

# LEXPIR: A VERB LEXICON

Montserrat Civià  
Centre de Recerca TALP  
Universitat Politècnica de Catalunya

Irene Castellón  
M<sup>a</sup> Antònia Martí  
Mariona Taulé  
Departament de Lingüística General  
Universitat de Barcelona

## ABSTRACT

*The present work lies within the area of Information Extraction (IE). Usually, IE systems deal with restricted semantic domains and have been mostly developed for English. This paper describes LEXPIR, a Spanish verb lexicon that plays a central role in an IE system that is based on linguistic knowledge, and is capable of dealing with unrestricted domains. This verb lexicon has been developed following the theoretical proposal of the Piràpides project. The objective of this project is the definition of a theoretically founded model of verb lexical entry, from which to derive the predicate semantic classification. In addition to LEXPIR, the IE system architecture includes a syntactic analyser (TACAT), a morphological analyser (MACO) and a semantic net (EuroWordNet). The system helps to obtain a semantic representation for the text basic contents.*

**KEY WORDS:** *information extraction, semantic domain, semantic representation, verb lexicon*

## RESUMEN

*El presente trabajo se encuadra dentro del área de la Extracción de Información (EI). Habitualmente, los sistemas de EI se han limitado a dominios semánticos restringidos y al inglés. En este artículo describimos LEXPIR, un lexicon verbal para el español que constituye el núcleo de un sistema de EI basado en conocimiento lingüístico, capaz de operar en cualquier dominio. Este lexicon verbal ha sido desarrollado siguiendo las propuestas teóricas del proyecto Piràpides. El objetivo de este proyecto es la definición de un modelo de fundamentación teórico para entradas verbales, y del cual se pueda inferir una clasificación de orden semántico-predicativa. Además de LEXPIR, la arquitectura del sistema de EI cuenta con un analizador sintáctico (TACAT), un analizador morfológico (MACO) y una red semántica (EuroWordNet). El sistema permite obtener representaciones semánticas de los contenidos básicos del texto.*

**PALABRAS CLAVE:** *extracción de información, dominio semántico, representación semántica, lexicon verbal*

## I. BACKGROUND

Information Extraction is an application whose complexity is basically determined by three factors: the type of text that is to be processed, the richness and variety of the information within the application domain, and the adequacy of the templates for the information to be extracted.

The work carried out within the area of information extraction is twofold: first, systems that extract linguistic information in order to feed lexical databases (lexical knowledge acquisition), and secondly, systems that are oriented towards semantic extraction from corpora, and that are applied to the semantic indexing of documents, information retrieval, etc.

The former usually focus on the extraction of subcategorisation frames. Lapata (1999) extracts syntactico-semantic frames that are associated to certain prepositions (Basili et al. 1998) and (Briscoe et al. 1997) extract subcategorisation frames based on the identification of category sequences, *chunks*, within a corpus. In the same line, (Poznansky & Sanfilippo 1991) extract subcategorisation frames that are related to semantic classes, based on the diathesis alternation proposals in the work of Levin's (1993).

The extraction of semantic information from corpora has been strongly encouraged by the seven *MUC* (*Message Understanding Conference*) evaluation conferences. These conferences are conceived as a competition where the different IE systems presented are evaluated. At the initial stage, participants receive a training corpus together with the templates to fill in. After a six month period, during which each group tunes its own system's tools by testing them on the training corpus, the evaluation is performed on a new set of articles, which is referred to as *test set*. Results are evaluated by comparing them to some templates that have been manually filled in by experts. The evaluation follows two basic measures: *precision* (quality of extracted information) and *recall* (relation between the information extracted and that which should have been extracted).

Despite the fact that, generally speaking, most studies have been carried out for English, some research has already begun for other languages, such as Spanish. For instance, TURBIO (Turnio 1997) is an IE system based on pattern learning that can be ported to different domains, even if it cannot handle unrestricted corpora.

Bearing all this in mind, our proposal follows three basic aims:

extending IE methods to Spanish:

1. extending IE methods to Spanish
2. applying them to general domains: and
3. basing the methodology on a language theoretical model (Vazquez et al. 2000)

This article presents the two modules comprising the system, the analysis and extraction modules, and offers a detailed description of LEXPIR, the central module upon which the IE process is based.

