Combining Unsupervised Lexical Knowledge Methods for Word Sense Disambiguation *

German Rigau, Jordi Atserias
Dept. de Llenguatges i Sist. Informàtics
Universitat Politècnica de Catalunya
Barcelona, Catalonia
{g.rigau,batalla}@lsi.upc.es

Eneko Agirre
Lengoaia eta Sist. Informatikoak saila
Euskal Herriko Unibertsitatea
Donostia, Basque Country
jibagbee@si.ehu.es

Abstract

This paper presents a method to combine a set of unsupervised algorithms that can accurately disambiguate word senses in a large, completely untagged corpus. Although most of the techniques for word sense resolution have been presented as stand-alone, it is our belief that full-fledged lexical ambiguity resolution should combine several information sources and techniques. The set of techniques have been applied in a combined way to disambiguate the genus terms of two machine-readable dictionaries (MRD), enabling us to construct complete taxonomies for Spanish and French. Tested accuracy is above 80% overall and 95% for two-way ambiguous genus terms, showing that taxonomy building is not limited to structured dictionaries such as LDOCE.

1 Introduction

While in English the "lexical bottleneck" problem (Briscoe, 1991) seems to be softened (e.g. WordNet (Miller, 1990), Alvey Lexicon (Grover et al., 1993), COMLEX (Grishman et al., 1994), etc.) there are no available wide range lexicons for natural language processing (NLP) for other languages. Manual construction of lexicons is the most reliable technique for obtaining structured lexicons but is costly and highly time-consuming. This is the reason for many researchers having focused on the massive acquisition of lexical knowledge and semantic information from pre-existing structured lexical resources as automatically as possible.

As dictionaries are special texts whose subject matter is a language (or a pair of languages in the case of bilingual dictionaries) they provide a wide range of information about words by giving definitions of senses of words, and, doing that, supplying knowledge not just about language, but about the world itself.

One of the most important relation to be extracted from machine-readable dictionaries (MRD) is the hyponym/hypernym relation among dictionary senses (e.g. (Amsler, 1981), (Vossen and Serail, 1990)) not only because of its own importance as the backbone of taxonomies, but also because this relation acts as the support of main inheritance mechanisms helping, thus, the acquisition of other relations and semantic features (Cohen and Loiselle, 1988), providing formal structure and avoiding redundancy in the lexicon (Briscoe et al., 1990). For instance, following the natural chain of dictionary senses described in the *Diccionario General Ilustrado de la Lengua Española* (DGILE, 1987) we can discover that a *bonsai* is a cultivated plant or bush.

bonsai_1_2 *planta y arbusto así cultivado.*
(bonsai, plant and bush cultivated in that way)

The hyponym/hypernym relation appears between the entry word (e.g. *bonsai*) and the genus term, or the core of the phrase (e.g. *planta* and *arbusto*). Thus, usually a dictionary definition is written to employ a genus term combined with differentia which distinguishes the word being defined from other words with the same genus term[1].

As lexical ambiguity pervades language in texts, the words used in dictionary are themselves lexically ambiguous. Thus, when constructing complete disambiguated taxonomies, the correct dictionary sense of the genus term must be selected in each dictionary

*This research has been partially funded by CICYT TIC96-1243-C03-02 (ITEM project) and the European Comission LE-4003 (EuroWordNet project).

[1] For other kind of definition patterns not based on genus, a genus-like term was added after studying those patterns.

	DGILE		LPPL	
	overall	nouns	overall	nouns
headwords	93,484	53,799	15,953	10,506
senses	168,779	93,275	22,899	13,740
total number of words	1,227,380	903,163	97,778	66,323
average length of definition	7.26	9.68	3.27	3.82

Table 1: Dictionary Data

definition, performing what is usually called Word Sense Disambiguation (WSD)[2]. In the previous example *planta* has thirteen senses and *arbusto* only one.

Although a large set of dictionaries have been exploited as lexical resources, the most widely used monolingual MRD for NLP is LDOCE which was designed for learners of English. It is clear that different dictionaries do not contain the same explicit information. The information placed in LDOCE has allowed to extract other implicit information easily, e.g. taxonomies (Bruce et al., 1992). Does it mean that only highly structured dictionaries like LDOCE are suitable to be exploited to provide lexical resources for NLP systems?

