

Using a Diathesis Model for Semantic Parsing

Jordi Atserias, Irene Castellón*, Montse Civit and German Rigau

Departament de Llenguatges i Sistemes Informàtics. Universitat Politècnica de Catalunya
Jordi Girona Salgado, 1-3. Barcelona 08034 Spain

*LARELC, Secció de Lingüística General. Universitat de Barcelona
Gran Via de les Corts Catalanes, 585. Barcelona 08007 Spain.

Abstract

This paper presents a semantic parsing approach for non domain-specific texts. Semantic parsing is one of the major bottlenecks of Natural Language Understanding (NLU) systems and usually requires the building of expensive resources not easily portable to a different domain. Our approach obtains a case-role analysis, in which the semantic roles of the verb are identified. In order to cover all the possible syntactic realisations of a verb, our system combines their argument structure with a set of general semantic labelled diatheses models. Combining them, the system builds a set of syntactic-semantic patterns with their own role-case representation. Once the patterns are build, we use an approximate tree pattern-matching algorithm to identify the most reliable pattern for a sentence. The pattern matching is performed between the syntactic-semantic patterns and the feature-structure tree representing the morphological, syntactical and semantic information of the analysed sentence. For sentences assigned to the correct model, the semantic parsing system we are presenting identifies correctly more than 73% of possible semantic case-roles.

Keys:

Semantic Parsing, Semantic Interpretation, Information Extraction, NLU.

1 Introduction

Semantic parsing, seen as the mapping from words to semantics to produce a semantic interpretation of sentences (Hirst 87), is a major bottleneck of Natural Language Understanding (NLU) systems. Usually the parser and the semantic analysis are domain dependent. That is, it requires a high cost development of resources which are not easily portable to different domains. Although empirical machine learning methods have proven to be useful in reducing that cost for specific domains (Tou Ng et al. 97), more general

approaches are necessary in order to make the systems and the resources more portable to different domains.

Two of the main problems of the production of large-scale semantic analysis are the need to cover all possible semantic realisations of the concepts in a sentence and to produce the same conceptual representation. This task is crucial to obtain correct and complete case-role analysis, which the semantic roles of the verb such an agent, instrument, etc. are identified.

Our approach obtains a case-role analysis where the semantic roles of the verb are identified. Using general linguistic knowledge, the system automatically builds the syntactic-semantic patterns of all the possible realisations of the arguments of the verb (each verbal entry includes sub-categorisation structure, semantic roles and selectional restrictions). In order to cover all the possible syntactic realisations of a verb, our system combines their argument structure with a set of general Semantic Labelled Diathesis Models (SLDM). Combining them, the system builds a set of syntactic-semantic patterns with their own role-case representation. Once the patterns are built, we use an approximate tree pattern-matching algorithm to identify the most reliable pattern for a sentence. The pattern matching is performed between the syntactic-semantic patterns and the feature-structure tree representing the morphological, syntactical and semantic information of the analysed sentence. Currently, for sentences assigned to the correct model our system identifies correctly more than 80% of possible semantic case-roles.

Some Information Extraction Systems (such as FASTUS (Appelt 95) or PROTEUS (Grishman 95)) has begun to explore similar mechanisms based on "*meta-patterns*" to avoid multiple definition of the same extraction patterns due to the syntactic variations.

Our approach can be seen as a first step of a non-domain specific semantic parser. The system uses a large set of wide coverage tools and resources for Spanish. These tools and resources allow to build a feature-structure (FS) tree analysis of the sentence. This analysis contains

morphological, syntactical and semantic information provided by a wide coverage morphological analyser (Carmona et al.'98) and Tagger (Padró 98), a chart parser using a shallow grammar (Castellón et al. 98) and the Spanish EuroWordnet ontology (Farreres et al. 98) (Rodríguez et al. 98).

After this short introduction, Section 2 describes the system architecture and Section 3 explains the preliminary processes performed to produce the complete feature-structure tree analysis. Section 4 is devoted to the construction of the syntactic-semantic patterns by means of the SLDM and the verbal entries. Section 5 focus on the pattern-matching algorithm used to map the feature-structure analysis to the syntactic-semantic patterns. Section 6 describes the experiments carried out and the results achieved. Finally, Section 7 summarises some conclusions and possible further work.