## II. SYSTEM ARCHITECTURE

The system here proposed can be classified as belonging to the group of IE systems using linguistic knowledge. In general, these systems function in two stages: a first stage of analysis and a second one of extraction, which is also the approach followed by our proposal. Closely related to most IE systems, the processing takes place in two separate modules (cf. Figure 1). The analysis module prepares the text for extraction by tagging it first (MACO; Atserias et al. 1998a and RELAX; Padró 1997), and then, partially parsing its output (TACAT: Atserias et al. 1998b and Castellón et al. 1998).

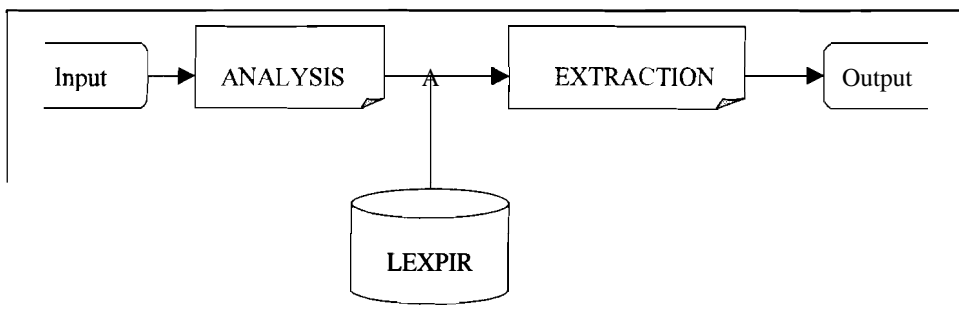


Figure 1: Information extraction system

The aim of the extraction module is to obtain the semantic interpretation of the sentences. In order to do so, two knowledge resources are used: EuroWordNet (EWN) and LEXPIR. EWN is used for the semantic tagging (EWN: Alonge et al. 1998), and LEXPIR allows the integration of all the morphological, syntactic and semantic information.

### II.1. Analysis Module

The main objective of this module is the morpho-syntactic interpretation of the text units. The module components behave in a sequential manner, enriching the sentence linguistic

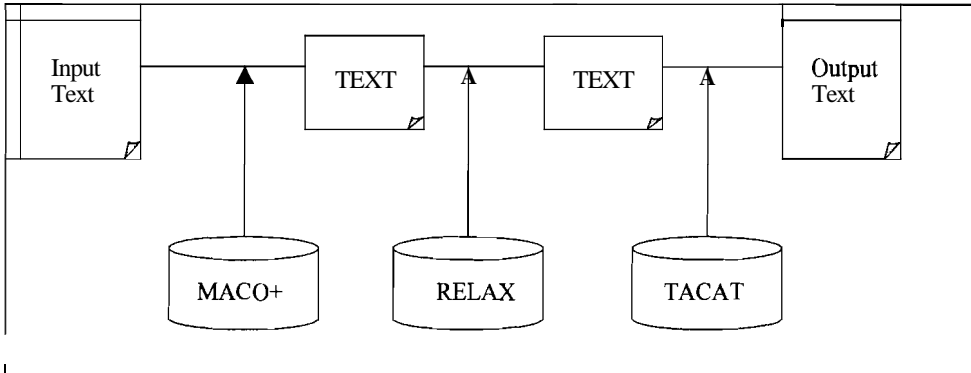


Figure 2: Operational behaviour of the analysis module

information by means of consecutive applications. Figure 2 illustrates the operational behaviour of these components.

### II.1.1. MACO and RELAX

MACO+ (Morphological Analyzer Corpus Oriented) is a morphological analyser whose linguistic knowledge is organised into classes and inflection paradigms. The forms contained in it are considered from an orthographic point of view. The analyser comprises about 90.000 base forms (Arévalo et al. 2000), which enable the analysis of about a million forms<sup>1</sup>.

MACO+ can be divided into several specialised recognisers. This modular organisation allows the recognition of abbreviations, punctuation marks, proper names, compounds, dates, etc. The result obtained from MACO+ is a set of feasible tags per word and its corresponding root form. These labels carry morphological information regarding grammatical category, gender, number, person, etc. Figure 3 shows the result achieved with the application of the tagger to the following sentence: "el invitado trajo una botella de vino"<sup>2</sup>.