We explored this question probing two disparate dictionaries: *Diccionario General Ilustrado de la Lengua Española* (DGILE, 1987) for Spanish, and *Le Plus Petit Larousse* (LPPL, 1980) for French. Both are substantially poorer in coded information than LDOCE (LDOCE, 1987)[3]. These dictionaries are very different in number of headwords, polysemy degree, size and length of definitions (c.f. table 1). While DGILE is a good example of a large sized dictionary, LPPL shows to what extent the smallest dictionary is useful.

Even if most of the techniques for WSD are presented as stand-alone, it is our belief, following the ideas of (McRoy, 1992), that full-fledged lexical ambiguity resolution should combine several information sources and techniques. This work does not address all the heuristics cited in her paper, but profits from techniques that were at hand, without any claim of them being complete. In fact we use unsupervised techniques, i.e. those that do not require hand-coding of any kind, that draw knowledge from a variety of sources – the source dictionaries, bilingual dictionaries and WordNet – in diverse ways.

[2]Called also Lexical Ambiguity Resolution, Word Sense Discrimination, Word Sense Selection or Word Sense Identification.

[3]In LDOCE, dictionary senses are explicitly ordered by frequency, 86% dictionary senses have semantic codes and 44% of dictionary senses have pragmatic codes.

This paper tries to proof that using an appropriate method to combine those heuristics we can disambiguate the genus terms with reasonable precision, and thus construct complete taxonomies from any conventional dictionary in any language.

This paper is organized as follows. After this short introduction, section 2 shows the methods we have applied. Section 3 describes the test sets and shows the results. Section 4 explains the construction of the lexical knowledge resources used. Section 5 discusses previous work, and finally, section 6 faces some conclusions and comments on future work.

2 Heuristics for Genus Sense Disambiguation

As the methods described in this paper have been developed for being applied in a combined way, each one must be seen as a container of some part of the knowledge (or heuristic) needed to disambiguate the correct hypernym sense. Not all the heuristics are suitable to be applied to all definitions. For combining the heuristics, each heuristic assigns each candidate hypernym sense a normalized weight, i.e. a real number ranging from 0 to 1 (after a scaling process, where maximum score is assigned 1, c.f. section 2.9). The heuristics applied range from the simplest (e.g. heuristic 1, 2, 3 and 4) to the most informed ones (e.g. heuristics 5, 6, 7 and 8), and use information present in the entries under study (e.g. heuristics 1, 2, 3 and 4) or extracted from the whole dictionary as a unique lexical knowledge resource (e.g. heuristics 5 and 6) or combining lexical knowledge from several heterogeneous lexical resources (e.g. heuristic 7 and 8).

2.1 Heuristic 1: Monosemous Genus Term

This heuristic is applied when the genus term is monosemous. As there is only one hypernym sense candidate, the hyponym sense is attached to it. Only 12% of noun dictionary senses have monosemous genus terms in DGILE, whereas the smaller LPPL reaches 40%.

2.2 Heuristic 2: Entry Sense Ordering

This heuristic assumes that senses are ordered in an entry by frequency of usage. That is, the most used and important senses are placed in the entry before less frequent or less important ones. This heuristic provides the maximum score to the first sense of the hypernym candidates and decreasing scores to the others.

2.3 Heuristic 3: Explicit Semantic Domain

This heuristic assigns the maximum score to the hypernym sense which has the same semantic domain tag as the hyponym. This heuristic is of limited application: LPPL lacks semantic tags, and less than 10% of the definitions in DGILE are marked with one of the 96 different semantic domain tags (e.g. *med.* for medicine, or *der.* for law, etc.).

2.4 Heuristic 4: Word Matching

This heuristic trusts that related concepts will be expressed using the same content words. Given two definitions – that of the hyponym and that of one candidate hypernym – this heuristic computes the total amount of content words shared (including headwords). Due to the morphological productivity of Spanish and French, we have considered different variants of this heuristic. For LPPL the match among lemmas proved most useful, while DGILE yielded better results when matching the first four characters of words.

