Evaluating Lexical Resources for WSD

M. Taulé, M. García, N. Artigas, M.A. Martí
CLiC (Centre de Llenguatge i Computació)
Department of Linguistics, Universitat de Barcelona
Gran Via de les Corts Catalanes, 585, 08007, Barcelona.
mtaule@ub.edu,{mar, nuripa}@clic.fil.ub.es, amarti@ub.edu
http://clic.fil.ub.es

Abstract
In this paper we present a methodology developed for evaluating the quality of the dictionaries for Word Sense Disambiguation (WSD) tasks. This methodology is based on the parallel tagging of a single corpus with three lexical sources (with different characteristics) and three annotators.
As far as the corpus tagging is concerned, strict criteria have been established regarding the lexicographic training of the annotators as well as the process itself, which has been developed under the guidance of a handbook specially elaborated for this task. The evaluation of the annotators' agreement degree has been automatically drawn by means of previously defined agreement measurements (total agreement, partial agreement, minimum agreement, and disagreement). As a result, two new tested resources have been obtained for the treatment of Spanish language: a lexical source, *MiniDir.2.1*, suitable for WSD tasks, and a semantically tagged Spanish corpus, *MiniCors*.

1. Introduction

A central and most difficult point that Natural Language Processing (NLP) applications have to solve is, without any doubt, ambiguity, and particularly, semantic (or lexical) ambiguity. The difficulty of its resolution is an underlying problem in many NLP applications, especially in Machine Translation and Information Retrieval systems (Ide & Véronis, 1998).
The goal of WSD systems is to assign the correct semantic interpretation to each word in a text, which basically implies the automatic identification of its sense. In order to be able to carry out the WSD task, electronic dictionaries and lexicons, and semantically tagged corpora are needed. These linguistic resources are fundamental to successfully carry out this task.
In the field of WSD two basic disambiguation techniques or strategies are developed: a) unsupervised or knowledge based systems, which make use of external explicit information resources, like machine readable dictionaries, ontologies or lexicons, and b) supervised or corpus-based systems, which apply machine learning techniques on previously semantically and manually tagged corpora (Márquez, 2002). However, in both cases the systems need information from a lexical resource to carry out the semantic tagging of the words to be analyzed.
In this sense, we consider that there are two critical points in the WSD process which have been neglected, and which are determinant when good results want to be reached: first, the quality of lexical sources and, second, the quality of the manually tagged corpora. Moreover, the quality of these corpora is determined, to a large extent, by the quality of the lexical source used for carry out the tagging process.

Our research has focused on the development of a semantically tagged Spanish corpus (*MiniCors*, (Taulé et al., 2004)), as well as on the development of a suitable lexicon (*Minidir.2.1*) for WSD tasks[1].
In this paper we present the methodology followed for the evaluation of the quality of dictionaries oriented to WSD tasks. This methodology is based on the parallel tagging of a single corpus with three lexical sources by three different annotators for each source. The annotators' agreement degree will be used for measuring the lexical source quality: the more agreement there is, the more quality the source will have. Thus, a high agreement would indicate that the senses in the lexical source are clearly defined and have a wide coverage. This methodology guarantees objectivity in the treatment of senses.
We assume the hypothesis that there is not a single way of dividing the continuum of meaning and, therefore, the underlying problem is the treatment of polysemy. It is an open question which can be enriched with the contributions made from NLP.

1.1. State of the art

In general, in WSD systems, algorithms are evaluated, while the quality of lexical sources and tagged corpora is disregarded. However, several authors have carried out studies with the aim of proposing specific models and methodologies for the elaboration of lexical sources oriented to WSD tasks.
A very outstanding proposal is Véronis' (2001), in which the validity of traditional lexical representation of senses is questioned. This author proposes a model of lexical source suitable for WSD based mainly on syntactic criteria (see section 2.1.).
Hanks (2000) proposes a model of phraseological dictionary which shows how different sides of word senses are activated depending on the context, taking into account variability and vagueness. With regard to verbs, Palmer (1998) proposes a model of lexicon, based on the generalizations on verbal classes and not on explicit lists of senses. These authors consider that contextual and syntagmatic information is fundamental for the characterization of senses.
As to the evaluation of the tagging process, some experiments have been carried out with the aim to determine which aspects are important. Kilgarriff (1999) developed an experiment on semantic tagging, with the aim to define the upper-bound in manual tagging, which has been established on the 95% of annotators' agreement. Krishnamurthy and Nichols (2000) analyze the process of the gold-standard corpus tagging for Senseval-2, highlighting the most common inconsistencies of dictionaries: incorrect sense division, definition errors, etc.
Fellbaum et al. (1997) analyze the process of semantic tagging with a lexical resource such as WordNet, but they focus on those features they consider as a source of difficulty: the lexical category, the order of the senses in the lexical source, and the annotators' profile. All the authors highlight the importance of the lexical source as an essential factor in order to obtain quality results.