2 System Architecture

The Semantic Parsing system we are presenting consists of three different modules:

1. The **Sentence Analyser** performs full syntactic analysis of the sentences. This module is described in next Section.
2. The **Pattern Builder** builds the sentence models using the verb sub-categorisation and the Semantic Labelled Diathesis Models (SLDM). Section 4 describes this module.
3. The **Tree-Pattern Matcher** chooses the best model for the parsed sentence and builds the final semantic case-role representation. This module is explained in Section 5.

3 Sentence Analyser

This module produces parsed trees for general domain texts. The Sentence Analyser involves a set of partial steps (i.e. tokenization, morphological analysis and tagging, syntactic parsing and semantic labelling). This process obtains a complete parsed tree for each sentence. Its nodes are lexical features containing lexical information provided by a wide coverage morphological analyser (Carmona et al.'98) and Tagger (Padró 98), and the semantic information from the Spanish Wordnet and the EuroWordnet Top Ontology (Farreres et al. 98). Next example shows for each word form the result of the Sentence Analyser before the parsing process. For each word form we obtained the lemma, disambiguated POS and EuroWordNet semantic labels.

Example :

Con	con	SPS00	*
sus	su	DP3CP00	*
labios	labio	NCMP000	BodyPart ...
,	,	Fc	*
fue	ser	VAIS3S0	Stative ...
susurrado	susurrar	VMPP0SM	Comm.Event ...
el	el	TDMS0	*
secreto	secreto	NCMS000	Meaning ...
.	.	Fp	*

The final syntactic analysis is produced by a chart parser that uses a wide coverage grammar (Castellón et al. 98).

4 Pattern Builder

This module completes the Semantic Labelled Diathesis Models (SLDM) using the verb specific information of the sentence to build tree patterns. Any missing information from SLDM is filled with the corresponding information of the verbal entry. If there is no specific information for that verb, the class information is used.

4.1 The Semantic Labelled Diathesis Models (SLDM)

We use the theoretical model of diathesis developed in the Pirapides project (Fernández et al. 98). The aim of the Pirapides project is to establish a wide coverage classification for verbs in Spanish and Catalan. The project works in the definition of a theoretical model for verbal entries based on three main components:

- **Eventual structure:** Based on (Pustejovsky 95), these structures can be classified in simple (event, state) and complex (when there is more than one even involved).
- **Meaning Components:** (Fernández et al 98). They are determined by each element the verb subcategorises (including the subject). They are more abstract than the thematic roles (i.e. *iniciador* (starter) includes the roles agent, source and the role denoting who performs the experience).
- **Diatheses:** They are the syntagmatic expression of the different semantic oppositions.

4.1.1 The diatheses in the Pirapides project

In Pirapides, diatheses are syntactic schemes related to semantic opposition. There are three different oppositions defined: change of focus, under-specification and aspectual opposition. **Change of focus** appears when there is a change in the point of view between the elements subcategorised by the verb.

i.e. *Los arquitectos construyeron el puente*
(The architects build the bridge).
El puente fue construido por los arquitectos
(The bridge was build by the architects)

The **under-specification** is produced when a verbal argument is omitted,

i.e. *El profesor dicta ejercicios a los alumnos*
(The teacher dictate exercises to the pupils)
El profesor dicta ejercicios
(The teacher dictate exercises)

and finally, the **aspectual opposition** implies a change between an event and an state, usually when a verbal modifier refers to an state.

i.e: *Ana bailó el tango.*
(Ana danced tango)
Ana baila el tango muy bien
(Ana is a good dancing tango)

In Pirapides, diatheses are defined as the syntagmatic expressions of a semantic opposition. This diathesis alternations are pairs of structures related to each other by one of those oppositions. From this point of view, the meanings of two sentences expressed with a pair of diatheses, related to an opposition, don't necessary share the same meaning.