2.5 Heuristic 5: Simple Cooccurrence

This heuristic uses cooccurrence data collected from the whole dictionary (see section 4.1 for more details). Thus, given a hyponym definition (O) and a set of candidate hypernym definitions, this method selects the candidate hypernym definition (E) which returns the maximum score given by formula (1):

$$SC(O,E) = \sum_{w_i \in O \wedge w_j \in E} cw(w_i, w_j) \qquad (1)$$

The cooccurrence weight (cw) between two words can be given by Cooccurrence Frequency, Mutual Information (Church and Hanks, 1990) or Association Ratio (Resnik, 1992). We tested them using different context window sizes. Best results were obtained in both dictionaries using the Association Ratio. In DGILE window size 7 proved the most suitable, whereas in LPPL whole definitions were used.

2.6 Heuristic 6: Cooccurrence Vectors

This heuristic is based on the method presented in (Wilks et al., 1993) which also uses cooccurrence data collected from the whole dictionary (c.f. section 4.1). Given a hyponym definition (O) and a set of candidate hypernym definitions, this method selects the candidate hypernym (E) which returns the maximum score following formula (2):

$$CV(O,E) = sim(V_O, V_E) \qquad (2)$$

The similarity (sim) between two definitions can be measured by the dot product, the cosine function or the Euclidean distance between two vectors (V_O and V_E) which represent the contexts of the words presented in the respective definitions following formula (3):

$$V_{Def} = \sum_{w_i \in Def} civ(w_i) \qquad (3)$$

The vector for a definition (V_{Def}) is computed adding the cooccurrence information vectors of the words in the definition ($civ(w_i)$). The cooccurrence information vector for a word is collected from the whole dictionary using Cooccurrence Frequency, Mutual Information or Association Ratio. The best combination for each dictionary vary: whereas the dot product, Association Ratio, and window size 7 proved best for DGILE, the cosine, Mutual Information and whole definitions were preferred for LPPL.

2.7 Heuristic 7: Semantic Vectors

Because both LPPL and DGILE are poorly semantically coded we decided to enrich the dictionary assigning automatically a semantic tag to each dictionary sense (see section 4.2 for more details). Instead of assigning only one tag we can attach to each dictionary sense a vector with weights for each of the 25 semantic tags we considered (which correspond to the 25 lexicographer files of WordNet (Miller, 1990)). In this case, given an hyponym (O) and a set of possible hypernyms we select the candidate hypernym (E) which yields maximum similarity among semantic vectors:

$$SV(O,E) = sim(V_O, V_E) \qquad (4)$$

where sim can be the dot product, cosine or Euclidean Distance, as before. Each dictionary sense has been semantically tagged with a vector of semantic weights following formula (5).

$$V_{Def} = \sum_{w_i \in Def} swv(w_i) \qquad (5)$$

The salient word vector (swv) for a word contains a saliency weight (Yarowsky, 1992) for each of the 25 semantic tags of WordNet. Again, the best method differs from one dictionary to the other: each one prefers the method used in the previous section.

2.8 Heuristic 8: Conceptual Distance

Conceptual distance provides a basis for determining closeness in meaning among words, taking as reference a structured hierarchical net. Conceptual distance between two concepts is essentially the length of the shortest path that connects the concepts in the hierarchy. In order to apply conceptual distance, WordNet was chosen as the hierarchical knowledge base, and bilingual dictionaries were used to link Spanish and French words to the English concepts.

Given a hyponym definition (O) and a set of candidate hypernym definitions, this heuristic chooses the hypernym definition (E) which is closest according to the following formula:

$$CD(O, E) = dist(headword_O, genus_E) \tag{6}$$

That is, Conceptual Distance is measured between the headword of the hyponym definition and the genus of the candidate hypernym definitions using formula (7), c.f. (Agirre et al., 1994). To compute the distance between any two words (w_1,w_2), all the corresponding concepts in WordNet (c_{1_i}, c_{2_j}) are searched via a bilingual dictionary, and the minimum of the summatory for each concept in the path between each possible combination of $c_{1_i}$and c_{2_j} is returned, as shown below:

$$dist(w_1, w_2) = \min_{\substack{c_{1_i} \in w_1 \\ c_{2_j} \in w_2}} \sum_{\substack{c_k \in \\ path(c_{1_i}, c_{2_j})}} \frac{1}{depth(c_k)} \tag{7}$$

Formulas (6) and (7) proved the most suitable of several other possibilities for this task, including those which included full definitions in (6) or those using other Conceptual Distance formulas, c.f. (Agirre and Rigau, 1996).