2. Methodology for the Lexical Resources Evaluation

Our research consists in the evaluation of different lexical sources for WSD tasks. With this aim, we have developed a methodology which consists in the manual semantic tagging of a single corpus with three different lexical sources. The tagging process has been carried out

by different annotators. This methodology allows us to analyze comparatively the results obtained for each of the lexical sources and, therefore, to determine which of them is the most suitable for WSD tasks. Our starting point is the hypothesis that the annotator agreement degree is proportional to the quality level of the lexical resource: the more agreement there is, the more quality has the lexical source.

The evaluated lexical sources present very different characteristics. Firstly, we have used the *Diccionario de la Real Academia Española* (*DRAE*), as it is the reference and normative dictionary of Spanish language. Secondly, we have developed two lexicons designed specifically for WSD tasks: *MiniDir.2.1* (Artigas et al., 2003b) and the *Véronis Model* (Véronis 2001). These lexical sources contain a limited number of entries (49 in the *MiniDir.2.1* and 4 in the *Véronis Model*) which have been elaborated specifically for this experiment.

The comparative evaluation with the *DRAE* has been carried out over a subset of the *MiniDir.2.1* entries: 9 nouns (*arte*, *autoridad*, *circuito*, *columna*, *corazón*, *gracia*, *letra*, *naturaleza*, *pasaje*), 3 adjectives (*ciego*, *claro*, *natural*) and 8 verbs (*apoyar*, *apuntar*, *canalizar*, *explotar*, *saltar*, *tocar*, *tratar*, *volar*). Regarding the comparison with the *Veronis Model*, it has been carried out with a subset of the DRAE entries (the nouns *pasaje* and *autoridad*; the adjective *brillante* and the verb *tratar*). We have treated a reduced number of entries because of the difficulty and complexity of this lexical entry model.

2.1 The lexical sources

MiniDir.2.1 presents a set of discrete senses which are clearly distinguishable[2] in order to avoid the semantic overlapping of the traditional lexical sources. We consider that in nouns we find the most clear cases of polysemy, because is the category with the most clear referential value. In the case of adjectives, their sense is characterized by underspecification: they have potential senses which are modulated by the context, hence the importance we have given in this dictionary to the collocative information and to the low granularity in the adjectival entries[3]. As far as verbs are concerned, we have defined a set of categories which reflects the diatheses alternations related to each sense.

In the development of *MiniDir.2.1* we have basically taken into account information extracted from corpora. We have used the corpora from the newspapers *El Periódico* and *La Vanguardia*, with a total of 3.5 millions and 12.5 millions of words respectively, and *Lexesp* (Sebastián et al., 2000), a balanced corpus of 5.5 millions of words, which includes texts on different topics (science, economics, justice, literature, etc.), written in different styles (essay, novel, etc.) and different language registers (standard, technical, etc.). All these corpora are morphologically tagged and disambiguated. The corpora provide quantitative and qualitative information which is essential to differentiate senses and to determine the lexicalization degree.

Apart from the information extracted from corpora, in order to establish and to define the senses we have consulted different traditional lexical sources and two lexical conceptual knowledge bases: *WordNet 1.5* (Miller, 1995) and *EuroWordNet* (Vossen, 1999). The criteria used in the elaboration of *MiniDir.2.1* are listed in (Castelló et al., 2003).

As regards the information of the entries of the dictionary, every sense is organized in the nine following lexical fields:

LEMMA#CATEGORY#SENSE#DEFINITION#EXAMPLE#SYNONYMS#(ANTONYMS)#COLLOCATIONS#SYNSETS#

The lexical category is represented by the Eagle tags (Eureka 1989-1995) which have been abridged. In the verbal entries we have also included an additional field with a syntactic category that indicates a classification based on the diathesis alternations that the verb admits.
As regards the field of antonyms, it is only filled in the adjective entries. In the field SYNSET we have established the mapping between each sense and the *synset* number in the semantic net *EuroWordNet* (Vossen, 1999).