<sup>1</sup> Its coverage is currently being increased so as to treat proper names and clitics  
 "The guest brought a bottle of wine".

```

El el TDMSO
invitado invitado NCMSOO invitar VMPP0SM
trajo traer VMIS3S0
una un TIFSO uno PI0FS000 un MCFSOO una NCFSOOO unir VMSP3S0
      unir VMSPIS0 unir VMMP3S0
botella botella NCFSOOO
de de SPSOO de NCFSOOO
vino venir VMIS3S0 vino NCMSOOO
.. Fp
    
```

Figure 3: Result of MACO+ analysis

The aim of RELAX (Relaxation Labelling Based Tagger; Padró 1997) is the disambiguation of these labels and, therefore, the obtaining of a unique morphological interpretation. Figure 4 shows the result of applying the disambiguation process to the previous sentence:

```

El el TDMSO
invitado invitado NCMSOOO
trajo traer VMIS3S0
una un TIFSO
botella botella NCFSOOO
de de SPSOO
vino vino NCMSOOO
.. Fp
    
```

Figure 4: Morphological disambiguation of the text

### II.1.2. TACAT

TACAT aims at obtaining the syntactic labelling of the corpus. This tool performs a superficial analysis, taking as input the tagged text provided by MACO+ and RELAX, and giving as output a constituent phrase grouping that does not make explicit any internal dependencies (Civit et al. 1998). This system offers the possibility of performing analysis at different levels given that it can interact sequentially with several grammars<sup>1</sup>. Figure 5 shows the result of the analysis done on the previous sentence.

---

<sup>1</sup> In order to avoid an extremely overloaded tree representation of the phrases, TACAT allows the flattening of the analysis by “hiding” some of the intermediate labels that must be declared in lists.

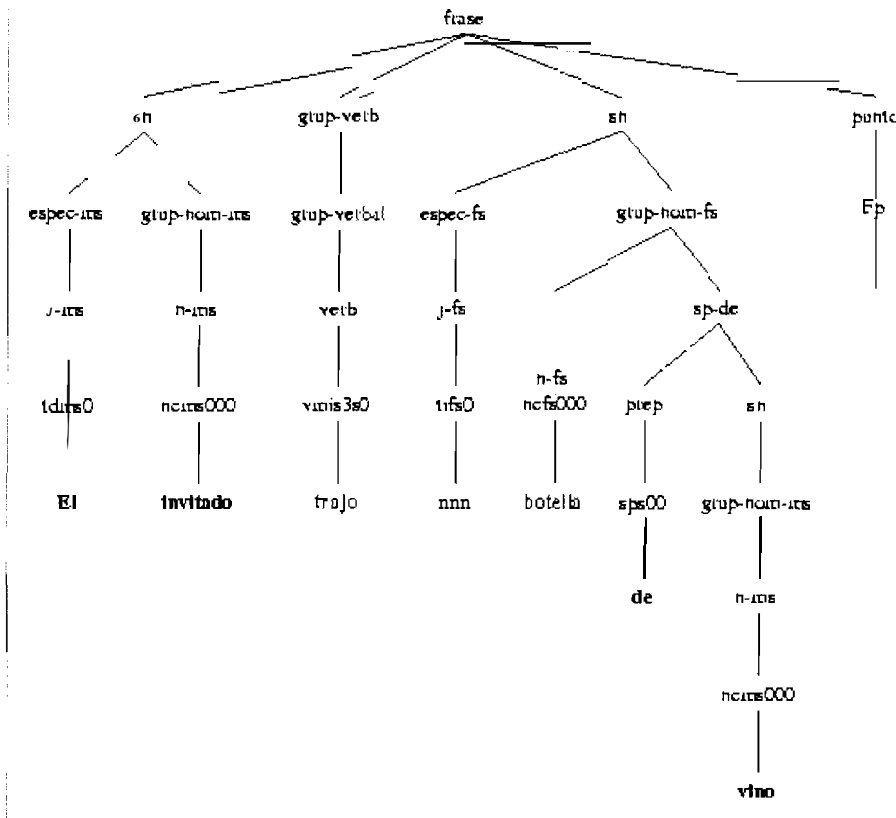


Figure 5: Syntactic analysis of the sentence

2. Extractinn Module

The extraction module aims at obtaining semantic interpretations from analysed text. The components of this module are EWN and LEXPTR, which contain semantic and syntactico-semantic information, respectively. The interaction between both components is performed in a dynamic manner by means of LEXPIR's consultation with EWN: EWN provides the semantic labelling, while LEXPIR integrates all the information acquired in the previous phases of the process and acts as interface between the analysis and the extraction. The system architecture can be seen in figure 6:

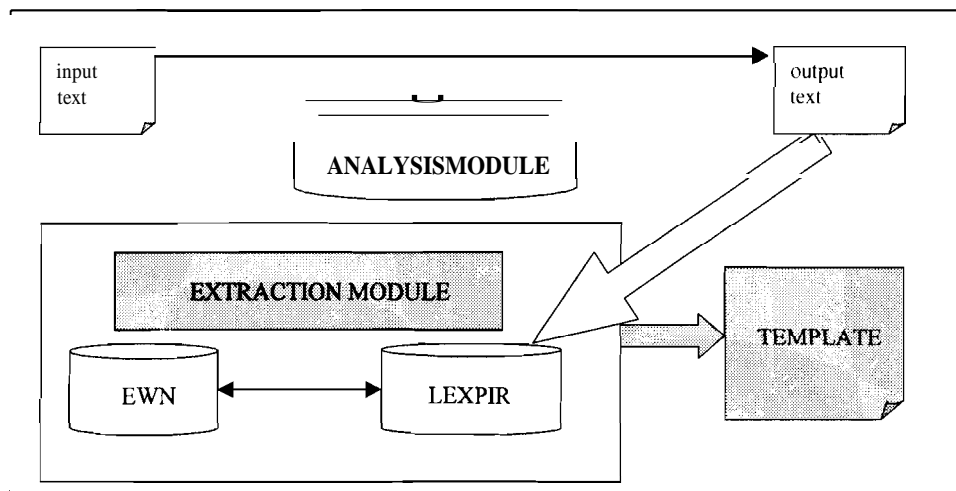


Figure 6: System architecture

### II.2.1. EuroWordNet

The Spanish WordNet (SpWordNet) is one of the several wordnets that has been integrated into the multilingual lexical database EuroWordNet<sup>1</sup> (Vossen 1998). Similarly to the other languages represented in EWN, the SpWordNet initially follows the WordNet model developed at Princeton University (Fellbaum 1998)<sup>2</sup>.

Wordnets are ontologies that are semantically organised around the notion of synset. A synset is a set of synonyms that have been assigned the same part-of-speech (POS) and represent a unique underlying lexical concept. These synsets are linked to each other by using semantic relations such as hypernymy, hyponymy, meronymy, antonymy and so on.

The POS categories represented in SpWordNet are nouns, verbs and adjectives. Adverbs could also be represented, but they have not been treated yet. To present, the SpWordNet covers the basic and most general vocabulary of the Spanish language.

Regarding cross-language relations, each Spanish synset is linked to its equivalent, or closest, English synset in a direct or indirect manner. In fact, the synsets in WordNet 1.5

<sup>1</sup> EuroWordNet (LE 2-4003 and LE 8328) has been funded by the European Community. The project started in 1996 and ended in 1999. The languages represented in EuroWordNet are Dutch, Italian, English, Spanish, French, German, Estonian and Czech.

<sup>2</sup> To be precise, it should be mentioned that the version used has been WordNet 1.5.

function as a sort of interlingual index that ensure the multilinguality of the EuroWordNet.

The following table details the total amount of synsets that are currently represented in the SpWordNet.

<b>Part-of-speech</b>	<b>Synsets</b>	<b>Variants</b>	<b>Words</b>
Noun	43.333	61.794	46.807
Verb	7.933	12.518	8.484
Adjective	12.148	16.327	5.313
Total	63.414	90.139	60.604

Table 1: Data Records

The usefulness of a resource such as WordNet stands out immediately since it can be used to make semantic inferences in a variety of research areas. For instance, it can be very useful for (cross-language) information retrieval applications, and also as a lexical resource for machine translation, word sense disambiguation (Escudero et al.2000), or language learning systems (Morante 2000).

The main objective of EWN within the system described here is that of helping to identify the semantics behind the verbal arguments. This will be explained in section 3. Furthermore, section 3 will also provide a detailed description of the component LEXPIR, since this part is the focus of the current paper.

### III. LEXPIR

Text parsing can be performed with context sensitive grammars (CSG), which help to obtain good results with restricted domains, but cannot be applied to general language texts. Alternatively, parsing can also take place with context free grammars (CFG), which, despite being more robust, do not allow a very detailed analysis.