2.9 Combining the heuristics: Summing

As outlined in the beginning of this section, the way to combine all the heuristics in one single decision is simple. The weights each heuristic assigns to the rivaling senses of one genus are normalized to the interval between 1 (best weight) and 0. Formula (8) shows the normalized value a given heuristic will give to sense E of the genus, according to the weight assigned to the heuristic to sense E and the maximum weight of all the sense of the genus E_i.

$$vote(O, E) = \frac{weight(O, E)}{\max_{E_i}(weigth(O, E_i))} \tag{8}$$

The values thus collected from each heuristic, are added up for each competing sense. The order in which the heuristics are applied has no relevance at all.

	DGILE	LPPL
Test Sampling	391	115
Correct Genus Selected	382 (98%)	111 (97%)
Monosemous	61 (16%)	40 (36%)
Senses per genus	2.75	2.29
idem (polysemous only)	3.64	3.02
Correct senses per genus	1.38	1.05
idem (polysemous only)	1.51	1.06

Table 2: Test Sets

3 Evaluation

3.1 Test Set

In order to test the performance of each heuristic and their combination, we selected two test sets at random (one per dictionary): 391 noun senses for DGILE and 115 noun senses for LPPL, which give confidence rates of 95% and 91% respectively. From these samples, we retained only those for which the automatic selection process selected the correct genus (more than 97% in both dictionaries). Both test sets were disambiguated by hand. Where necessary multiple correct senses were allowed in both dictionaries. Table 2 shows the data for the test sets.

3.2 Results

Table 3 summarizes the results for polysemous genus.

In general, the results obtained for each heuristic seem to be poor, but always over the random choice baseline (also shown in tables 3 and 4). The best heuristics according to the recall in both dictionaries is the sense ordering heuristic (2). For the rest, the difference in size of the dictionaries could explain the reason why cooccurrence-based heuristics (5 and 6) are the best for DGILE, and the worst for LPPL. Semantic distance gives the best precision for LPPL, but chooses an average of 1.25 senses for each genus.

With the combination of the heuristics (Sum) we obtained an improvement over sense ordering (heuristic 2) of 9% (from 70% to 79%) in DGILE, and of 7% (from 66% to 73%) in LPPL, maintaining in both cases a coverage of 100%. Including monosemous genus in the results (c.f. table 4), the sum is able to correctly disambiguate 83% of the genus in DGILE (8% improvement over sense ordering) and 82% of the genus in LPPL (4% improvement). Note that we are adding the results of eight different heuristics with eight different performances, improving the individual performance of each one.

In order to test the contribution of each heuristic to the total knowledge, we tested the sum of all the heuristics, eliminating one of them in turn. The results are provided in table 5.

LPPL	random	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	Sum
recall	36%	-	66%	-	8%	11%	22%	11%	50%	73%
precision	36%	-	66%	-	66%	44%	61%	57%	76%	73%
coverage	100%	-	100%	-	12%	25%	36%	19%	66%	100%
DGILE										
recall	30%	-	70%	1%	44%	57%	60%	57%	47%	79%
precision	30%	-	70%	100%	72%	57%	60%	58%	49%	79%
coverage	100%	-	100%	1%	61%	100%	100%	99%	95%	100%

Table 3: Results for polysemous genus.

LPPL	random	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	Sum
recall	59%	35%	78%	-	40%	42%	50%	42%	68%	82%
precision	59%	100%	78%	-	93%	82%	84%	88%	87%	82%
coverage	100%	35%	100%	-	43%	51%	59%	48%	78%	100%
DGILE										
recall	41%	16%	75%	2%	41%	59%	63%	59%	48%	83%
precision	41%	100%	75%	100%	79%	65%	66%	63%	57%	83%
coverage	100%	16%	100%	2%	56%	95%	97%	94%	89%	100%

Table 4: Overall results.