Below (Figure 1), an entry of this dictionary is presented, where its different fields are shown. As it is a nominal entry, the fields corresponding to verbal subcategorization and antonymy are empty:

partido#NCMS#1#Organización política cuyos miembros comparten la misma ideología
#EJ:el principal partido del país; el partido en la oposición
#SIN:?
#COL:partido centrista, partido comunista, partido conservador, partido cristianodemócrata, partido de derechas, partido de izquierdas, partido de la oposición, partido demócrata, partido ecologista, partido estatal, partido fascista, ...
#SYNSET:05259394n#

partido#NCMS#2#Prueba deportiva en la que se enfrentan dos equipos o jugadores
#EJ:partido de baloncesto; partido de tenis; el mejor partido de la temporada
#SIN:?
#COL:partido amistoso, partido de consolación, partido de desempate, ...
#SYNSET:04780657n#

Figure 1: *MiniDir 2.1.* lexical entry of *partido*

In the case of the *DRAE* entries we have adapted the format required by the semantic tagging editor (Artigas et al., 2003a). *DRAE* is a normative dictionary of Spanish language which has not been designed for any specific WSD task.
The third lexical source we have used is the dictionary developed according to the entry model proposed by Véronis (2001), which includes syntactic information (the context), paradigmatic information (hyperonymy and synonymy), and coocurrence information (collocations). The figure below (figure 2) shows an entry of this dictionary.

2.2 Tagging Process

The tagging process has been carried out by experienced lexicographers. It has been developed individually, so as to avoid interferences. Also, the authors of the dictionary have not participated in the tagging process. In order to systematize the process to the utmost, in

an initial phase we have designed a tagging handbook (Artigas et al., 2003a) in which the annotation criteria are specified.

	Info. sintax.	Info. para. sigmática	Información de concurrencias		
1	tratar (con)	portarse	X tratar Y (con Z)		
			X= entidad	**Y= entidad** persona, organización, equipo	**(Z=manera)** **sp (con)**: amabilidad, cariño, consideración, corrección, desprecio, familiaridad, indiferencia, respeto, severidad **sp(de)**: de igual a igual **adv**: cortésmente, bien, mal, así
	tratar (como)		X tratar Y como a / si Z		
			X=entidad	**Y=entidad** persona, organización, elemento, tierra	**Z=manera** **adv (como)**: ciudadanos, miembros, una colonia, sospechoso, una reina, un mentiroso, un asunto beneficioso **adv (como a)**: una extraña, uno más, una república bananera, un hijo, un delincuente, una piltrafa **adv (como si)**: hubiera escapado, fuera su novia, fuera una adulta, fueran parientes, fuese el más pequeño
2	tratar (con)	procesar con sustancia/ tratamiento	X tratar Y (con Z)		
			X=entidad depuradora, doctor, hospital, médico, planta	**Y=entidad** enfermedad, infección, síntoma, fruta, verdura, residuos, materiales, productos, persona	**(Z=sustancia/tratamiento)** medicamento, sustancia, tratamiento, fármaco, fibra, hierba, hormona, placebo, procedimientos, productos químicos, quimioterapia
3	tratar (con)	procesar información	X tratar Y (con Z)		
			X=entidad persona, empresa, ordenador	**Y=información** datos, información, bases de datos, corpus, imagen, fotografía, música, sonido,	**(Z=manera)** **sp (con)**: ordenador, informática, métodos informáticos, programa

Figure 2: *Véronis Model* entry of *tratar*

In order to systematize and simplify the annotation process, an interface has been designed specifically for this task. The interface has also been used during the annotation process, as it allows to visualize the examples according to the type of agreement and to select the sense to be assigned in cases of disagreement. The editor is language-independent and it can be used in the annotation process of other languages.

We have tagged 13,477 occurrences *with MiniDir.2.1*[4], 4,000 with *DRAE* and 800 with the *Véronis Model*. Consequently, there is a total of 800 occurrences tagged with the three lexical sources. From this tagging, a comparative study of the lexical sources has been developed.

The annotated corpus is *MiniCors*, a semantically tagged corpus where only a sample of words has been tagged (those selected in *Minidir 2.1.*). It has been obtained from the corpus of the EFE Spanish News Agency, which includes 289,066 news spanning from January to December of 2000[5].

