with the WN second sense, and eight times with the WN seventh sense; therefore, the ranking of its senses has been changed to the following: 6 2 7 1 3 4 5. In table 5 we show the POS-specific results from the total ones, in order to highlight the superior performance over nouns.

	Coarse-grained	Fine-grained
Precision	.591	.501
Recall	.493	.417
Coverage	83.39%	83.39%

Table 4: Coarse and fine-grained scores for the upv-unige-CIAOSENSO-Is system in the english lexical sample task.

	Nouns	Adjectives	Verbs
Precision	.612	.589	.568
Recall	.552	.433	.443
Coverage	90.26%	73.58%	77.90%

Table 5: POS-specific results (coarse-grained) for the upv-unige-CIAOSENSE-1s system in the english lexical sample task.

5 The WSD of WordNet Glosses Task

The *upv-unige-CIAOSENSE-gl* system is an optimized version for this task, of the *upv-unige-CIAOSENSE2-eaw* which participated in the all-words task. The optimization has been done on the basis of the work we carried out over WordNet glosses during the testing of the disambiguation of adjectives over the SemCor corpus. During that work, we tried to extract from adjective glosses the nouns to be used to calculate additional MDWs, and we obtained a precision of 61.11% for the adjectives in the whole SemCor using the disambiguated glosses, against a 57.10% of precision with the undisambiguated glosses.

This improvement led us to further investigate the structure of wordnet glosses, investigation that took us to apply the following “corrections” to the original system for the SENSEVAL-3 gloss disambiguation task. First of all, it has been noted that noun glosses often contains references to the direct hypernym and/or the direct hyponyms (e.g. *command(1)* in the gloss of *behest*: “an authoritative command or request”), and its meronyms and holonyms too (e.g. *jaw(3)* in the gloss of *chuck(3)*: “a holding device consisting of adjustable jaws...”). Therefore, we added a weight of 0.7 for the noun senses being direct hypernyms, or direct hyponyms, of the synset to which belongs the gloss (*head* synset), and 0.5 for the senses being meronyms or holonyms of the head synset. Then, it has been noted that verb glosses often contains references to the direct hypernym (e.g. *walk(1)* in the gloss of *flounce*: “walk emphatically”), thus a weight of 0.5 is added for the verb senses being direct hypernym of the head verb synset. A weight 0.5 is also added when an attribute or pertainymy relationship with the head synset is found. Finally, we used WordNet Domains to assign extra weights to the senses having the same domain of the head synset (e.g. *heart(2)* in the gloss of *blood(1)*: “the fluid that is pumped by the heart”). The assigned weight is 1.0 if the domain is different than *Factotum*, 0 otherwise. E.g. *blood(1)* belongs to the domain *Medicine*; of the ten senses of *heart* in WordNet, only the second is in

the domain *Medicine*, therefore the second sense of *heart* gets a weight of 1.0 (we gave intentionally an higher weight than the other relationships because it seemed to us more meaningful than the other ones).

Although we participated in this task only with the optimized version, we tried to use the standard system for the same task in order to see the difference between them. The results show that the optimized version performs much better for the gloss disambiguation task than the standard one:

	Optimized	Standard
Precision	.534	.514
Recall	.405	.364
Coverage	76.0%	70.7%

Table 6: Comparison of optimized (*upv-unige-CIAOSENSE-gl*) and standard versions of the CIAOSENSE WSD system in the WordNet gloss disambiguation task.

6 Conclusions and Further Work

The results we obtained in the three tasks of the SENSEVAL-3 we participated in are roughly comparable with those attained in our previous work over the SemCor. In other words, it seems that our system better disambiguate nouns in comparison to words of the others morphological categories.

A further research direction we plan to investigate is the role of WordNet glosses in the disambiguation, by using the Web as resource to retrieve additional sample sentences, in order to integrate a “leskian” approach within our system. We aim to enhance the performance over verbs, that is the morphological category in which we are facing most difficulty.

We also took part in the english all-words and english lexical sample tasks in the integrated R2D2-Team system, together with other (un)supervised methods based on Maximum Entropy and Hidden Markov Models, obtaining the following results:

	EAW	LS(coarse)	LS (fine)
Precision	.626	.697	.634
Recall	.626	.573	.521
Coverage	100.0%	82.12%	82.12%

Table 7: Results of the R2D2-Team system. EAW: english all-words task, scores are both with U and w/o U. LS: Lexical Sample task.

The integration has been made by means of a voting technique. We plan to improve the integration by assigning a certain weight to each system.

Acknowledgements

This work was supported by the CIAO SENSO MCYT Spain-Italy project (HI 2002-0140) and by the R2D2 CICYT project (TIC2003-07158-C04-03). We are grateful to A. Molina and F. Pla for making the POS-tagger available.

References

- Eneko Agirre and German Rigau. 1995. A proposal for Word Sense Disambiguation using Conceptual Distance. *Proceedings of the International Conference on Recent Advances in NLP*, (RANLP).
- Bernardo Magnini and Gabriela Cavagli`a. 2000. Integrating Subject Field Codes into WordNet. *Proceedings of LREC-2000, Second International Conference on Language Resources and Evaluation*, pp. 1413-1418.
- Ferran Pla and Antonio Molina. 2001. Part-Of-Speech Tagging with Lexicalized HMM. *Proceedings of International Conference on Recent Advances in NLP*, (RANLP).
- Paolo Rosso, Francesco Masulli, Davide Buscaldi, Ferran Pla, Antonio Molina. 2003. Automatic Noun Disambiguation. *Lecture Notes in Computer Science, Springer Verlag*, (2588):273-276.