LPPL	Sum	-(1)	-(2)	-(3)	-(4)	-(5)	-(6)	-(7)	-(8)
recall	82%	73%	74%	-	73%	76%	77%	77%	78%
precision	82%	73%	75%	-	73%	76%	77%	77%	78%
coverage	100%	100%	99%	-	100%	100%	100%	100%	100%
DGILE									
recall	83%	79%	72%	81%	81%	81%	81%	81%	77%
precision	83%	79%	72%	82%	81%	81%	81%	81%	77%
coverage	100%	100%	100%	98%	100%	100%	100%	100%	100%

Table 5: Knowledge provided by each heuristic (overall results).

(Gale et al., 1993) estimate that any sense-identification system that does not give the correct sense of polysemous words more than 75% of the time would not be worth serious consideration. As table 5 shows this is not the case in our system. For instance, in DGILE heuristic 8 has the worst performance (see table 4, precision 57%), but it has the second larger contribution (see table 5, precision decreases from 83% to 77%). That is, even those heuristics with poor performance can contribute with knowledge that other heuristics do not provide.

3.3 Evaluation

The difference in performance between the two dictionaries show that quality and size of resources is a key issue. Apparently the task of disambiguating LPPL seems easier: less polysemy, more monosemous genus and high precision of the sense ordering heuristic. However, the heuristics that depend only on the size of the data (5, 6) perform poorly on LPPL, while they are powerful methods for DGILE.

The results show that the combination of heuristics is useful, even if the performance of some of the heuristics is low. The combination performs better than isolated heuristics, and allows to disambiguate all the genus of the test set with a success rate of 83% in DGILE and 82% in LPPL.

All the heuristics except heuristic 3 can readily be applied to any other dictionary. Minimal parameter adjustment (window size, cooccurrence weigth formula and vector similarity function) should be done to fit the characteristics of the dictionary, but according to our results it does not alter significantly the results after combining the heuristics.

4 Derived Lexical Knowledge Resources

4.1 Cooccurrence Data

Following (Wilks et al., 1993) two words cooccur if they appear in the same definition (word order in definitions are not taken into account). For instance, for DGILE, a lexicon of 300,062 cooccurrence pairs among 40,193 word forms was derived (stop words were not taken into account). Table 6 shows the first eleven words out of the 360 which cooccur with *vino* (wine) ordered by Association Ratio. From left to right, Association Ratio and number of occurrences.

The lexicon (or machine-tractable dictionary,

AR	#oc.	
11.1655	15	*tinto* (red)
10.0162	23	*beber* (to drink)
9.6627	14	*mosto* (must)
8.6633	9	*jerez* (sherry)
8.1051	9	*cubas* (cask, barrel)
8.0551	16	*licor* (liquor)
7.2127	17	*bebida* (drink)
6.9338	12	*uva* (grape)
6.8436	9	*trago* (drink, swig)
6.6221	12	*sabor* (taste)
6.4506	15	*pan* (bread)

Table 6: Example of association ratio for *vino* (wine).

MTD) thus produced from the dictionary is used by heuristics 5 and 6.

4.2 Multilingual Data

Heuristics 7 and 8 need external knowledge, not present in the dictionaries themselves. This knowledge is composed of semantic field tags and hierarchical structures, and both were extracted from WordNet. In order to do this, the gap between our working languages and English was filled with two bilingual dictionaries. For this purpose, we derived a list of links for each word in Spanish and French as follows.

Firstly, each Spanish or French word was looked up in the bilingual dictionary, and its English translation was found. For each translation WordNet yielded its senses, in the form of WordNet concepts (synsets). The pair made of the original word and each of the concepts linked to it, was included in a file, thus producing a MTD with links between Spanish or French words and WordNet concepts. Obviously some of this links are not correct, as the translation in the bilingual dictionary may not necessarily be understood in its senses (as listed in WordNet). The heuristics using these MTDs are aware of this.

For instance when accessing the semantic fields for *vin* (French) we get a unique translation, wine, which has two senses in WordNet: <wine,vino> as a beverage, and <wine, wine-coloured> as a kind of color. In this example two links would be produced (*vin*, <wine,vino>) and (*vin*, <wine, wine-coloured>). This link allows us to get two possible semantic fields for *vin* (noun.food, file 13, and noun.attribute, file 7) and the whole structure of the hierarchy in WordNet for each of the concepts.