Taking into account this alternations, verbs can be classified in three main classes according to whether they admit or not those oppositions.

By now, Pirapides has studied and defined the verbal classes of *change of state*, *attitude* and *transference*. The transference class has been divided in four sub-classes according to whether the verb can express both point of the trajectory (source, destination) or only one, and whether the verb can express a transfer done independently by the entity or not.

4.1.2 Semantic Labelling of the Pirapides diatheses

As described before, SLDM specifies syntactic alternations of verbs (active, passive, anti-causative, etc.) associated with a semantic opposition. Those alternations have been semantically labelled with role-names *-iniciador*

(starter), *entidad* (entity), *instrumento* (instrument)- and semantic constraints *-humano* (human), *animado* (animated), *instrumento* (instrument), *causa_natural* (natural_cause)-.

To obtain a full syntactic-semantic pattern of the verb argument structure, SLDM are combined with the syntactic and semantic information of the verb (the preposition that rules the argument, the selectional restrictions and the their possible syntactic realisations).

The elements of the SLDMs can contain the following information:

- Syntactic categories (and for PPs, preposition).
- Semantic constraints (selectional restriction).
- Morphological information (lemma, word form, gender, number, person).
- Coreference with other SLDM elements.
- Agreement with other SLDM elements.
- Optionality of the element.
- Role.

The diatheses have been classified according to the semantical transitivity of the verb. (semantically intransitive, transitive, transitive using PP).

For instance, Table 1 shows the SLDM for passive voice in transitive verbs. Note that the empty features do not constraint the SLDM.

4.2 The Verbal Entries

Verbal entries are described under a syntactic-semantic point of view and are logically organised in a hierarchy of classes. Each verbal entry specifies:

- The semantical transitivity of the verb.
- A list of its arguments/roles with:
 - The syntactic realisation of the role as noun phrase
 - The selectional restriction for the role
 - The preposition in case the role can appear as a PP.

For instance, table 2 shows the verbal entry for *susurrar* (whisper). As some features are also represented in SLDM both feature structures (FS) can be combined to build a more informed model.

4.3 Building a Pattern Model

Once verb is located in the sentence, its verbal entry is combined with the SLDM with the same transitivity to obtain the syntactic-semantic patterns.

Those patterns are built by completing the missing information in the verb-roles from SLDM

(the syntactic realisation, selectional restriction, PP preposition) and adding the specific roles of the verb, if any. For instance, table 3 shows how each role of the SLDM for passive voice (shown in table 1) is combined with the verbal entry *susurrar* (shown in Table 2) to make a tree-pattern. **Role entidad:** the verbal entry specifies that the entity role as a NP can be realised as a pronoun (npatons) as a NP (sn) or as a subordinate clause (prop). As neither **vaux** nor **event** appear in the verbal entry no information is added.

Role iniciador: This role realises syntactically as a PP with the preposition “*por/de*” (by). As no semantic constraint (selectional restriction) is

specified, the selectional restriction (Human) took from the verbal entry is added. **Role Meta:** As no preposition is present in the SLDM, this information is taken from the verbal entry “*a/al*” (to). In the same way, the selectional restriction *Human* is added as in the *iniciador*.

Once this process is completed, all roles from the verbal entry that do not appear in the SLDM and are not *entidad*, *iniciador*, *meta* are added as optional. So in the example of *susurra*, two more roles are added (**entidad.2** and **medio**).

Model	Trans		
Roles	NP realisation	Preposition for PP	Semantic
Iniciador (starter)	sn/%psubj		Human
Entidad (entity)	sn/spatons/prop		
Entidad.2 (entity.2)	sp	de/sobre	
Meta (goal)	sn/npatons	a/al	Human
Medio (instrument)	sp	con/por/a_través_de	Instrument

Table 1: Representation of the “*susurrar*” verbal entry.

Model	trans						
Meaning component	Syntax	Prep	Morph	Sem	Agreement	Co-refer.	Optional
entidad	sn				i		false
vaux	vser				i		false
event	vpart						false
iniciador	sp	por/de					true
meta	sp						true

Table 2: Table representation¹ of the “*susurrar*” verbal entry.