It is, therefore, a group of sentences which belong to standard language, and, in theory, deal about general subjects and topics. The objective was to obtain 200 sentences for each of the selected words, that is, to obtain a total of 200 examples per word. The context considered for each word is larger than a sentence, as the previous and the following sentences have also been included. For each word, we tried to obtain at least 15 occurrences per sense.

<EXAMPLE IDENT="343" DOCNO="14460" DATE="2000/02/17" WFORM="**partido**" VAL="OK"><CAT SCHEME="IPTC" CODE="14000000" /><CAT SCHEME="ANPA" CODE="SOC:SOCIEDAD-SALUD,COMUNICACION" /><SENSE SCHEME="MINIDIR" CODE="1" USER="Adria" /><SENSE SCHEME="MINIDIR" CODE="1" USER="Mar" /><SENSE SCHEME="MINIDIR" CODE="1" USER="Nuripa" /><**GS SENSE="{ 1 }" ACUERDO="1.1."** /><BODY><ANT>HASH(0x863cfc0)</ANT><ACT>El PSOE presentó hoy ante la Junta Electoral Central un nuevo escrito en el que denuncia el tratamiento informativo "absolutamente hostil" de TVE contra este <TARGET>**partido**</TARGET> durante los últimos días y acompañó el texto con testimonios del Comité de Empresa de esta cadena de televisión.</ACT><SEG>En el escrito, firmado por el representante legal de PSOE-Progresistas, los socialistas acusan a TVE de "violar los principios de neutralidad informativa, proporcionalidad y pluralismo político" con la eliminación de referencias informativas a la actividad de su partido y solicita a la Junta Electoral Central que requiera a esta cadena para que cese en su actividad "hostil".</SEG></BODY></EXAMPLE>

<EXAMPLE IDENT="851" DOCNO="13110" DATE="2000/04/17" WFORM="**partido**" VAL="OK"><CAT SCHEME="IPTC" CODE="15000000" /><CAT SCHEME="ANPA" CODE="DEP:DEPORTES,FUTBOL" /><SENSE SCHEME="MINIDIR" CODE="2" USER="Adria" /><SENSE SCHEME="MINIDIR" CODE="2" USER="Mar" /><SENSE SCHEME="MINIDIR" CODE="2" USER="Nuripa" /><**GS SENSE="{ 2 }" ACUERDO="1.1."** /><BODY><ANT>Aunque superar el 3-1 adverso del partido de ida se antoja una misión difícil, no es para nada imposible, y la prensa deportiva barcelonesa se ha dedicado estos días a rememorar anteriores históricas "remontadas" del equipo azulgrana en partidos de competiciones europeas, en los que en el Camp Nou se levantaron en tres ocasiones marcadores contrarios por 3-0 en la ida.</ANT><ACT>El entrenador barcelonista, Louis van Gaal, facilitó hoy una lista de 18 convocados para el partido contra el Chelsea, en la que destaca el regreso de Patrick Kluivert, ausente en el último <TARGET>**partido**</TARGET> contra el Oviedo, castigado por unas declaraciones en las que se mostraba crítico con el técnico.</ACT><SEG>Van Gaal, aunque ha recordado varias veces que perdona pero no olvida, ha levantado el castigo en esta ocasión a su delantero holandés para intentar cerrar filas en el vestuario y presentar el mejor equipo posible ante el Chelsea.</SEG></BODY></EXAMPLE>

Figure 3: A sample of *MiniCors*, a semantic tagged corpus

Figure 3 shows the results of the semantic tagging process. As the example shows, different types of data are registered: the source corpus (<EXAMPLE IDENT="343" DOCNO="14460" DATE="2000/02/17"); the word form to be tagged ("partido"); the subject field (<CAT SCHEME = "ANPA" CODE = "SOC: SOCIEDAD-SALUD, COMUNICACION" />); the sense assigned by the three different annotators (<SENSE SCHEME="MINIDIR" CODE="1" USER="Adria" / ...>), and the final result with the annotation agreement (<GS SENSE="{ 1 }" ACUERDO="1.1.").
The text includes the previous and the following sentence for each example to be tagged, with the word to be annotated highlighted with the <TARGET> tag.
Each word has been semantically annotated by three different annotators, in order to facilitate the manual arbitration phase, which has been reduced only to cases of disagreement.