5 Comparison with Previous Work

Several approaches have been proposed for attaching the correct sense (from a set of prescribed ones) of a word in context. Some of them have been fully tested in real size texts (e.g. statistical methods (Yarowsky, 1992), (Yarowsky, 1994), (Miller and Teibel, 1991), knowledge based methods (Sussna, 1993), (Agirre and Rigau, 1996), or mixed methods (Richardson et al., 1994), (Resnik, 1995)). The performance of WSD is reaching a high stance, although usually only small sets of words with clear sense distinctions are selected for disambiguation (e.g. (Yarowsky, 1995) reports a success rate of 96% disambiguating twelve words with two clear sense distinctions each one).

This paper has presented a general technique for WSD which is a combination of statistical and knowledge based methods, and which has been applied to disambiguate all the genus terms in two dictionaries.

Although this latter task could be seen easier than general WSD[4], genus are usually frequent and general words with high ambiguity[5]. While the average of senses per noun in DGILE is 1.8 the average of senses per noun genus is 2.75 (1.30 and 2.29 respectively for LPPL). Furthermore, it is not possible to apply the powerful "one sense per discourse" property (Yarowsky, 1995) because there is no discourse in dictionaries.

WSD is a very difficult task even for humans[6], but semiautomatic techniques to disambiguate genus have been broadly used (Amsler, 1981) (Vossen and Serail, 1990) (Ageno et al., 1992) (Artola, 1993) and some attempts to do automatic genus disambiguation have been performed using the semantic codes of the dictionary (Bruce et al., 1992) or using cooccurrence data extracted from the dictionary itself (Wilks et al., 1993).

Selecting the correct sense for LDOCE genus terms, (Bruce et al., 1992)) report a success rate of 80% (90% after hand coding of ten genus). This impressive rate is achieved using the intrinsic char-

[4] In contrast to other sense distinctions Dictionary word senses frequently differ in subtle distinctions (only some of which have to do with meaning (Gale et al., 1993)) producing a large set of closely related dictionary senses (Jacobs, 1991).

[5] However, in dictionary definitions the headword and the genus term have to be the same part of speech.

[6] (Wilks et al., 1993) disambiguating 197 occurrences of the word bank in LDOCE say "was not an easy task, as some of the usages of bank did not seem to fit any of the definitions very well". Also (Miller et al., 1994) tagging semantically SemCor by hand, measure an error rate around 10% for polysemous words.

acteristics of LDOCE. Furthermore, using only the implicit information contained into the dictionary definitions of LDOCE (Cowie et al., 1992) report a success rate of 47% at a sense level. (Wilks et al., 1993) reports a success rate of 45% disambiguating the word bank (thirteen senses LDOCE) using a technique similar to heuristic 6. In our case, combining informed heuristics and without explicit semantic tags, the success rates are 83% and 82% overall, and 95% and 75% for two-way ambiguous genus (DGILE and LPPL data, respectively). Moreover, 93% and 92% of times the real solution is between the first and second proposed solution.

6 Conclusion and Future Work

The results show that computer aided construction of taxonomies using lexical resources is not limited to highly-structured dictionaries as LDOCE, but has been succesfully achieved with two very different dictionaries. All the heuristics used are unsupervised, in the sense that they do not need hand-codding of any kind, and the proposed method can be adapted to any dictionary with minimal parameter setting.

Nevertheless, quality and size of the lexical knowledge resources are important. As the results for LPPL show, small dictionaries with short definitions can not profit from raw corpus techniques (heuristics 5, 6), and consequently the improvement of precision over the random baseline or first-sense heuristic is lower than in DGILE.

We have also shown that such a simple technique as just summing is a useful way to combine knowledge from several unsupervised WSD methods, allowing to raise the performance of each one in isolation (coverage and/or precision). Furthermore, even those heuristics with apparently poor results provide knowledge to the final result not provided by the rest of heuristics. Thus, adding new heuristics with different methodologies and different knowledge (e.g. from corpora) as they become available will certainly improve the results.