Meaning Component	Syntax	Prepositon	Morph	Semantic	Agr.	Cor.	Opt.
Entidad	<i>sn/patons/prop</i>				i		false
Vaux	vser				i		false
Event	vpart						false
Iniciador	sp	por/de		<u>Human</u>			true
Meta	sp	a/al		<u>Human</u>			true
<u>Entidad.2</u>	<u>sp</u>	<u>de/sobre</u>					<u>true</u>
<u>Medio</u>	<u>sp</u>	<u>con/por/a_través_de</u>		<u>Instrument</u>			<u>true</u>

Table 3: Pattern combination of the SLDM for semantic voice and the “*susurrar*” verbal entry.

¹ **Syntactic categories:** vser (auxiliar form), vpart (verb in past participle) sn (Noun Phrase), sp (Prepositional Phrase), patons (pronoun), npatons (pronoun except "se"). **Agr. :** Agreement in number and person. In the SLDM only a chain of agreement appears. **Coref:** Co-reference identifies elements referring to the same entity. In SLDM only one chain co-reference appears.

5 Tree-Pattern Matcher

This module determines which of the patterns created by the pattern-builder best fits a parse tree. In order to improve coverage, an inexact tree pattern matching algorithm is used. The method we propose is based on the definition of a similarity measure using tree edit operations. We adapt the method proposed by (Tsong-Li et al.'94) to retrieve similar syntactically labelled trees from a Tree Bank for comparing FS parse-trees with tree patterns. The major differences with our approach are:

- The trees are FS trees. The Tree pattern contains expressions such as (*or, not, substring*)
- The tree edit operation **-Relabel-** is applied on the features of the structure.
- A new tree model operation **-Move-** allowing disordering of the siblings.
- A new constraint in the pattern matching algorithm (the **structural criterion**) to avoid the deletion of some structures than are relevant from a linguistic point of view.
- The addition of some structural heuristics to the cost function.
- Our similarity measure is not a distance because we consider the **insert** and **delete** edit operations not always symmetrical. As the **insert** operation adds any kind of information not present in the tree pattern and the **delete** operation could remove relevant verbal arguments, both operations have different cost.

5.1 Adapting an approximate tree pattern-matching algorithm to FS parsed-trees

A parsed tree is an ordered tree whose features have lexical information. The patterns have also FS as nodes whose values can be an expression (not, or, prefix), a variable (for instance, to force agreement) or a constant value. We will define the mapping between a tree and a pattern as the function, resulting from the pattern matching, that assigns FS from the tree to another in the pattern.

A FS of a tree-pattern can be *unified* to another of the parse tree if and only if each feature *unifies* to the same feature of the parsed-tree. Moreover, we impose two restrictions to the tree pattern matching, the **ancestor criteria** and the **structural criteria**.

- **Ancestor-criteria:** The mapped FS of both trees must have the same ancestor relation.
- **Structural-criteria:** This criterion preserves the structures of the parsed-tree through the different levels of the mapping in order to avoid partial structure mappings. For instance, if we are looking for a NP in S, this criterion will avoid the mapping to a NP inside a PP (even if the PP is in S).

As explained before, we define a similarity measure to choose the best (one or more) of all possible matching between the parsed-tree and the pattern. This measure is defined as the minimum cost of all possible sequences of tree editing operations that transform one tree to the other. The cost of a sequence of operations is the addition of the cost of each operation. We defined the following tree edit operations:

- **Re-label**, this operation changes the value of a feature.
- **Delete**, this operation removes a FS of the pattern tree. There are two kinds of deleting, cutting (just this FS) or pruning (the node and all its descendants).
- **Move**, this operation changes the order of siblings.
- **Insert**, this operation adds a FS in the tree.

The cost function assigns a non-negative integer to each edit operation. To make the similarity measure between trees more general, the cost function not only depends on the type of the operation but also on the position in the tree. That is, whether a FS is a leaf (as leaves contain the word forms) or whether the FS does not have an ancestor mapped.