2.3 Evaluation and arbitration

Once the corpus has been tagged, we have carried out the comparison among the different annotations and the subsequent evaluation of the results in order to obtain a disambiguated corpus to begin with the evaluation of the lexical sources. Since each word has been tagged three times for each lexical source, the subsequent process of arbitration has been reduced to those cases of disagreement among annotators.
Each agreement degree receives a different tag: total agreement, partial agreement, minimum agreement, or disagreement. Total agreement takes place when the three annotations match (e.g.: 1, 1, 1 = 1). When not all the annotations match but there is a predominant annotation we get partial agreement (e.g.: 1, 1, 1/2= 1). Minimum agreement occurs when two annotations match but the other one is different (e.g.: 1, 1, 2 = 1). Finally, disagreement is produced when any of the annotations match. All the cases of agreement, either total, partial or minimum, are validated automatically according to the pattern we have previously defined. Only cases of disagreement undergo a manual arbitration phase.

We have also considered other parameters of analysis:
a) Total minimum agreement that counts all the cases of total agreement among the annotators, and the maximum total agreement, which counts the cases of total agreement and partial agreement among the annotators.
b) Pairwise agreement, which counts the degree of agreement between each pair of annotators. In this case, we have also distinguished among minimum pairwise agreement (cases of total agreement among every pair of annotators) and maximum pairwise agreement (cases of partial agreement among each pair of annotators).

The table below shows the results obtained for each dictionary for the subset of common words:

nouns	MinTA.	MaxTA.	MinA.	Dis
MiniDir.2.1	0.86	0.88	0.11	0.01
DRAE	0.51	0.76	0.23	0.01
Véronis	0.95	0.96	0.04	0.00

adjectives	MinTA.	MaxTA.	MinA.	Dis
MiniDir.2.1	0.83	0.88	0.12	0.01
DRAE	0.57	0.74	0.24	0.02
Véronis	0.99	1	0.00	0.00

verbs	MinTA.	MaxTA.	MinA.	Dis
MiniDir.2.1	0.81	0.84	0.15	0.02
DRAE	0.64	0.70	0.27	0.03
Véronis	0.95	0.95	0.04	0.00

MinTA = Minimum Total Agreement
MaxTA = Maximum Total Agreement
MinA = Minimum Agreement
Dis = Disagreement

Table 1: Agreement levels

3. Analysis of the results

The tagging experiments show that the two lexical sources which have been designed with specific criteria for WSD, *MiniDir.2.1* and the *Véronis Model*, reach higher agreement levels in the manual tagging of corpus than a lexical source of common use such as *DRAE*. Therefore, *DRAE* has been rejected as a valid source for WSD tasks, since it presents a high level of granularity. The annotators that used this source have chosen in many case multiple annotation, that is, they have assigned more than a sense for example. This fact highlights the high level of overlapping among definitions.

The annotation with the *Véronis Model* and with *MiniDir.2.1* reaches results considerably acceptable that prove their adequacy for WSD tasks. The agreement obtained both with *Minidir.2.1* and with the *Véronis model* guarantee the quality of the tagged corpora.Nevertheless, the development of entries according to the proposal of Véronis has a high development cost. Besides, this model is not suitable for the treatment of those entries which do not present clear syntactic alternations (e.g. *brillante*).

In fact, in the *Véronis Model* the highest levels of agreement have been achieved in nouns and adjectives (the difference in verbs is very small). From these results, we can conclude that both syntagmatic and cooccurence information constitute determining factors in order to

achieve higher levels of agreement. It would be useful to carry out a large-scale annotation to study if these levels of agreement are achieved.
Nevertheless, the development of the entries proposed by Véronis have a high development cost, and, therefore, it is not a feasible lexical source at short-term.
In this sense, *MiniDir.2.1* seems to be the most suitable option, both for the quality of its results and for the feasibility of the lexical source. Also, some of the features of the *Véronis Model*, such as the syntagmatic information, could be included in those entries of *MiniDir.2.1* where this kind of information is important.
Finally, if we evaluate the results according to lexical categories, nouns achieve the highest levels of agreement: its referents are more stable and clearly identifiable. As regards adjectives, the levels of agreement are also high; since we have reduced the number of senses significantly (we have treated them as cases of vagueness instead of cases of polysemy). It is in verbs where there are more problems to distinguish senses.