Needless to say, several improvements can be done both in individual heuristic and also in the method to combine them. For instance, the cooccurrence heuristics have been applied quite indiscriminately, even in low frequency conditions. Significance tests or association coefficients could be used in order to discard low confidence decisions. Also, instead of just summing, more clever combinations can be tried, such as training classifiers which use the heuristics as predictor variables.

Although we used these techniques for genus disambiguation we expect similar results (or even better taken the "one sense per discourse" property and lexical knowledge acquired from corpora) for the WSD problem.

7 Acknowledgments

This work would not be possible without the collaboration of our colleagues, specially Jose Mari Arriola, Xabier Artola, Arantza Diaz de Ilarraza, Kepa Sarasola and Aitor Soroa in the Basque Country and Horacio Rodríguez in Catalonia.

References

[Ageno et al.1992] Alicia Ageno, Irene Castellón, Maria Antonia Martí, Francesc Ribas, German Rigau, Horacio Rodríguez, Mariona Taulé and Felisa Verdejo. 1992. SEISD: An environment for extraction of Semantic information from on-line dictionaries. In *Proceedings of the 3th Conference on Applied Natural Language Processing (ANLP'92)*, Trento, Italy.

[Agirre et al.1994] Eneko Agirre, Xabier Arregi, Xabier Artola, Arantza Díaz de Ilarraza and Kepa Sarasola. 1994. Conceptual Distance and Automatic Spelling Correction. In *Proceedings of the workshop on Computational Linguistics for Speech and Handwriting Recognition*, Leeds, United Kingdom.

[Agirre and Rigau1996] Eneko Agirre and German Rigau. 1996. Word Sense Disambiguation using Conceptual Density. In *Proceedings of the 16th International Conference on Computational Linguistics (Coling'96)*, pages 16–22. Copenhagen, Denmark.

[Amsler1981] Robert Amsler. 1981. A Taxonomy for English Nouns and Verbs. In *Proceedings of the 19th Annual Meeting of the Association for Computational Linguistics*, pages 133–138. Stanford, California.

[Artola1993] Xabier Artola. 1993. *Conception et construccion d'un systeme intelligent d'aide diccionariale (SIAD).* PhD. Thesis, Euskal Herriko Unibertsitatea, Donostia, Basque Country.

[Briscoe et al.1990] Eduard Briscoe, Ann Copestake and Branimir Boguraev. 1990. Enjoy the paper: Lexical Semantics via lexicology. In *Proceedings of the 13th International Conference on Computational Linguistics (Coling'90)*, pages 42–47.

[Briscoe1991] Eduard Briscoe. 1991. Lexical Issues in Natural Language Processing. In Klein E. and Veltman F. eds. *Natural Language and Speech.* pages 39–68, Springer-Verlag.

[Bruce et al. 1992] Rebecca Bruce, Yorick Wilks, Louise Guthrie, Brian Slator and Ted Dunning. 1992. NounSense - A Disambiguated Noun Taxonomy with a Sense of Humour. *Research Report MCCS-92-246*. Computing Research Laboratory, New Mexico State University. Las Cruces.

[Church and Hanks 1990] Kenneth Church and Patrick Hanks. 1990. Word Association Norms, Mutual Information, and Lexicography. *Computational Linguistics*, vol. 16, ns. 1, 22-29.

[Cohen and Loiselle 1988] P. Cohen and C. Loiselle. 1988. Beyond ISA: Structures for Plausible Inference in Semantic Data. In *Proceedings of 7th Natural Language Conference AAAI'88.*

[Cowie et al. 1992] Jim Cowie, Joe Guthrie and Louise Guthrie. 1992. Lexical Disambiguation using Simulated Annealing. In *Proceedings of DARPA WorkShop on Speech and Natural Language*, pages 238-242, New York.

[DGILE 1987] DGILE 1987. *Diccionario General Ilustrado de la Lengua Española VOX.* Alvar M. ed. Biblograf S.A. Barcelona, Spain.

[Gale et al. 1993] William Gale, Kenneth Church and David Yarowsky. 1993. A Method for Disambiguating Word Senses in a Large Corpus. *Computers and the Humanities* 26, pages 415-439.