FS	Syntactic category	Preposition	Morph	Semantic	Agr. with the verb
0	sp	con		BodyPart	true
1	sn			BodyPart	true
2	fc				
3	grup-verbal (verbal group)				
4	vser				
5	vpart				
6	sn				true

Table 4: Main FSs of the parse-tree involved in the matching process

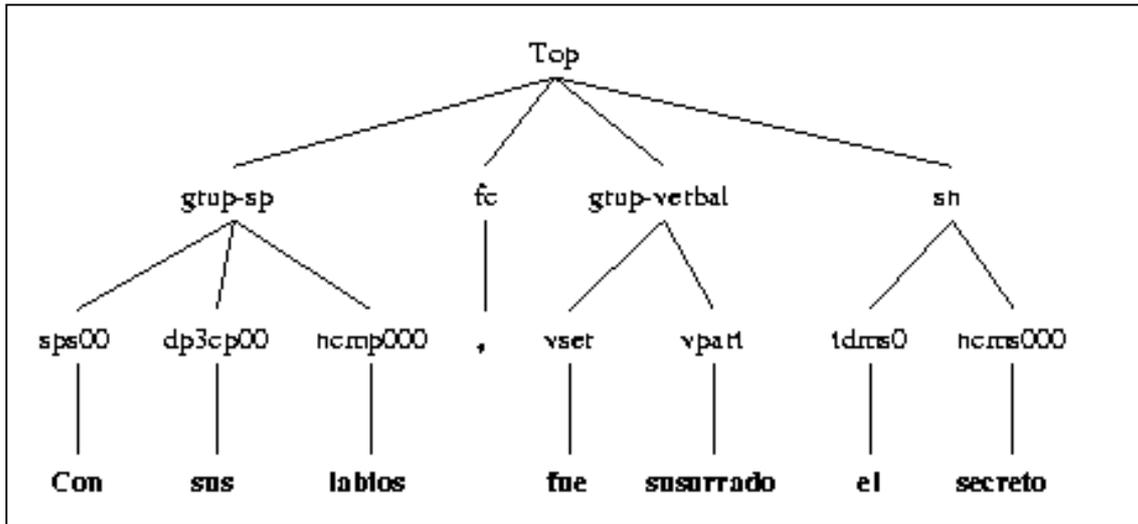


Fig 1: Parse tree: "Con sus labios, fue susurrado el secreto"

For instance, given the sentence "Con sus labios, fue susurrado el secreto" (literally, with his/her lips, the secret was whispered), the application of the syntactic-semantic pattern for the passive voice of the verb "susurrar" performs as follows.

The matching process tries to assign each FS in the pattern to another in the parse tree. Figure 1 shows the syntactic structure of the parsed-tree and table 4 its main FSs. The first role in the pattern (shown in table 3), the role *entidad* (an NP in agreement with the verb) could be mapped to the FSs 1 or 6 in the parse-tree. But the structural criterion will prevent from taking the NP 1, *labios*, from inside a PP. FSs with roles *iniciador* and *meta* are deleted as they can not be mapped to any FSs of the parsed-tree. The meaning components, *vser* and *vpart* can only be mapped respectively to FSs 4 and 5 in the parsed-tree. The role *medio* can only be mapped to FS 0 using the move operation and re-labelling the semantic feature from BodyPart to Instrument. Finally, the algorithm insert the FSs 3,2 and the leaves from the parsed-tree to the pattern as it can not be mapped to any other FS in the pattern.

The resulting similarity between the parsed-tree and the pattern from the SLDM for passive voice

of the verb *susurrar*, will be calculated by adding the costs of the two deletes, the move, the re-label and the nine insert operations.

6 Experiments & Results

As one of the major goals of this work has been to test not only the feasibility of the method but also the linguistic data, we have performed a complete experiment for Spanish using 42 SLDM for the verbs of the **Trajectory Class** developed in **Pirapides** (Morante et al. 98) and ten verbal entries. Eight of these entries from the communication class which is a sub-class of the trajectory class, *-explicar* (to explain), *charlar* (to chat), *decir* (to say), *hablar* (to talk), *murmurar* (to murmur), *susurrar* (to whisper), *discutir* (to discuss), *criticar* (to criticise)-, and two from other verbs belonging to other classes *-reprender* (to reprimand), *invitar* (to invite)-.