Moreover, the analysis of Spanish adds a further specific problem, when compared to that of English: on the one hand, it is a free constituent-order language and, on the other, its constituents do not always occur in an explicit manner. However, for an information extraction system, it is essential to define clearly which constituents within a sentence are arguments and which are adjuncts, as well as what kind of relation a verbal predicate establishes with its arguments. Therefore, the establishment of the dependencies among the elements in the sentence is a task that must incorporate another type of information, in addition to the purely



syntactic one.

LEXPIR is a verb lexicon based on the theoretical proposal of the *Pirápides* project (Fernández et al. 1999). It provides a model for the verb lexical entry where both syntactic and semantic information are integrated.

A verb entry is structured in relation to the following information modules:

- 1) **Meaning components.** This refers to semantic units that are contained within the lexical items, and that are organised in an ontology with a discriminating nature towards predicate classification. These components are placed at a higher level of abstraction than thematic roles, so that a meaning component can contain two or more thematic roles. Further, the following should also be taken into account: they can occur as part of the verb root itself, they might not occur explicitly, or one lexical item might express two components (co-indexing).
- 2) **Event structure.** This describes the internal time distribution of the predicate. The types of event structure that have been proposed within the work of *Pirápides* are basically two: *states* and *events*. The difference between them is established by using the feature [ $\pm$  dynamic], where *dynamic* stands for the progression of a situation in time. Both basic types have a correlate in the meaning components, given that while events imply the existence of a initiator, states are defined as properties described in relation to an entity.
- 3) **Diathesis component.** Within the *Pirápides* model, a diathesis can be defined as the phrase expression of different semantic oppositions that are motivated by various communication strategies. These constructions are considered in alternation pairs, where each structure expresses one of the meanings in the opposition. Therefore, the diathesis associates a syntactic structure to a semantic interpretation (either an event or a state, with its participants).

In a first approximation to the predicate classification based on this model, three large semantic classes have been defined: *change*, *trajectory* and *attitude* (Fernández et al. 1996, Morante et al. 1998). Each verb sense is associated to a class, and consequently, a verb form will occur as many times as senses has had identified.

As it will be seen in figure 7, only the semantic and diathesis components have been formalised in LEXPIR.

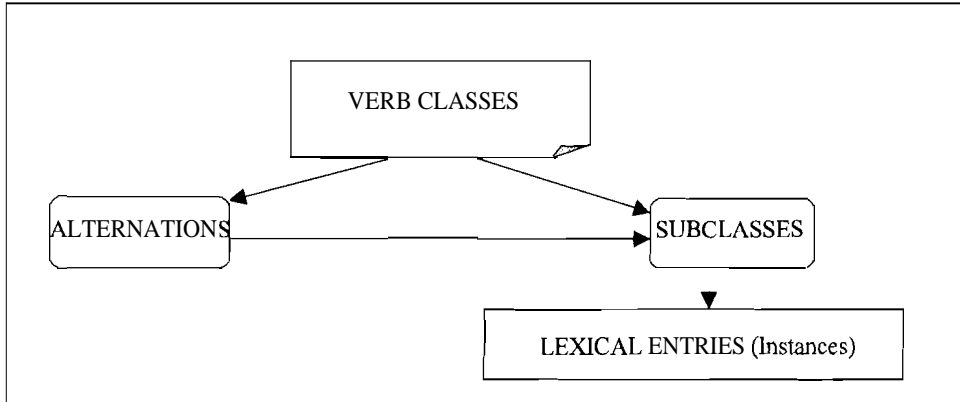


Figure 7: LEXPIR's framework

Below follows the description of the internal organisation of the information in LEXPIR. i.e., the verb hierarchy and the lexical entries.

**The verb hierarchy:** each verb class has some meaning components and diathesis alternations associated. Furthermore, this application also considers information regarding the specific syntactic structure of the components, the prepositions that can mark them and their particular semantics, as well as their possible agreement with the verb and its optionality. It might occur that this information is not specified if, due to the case variety offered by the verb forms included in a class, it cannot be made explicit. The transmission of information along the hierarchy takes place in a top-down fashion, i.e., starting from the class and coming down to the verb instances, and applying simple monotonic inheritance by default.

**The verb instances:** information is propagated, firstly, from the classes to the subclasses, and then, to the specific verb entries. When occurring in any of these latter levels, the information can be either made explicit or modified, depending on the subclass or verb specific characteristics.