[Grishman et al. 1994] Ralph Grishman, Catherine Macleod and Adam Meyers. 1994.. Comlex syntax: building a computational lexicon. In *Proceedings of the 15th Annual Meeting of the Association for Computational Linguistics, (Coling'94).* 268-272. Kyoto, Japan.

[Grover et al. 1993] Claire Grover, John Carroll and John Reckers. 1993. The Alvey Natural Language Tools grammar (4th realese). *Technical Report 284.* Computer Laboratory, Cambridge University, UK.

[Jacobs 1991] Paul Jacobs. 1991. Making Sense of Lexical Acquisition. In Zernik U. ed., *Lexical Acquisition: Exploiting On-line Resources to Build a Lexicon*, Lawrence Erlbaum Associates, publishers. Hillsdale, New Jersey.

[LDOCE 1987] LDOCE 1987. *Longman Dictionary of Contemporary English.* Procter, P. ed. Longman, Harlow and London.

[LPPL 1980] LPPL 1980. *Le Plus Petit Larousse.* Gougenheim, G. ed. Librairie Larousse.

[McRoy 1992] Sussan McRoy. 1992. Using Multiple Knowledge Sources for Word Sense Discrimination. *Computational Linguistics* 18(1).

[Miller 1990] George Miller. 1990. Five papers on WordNet. *Special Issue of International Journal of Lexicography* 3(4).

[Miller and Teibel 1991] George Miller and David Teibel. 1991. A proposal for Lexical Disambiguation. In *Proceedings of DARPA Speech and Natural Language Workshop*, 395-399, Pacific Grave, California.

[Miller et al. 1994] George Miller, Martin Chodorow, Shari Landes, Claudia Leacock and Robert Thomas. 1994. Using a Semantic Concordance for sense Identification. In *Proceedings of ARPA Workshop on Human Language Technology.*

[Resnik 1992] Philip Resnik. 1992. WordNet and Distributional analysis: A class-based approach to lexical discovery. In *Proceedings of AAAI Symposyum on Probabilistic Approaches to NL*, San Jose, California.

[Resnik 1995] Philip Resnik. 1995. Disambiguating Noun Groupings with Respect to WordNet Senses. In *Proceedings of the Third Workshop on Very Large Corpora*, MIT.

[Richardson et al. 1994] R. Richardson, A.F. Smeaton and J. Murphy. 1994. Using WordNet as a Knowledge Base for Measuring Semantic Similarity between Words. *Working Paper CA-1294*, School of Computer Applications, Dublin City University. Dublin, Ireland.

[Sussna 1993] Michael Sussna. 1993. Word Sense Disambiguation for Free-text Indexing Using a Massive Semantic Network. In *Proceedings of the Second International Conference on Information and knowledge Management.* Arlington, Virginia.

[Vossen and Serail 1990] Piek Vossen and Iskander Serail. 1992. Word-Devil, a Taxonomy-Browser for Lexical Decomposition via the Lexicon. *Esprit BRA-3030 Acquilex Working Paper n. 009.*

[Wilks et al. 1993] Yorick Wilks, Dam Fass, Cheng-Ming Guo, James McDonald, Tony Plate and Brian Slator. 1993. Providing Machine Tractable Dictionary Tools. In Pustejowsky J. ed. *Semantics and the Lexicon*, pages 341–401.

[Yarowsky 1992] David Yarowsky. 1992. Word-Sense Disambiguation Using Statistical Models of Rogets Categories Trained on Large Corpora. In *Proceedings of the 14th International Conference on Computational Linguistics (Coling'92)*, pages 454-460. Nantes, France.

[Yarowsky 1994] David Yarowsky. 1994. Decision Lists for Lexical Ambiguity Resolution. In *Proceedings of the 32th Annual Meeting of the Association for Computational Linguistics, (ACL'94).* Las Cruces, New Mexico.

[Yarowsky 1995] David Yarowsky. 1995. Unsupervised Word Sense Disambiguation Rivaling Supervised Methods. In *Proceedings of the 33th Annual Meeting of the Association for Computational Linguistics, (ACL'95)*. Cambridge, Massachussets.