In order to test the generality and soundness of the method we also used a corpus that not only contains verbs of the trajectory class. The sentences of this corpus contain prototypical diathesis alterations.

6.1 Semantic Representation

To perform wide-coverage semantic (neither domain specific nor language specific), the 79 semantic labels defined in the preliminary version of the Top Ontology was chosen as a common semantic representation for SLDM, verbal entries and the parse tree. The Top Ontology was developed inside the EuroWordNet project as an ontology for clustering the common base concepts defined for the different languages involved in the project (Vossen et al'97).

6.2 Corpus

We have divided the whole corpus in two. The first part has been used for tuning the SLDM models (the tuning corpus) and the second one (the test corpus) for testing the process independently. During the tuning process, we modify the 31 original models adding ten more SLDM produced by splitting the initial models or by taking into account new models. Table 5 summarises some figures of corpus used in the experiment.

	Tuning	Test
<i>Sentences</i>	257	47
<i>Words</i>	1557	274
<i>Communi. verbs</i>	186 (72%)	26 (56%)
<i>Other verbs</i>	71 (28%)	21 (44%)

Table 5: Figures of the corpus.

6.3 Syntactic Analysis of the corpus

The corpus was processed to obtain a complete parsed tree for each sentence. The nodes of the parse-trees contain lexical features provided by a wide coverage morphological analyser (Carmona et al.'98) and Tagger (Padró 98), and semantic information from the EuroWordnet Top Ontology (Farreres et al. 98). See Section 3.

6.4 Results

Although the project is in progress, performing a cycling tuning process on the linguistic data and algorithm, our initial figures seem to be very promising. The current version achieves with a total coverage out of 96%, a precision of 72% in the test corpus SLDM identification task, and a precision out of 73% in the semantic-role identification task. Moreover, due to slightly differences between models, even when an incorrect SLDM has been selected as a solution, the semantic-role identification is correctly performed. Table 6 shows the results in terms of recall and precision focusing on the model.

	Model	
	Rec.	Prec.
Tuning	85%	88%
Test	66%	72%

Table 6: Recall and precision for the model identification task.

The evaluation criterion for the roles is the exact string equality, for instance the value "*mismo emperador*" (the emperor himself) to fill a role with value "*emperador*" (emperor) will be counted as a miss. Also missing roles or roles that are not in the solution are counted as errors. As multiple answers and solutions are possible for a sentence, table 7 presents the results per model and role focusing on the sentence according to four evaluation criteria. At least one answer is correct, all the answers are correct, all the answers include all the solutions and all the answers and solutions are the same (there are not more answers than solutions and all are correct).

6.5 Analysing errors

The main sources of misleading information came from the **Sentence Analyser** module (morphological, semantic and syntactic errors). **Morphological errors** are produced mainly, when the verb is not recognised as a verb. Moreover, errors in POS tagging can produce incorrect syntactic groups during the parsing and furthermore incoherent structures for the SLDM. **Semantic errors** are produced when no semantic labels are found for some words, converting several SLDM to the same patterns and producing an over-generation of solutions. **Syntactic Errors** produced during the parsing process introduce noise in the result. The main causes of syntactic mistakes are produced by noun modifiers, PP-attachment and bad identification of sentence boundaries.

7 Conclusions & Further Work

This paper has presented a semantic parsing approach for non domain-specific texts. Our approach obtains a case-role analysis, in which the semantic roles of the verb are identified using general domain resources (taggers, shallow parsers and semantic ontologies). In order to cover all the possible syntactic realisations of a verb (or the class model of the verb), our system combines their argument structure with a set of general semantic labelled diathesis models. Combining them, the system builds a set of syntactic-semantic patterns with their own role-case representation. Once the patterns are build, we use an approximate tree pattern-matching algorithm to identify the most

reliable pattern for a sentence. The pattern matching is performed between the syntactic-semantic patterns and the FS tree representing the morphological, syntactical and semantic information of the analysed sentence. For sentences assigned to the correct model, the semantic parsing system we are presenting identifies correctly more than 73% of possible semantic case-roles.