4. Conclusions

In this study we have evaluated different lexical sources in order to determine the most adequate one for WSD tasks. The evaluation has consisted in the tagging of a single corpus with three different dictionaries and different annotators. The agreement degree among the annotators has been the determining criteria to establish the quality of the lexical source.
As a result of this study we have a lexical source, *MiniDir.2.1*, suitable for WSD tasks which constitutes the starting point for the development of a larger dictionary. Besides, we also have a Spanish corpus available, *MiniCors*, which has been manually and semantically tagged and which can be used as training corpus for machine learning systems in any NLP application. The high level of agreement between annotators guarantees the quality of *MiniCors*. These two lexical resources, *MiniCors* and *MiniDir.2.1.*, will be used in Senseval-3 competition.

Acknowledgements

The realization of this work has been possible thanks to the help received for the project X-TRACT2: BFF2002-04226-C03.
The annotators' team is composed by members of the research group CLiC of the University of Barcelona and by members of the Natural Language group of the UNED.

Endnotes

[1] These resources will be used for the Spanish WSD tasks in the third edition of the international competition Senseval (http://www.senseval.org/).

[2] For example, entries like *apuntar*, *usar* or *volar* have 29, 8 and 18 senses respectively in the *DRAE*, whereas in the *MiniDir.2.1* the number of senses are reduced to 9, 3 and 6 respectively.

[3] For example, the adjective *claro* has 18 different senses in the *DRAE*, whereas in the *MiniDir.2.1* the number is reduced to 5.

[4] *MiniCors* is the corpus that has been tagged with *MiniDir.2.1*, and it consists of news of the EFE Agency.

[5] The available volume of the EFE corpus is 2,814,291 sentences, 95,344,946 words, with an average of 33.8 words per sentence.

References

Artigas, N., García, M., Martí, M.A., Taulé M. 2003a. *Manual de anotación semántica*, XTRACT-WP-03/03, Centre de Llenguatge i Computació (CLiC), Universitat de Barcelona.

Artigas, N., García, M., Martí, M.A., Taulé M. 2003b. *Diccionario MiniDir.2.1.*, X-TRACT2 WP-03/08, Centre de Llenguatge i Computació (CLiC), Universitat de Barcelona.

Fellbaum, Ch., Grabowsky, J., Landes Sh. 1997. 'Analysis of a hand-tagging task', in Proceedings of ANLP-97 *Workshop on Tagging Text with Lexical Semantics: Why, What, and How?*, Washington D.C., USA.

Hanks, P. 2000. 'Do Word Meanings Exist?', in *Computers and the Humanities. Special Issue: Evaluating Word Sense Disambiguation Programs*, 34 (1-2), 205-215.

Ide, N., Véronis, J. 1998. 'Introduction to the Special Issue on Word Sense Disambiguation: the State of the Art', in Ide, N., J. Véronis, (Eds.), (1998) *Special issue of Computational Linguistics on Word Sense Disambiguation*, 24(1): 1-40.

Kilgarriff, A. 1999. '95% Replicability for manual word sense tagging', in Proceedings of EACL'99, Morgan Kaufman Publishers, San Francisco.

Krishnamurthy, R, Nicholls, D. 2000. 'Peeling an Onion: The Lexicographer's Experience of Manual Sense-Tagging', in Computers *and the Humanities. Special Issue: Evaluating Word Sense Disambiguation Programs*, 34 (1-2):85-7.

Márquez, Ll. 2002. 'Aprendizaje automático y procesamiento del lenguaje natural', in Martí, M.A., J. Llisterri (Eds.), (2002) *Tratamiento del lenguaje Natural*, Edicions de la Universitat de Barcelona and Fundación Duques de Soria, Barcelona.

Palmer, M. 1998. 'Consistent criteria for sense distinctions', in *Computers and the Humanities. Special Issue: Evaluating Word Sense Disambiguation Programs*, 34.

Real Academia Española, Diccionario de la lengua española, 22ª ed., Madrid, 2001.

Taulé, M., Civit, M., Artigas, N., García, M., Màrquez, Ll., Martí, M.A., Navarro, B. 2004. 'MiniCors and Cast3LB: two semantically tagged Spanish corpora'. Forthcoming in LREC'04, Lisbon.

Véronis, J. 2001. 'Sense tagging: does it make sense?', in Proceedings of the *Corpus Linguistics Conference*, Lancaster, UK.

Vossen, P. 1999. *EuroWordNet General Document*, http://www.hum.uva.nl/~ewn.