Although the results of the experiments are promising for simple sentences, some tuning must be performed on the SLDM to achieve better performance. Improvements on the similarity measure adding statistical information or probabilities to the model could also be tried. Moreover, to design a more general framework, we are planning to formalise the pattern matching and models as a Consistency Labelling Problem (see (Padró 98)) in which different nominal and verbal models can compete for their case-roles assignment.

Acknowledgements

This research has been partially funded by the Spanish Research Department (ITEM Project TIC96-1243-C03-03 and Spontaneous-Speech Dialogue System In Limited Domains TIC98-423-C06), the Catalan Research Department (CREL Project), and the EU Comission (EuroWordNet LE4003).

References

- (Appelt 95) Douglas E. Appelt, Jerry R. Hobbs, John Bear, David Israel, Megumi Kameyama, Andy Kehler, David Martin, Karen Myers, Mabry Tyson *SRI International FASTUS system MUC-6, test results and analysis*, Proceedings of the MUC-6 conference.
- (Castellón et al. 98) Irene Castellón, Montse Civit Jordi Atserias *Syntactic Parsing of Unrestricted Spanish Text* in the 1th International Conference on Language Resources & Evaluation. May 1998. Granada Spain.
- (Carmona et al. 98) J. Carmona, S. Cervell, L. Màrquez, M.A. Martí, L. Padró, R. Placer, H. Rodríguez, M. Taulé & J. Turmo *An Environment for Morphosyntactic Processing of Unrestricted Spanish Text* First International Conference on Language Resources and Evaluation (LREC'98). Granada, Spain, 1998.
- (Farreres et al. 98) Farreres X., Rigau G. and Rodríguez H., *Using WordNet for Building WordNets*. Proceedings of COLING-ACL Workshop "Usage of WordNet in Natural Language Processing Systems". Montreal, Canada. 1998.
- (Fernández et al. 98) Fernández Montraveta, A., M. A. Martí, G. Vázquez, I. Castellón i R. Morante (1998) *Criterios de clasificación de las estructuras eventivas: el cambio y la transferencia*, AESLA 1998, La Rioja.
- (Grishman 95) R. Grishman *The NYU System for MUC6 or Where's the Syntax?* Proceedings of the MUC-6 conference.
- (Hirst 87) Hirst G., *Semantic Interpretation and the Resolution of Ambiguity*, Studies in NLP, Cambridge University Press, 1987.
- (Morante et al. 98) Morante, R., Castellón, I., Vázquez, G. (1998) *Los verbos de trayectoria*. Congreso de la SEPLN-98
- (Padró 98) L. Padró *An Hybrid Environment for Syntax-Semantic Tagging*. PhD Thesis, Departament de Llenguatges i Sistemes Informàtics. Universitat Politècnica de Catalunya. Barcelona, 1998.
- (Pustejovsky 95) Pustejovsky, J. *The generative lexicon*, The MIT Press, Cambridge, MA. 1995.
- (Rodríguez et al. 98) Rodríguez H., Climent S., Vossen P., Blocksma L., Peters W., Alonge A., Bertagna F., and Rovertini A. *The top-down strategy for building EUWN: Vocabulary coverage, base concepts and Top-Ontology*. In Computers and the humanities. N. Ide & D. Greenstein Eds. vol 32, n 2-3, 1998.
- (Tou Ng 97) Hwee Tou Ng and John Zelle *Corpus-Based approach to semantic interpretation in Natural Language Processing* AI magazine, 45-63, 1997
- (Tsong-li et al. 94) J. Tsong-Li, K. Zhang, K. Jeong D. Shasha *A system for Approximate Tree Matching*, IEEE transactions on knowledge and data engineering, Vol. 6, No 4. 1994.
- (Vossen et al. 97) Piek Vossen, Wim Peters, Pedro Díez *The multilingual design of the EuroWordnet Database* ECAI'97 workshop on multilinguality in the lexicon I.