### III.1. The Trajectory Class

This section deals, as a kind of example, with the *trajectory* class, explaining the general diagram for the class (figure 8), the alternations it presents (figure 9), the subclasses that constitute it (figure 10) and its instances (specific verb forms) (figure 12).

As it can be observed in figure 8, the trajectory class is semantically characterised by the *initiator*, *entity* and *trajectory* components: the latter being also subdivided into three further: *path*, *source* and *goal*. Each of these components presents the following characteristics within the diathesis structure here regarded as basic:

- a phrase structure:
- one (or more) prepositions that introduce it: and
- a particular semantics.

With regard to the diatheses that accept a predicate, these can alter both the order and omission of a component. Moreover, it is also indicated which component establishes agreement with the verb, when the former is syntactically manifested.

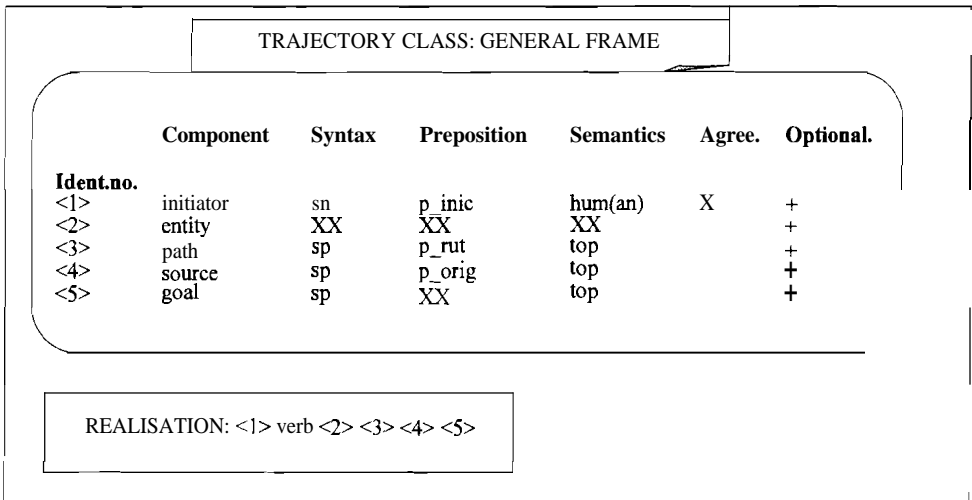


Figure 8: Information representation within the hierarchy: the class

The fields established in the LEXPIR Database (DB) are the following<sup>1</sup>:

- (i) Identification No.: numerical value that allows the identification of the components.
- (ii) Component: meaning component determined by the class.

---

<sup>1</sup> The information from the *Pirápidas* model has been implemented as a DB for this particular application.

- (iii) **Syntax**: phrase structure of each component. This information can be unspecified (such as component “2” in figure 8 can vary between “sn”<sup>2</sup> and “sp”<sup>3</sup>, depending on the semantic class). This is marked with value “XX”, since the specific value will be determined in the subclasses.
- (iv) **Preposition**: prepositions have been classified according to their meaning and the contexts in which they occur (See example. **p\_rut** contains values “por”<sup>4</sup>, “a través de”<sup>5</sup>. etc.).
- (v) **Semantics**: semantic class for the noun that occurs in this position: this characteristic is specific to each argument and its values are taken from the TopOntology in EWN.
- (vi) **Agreement**: “x” indicates which element must agree with the verb in the sentence.
- (vii) **Optionality**: “+” means that these are optional elements, that is, that they can either occur or not in the sentence.

For instance:

alguien <1> traslada algo <2> por X <3> de Y <4> a Z <5>  
*somebody <1> moves something <2> through X <3> from Y <4> to Z <5>*

Along with the basic diagram take place the alternations that are accepted by the corresponding verb class. The fact that optionality is treated in the same diagram allows the reduction of the diathesis alternations that have been established at the theoretical level (in particular in *Pirápidas*, the underspecification of some of the components is considered as an alternation). In these diagrams, only relevant information is pointed out, which refers to the information that offers some variation with respect to the one of the class. This can be observed in figure 9, where certain elements have been emphasised in bold so as to distinguish those that have been specified with regard to general information:

---

Noun phrase.

<sup>2</sup> Prepositional phrase.

<sup>4</sup> Along/through.

<sup>5</sup> Across/through.

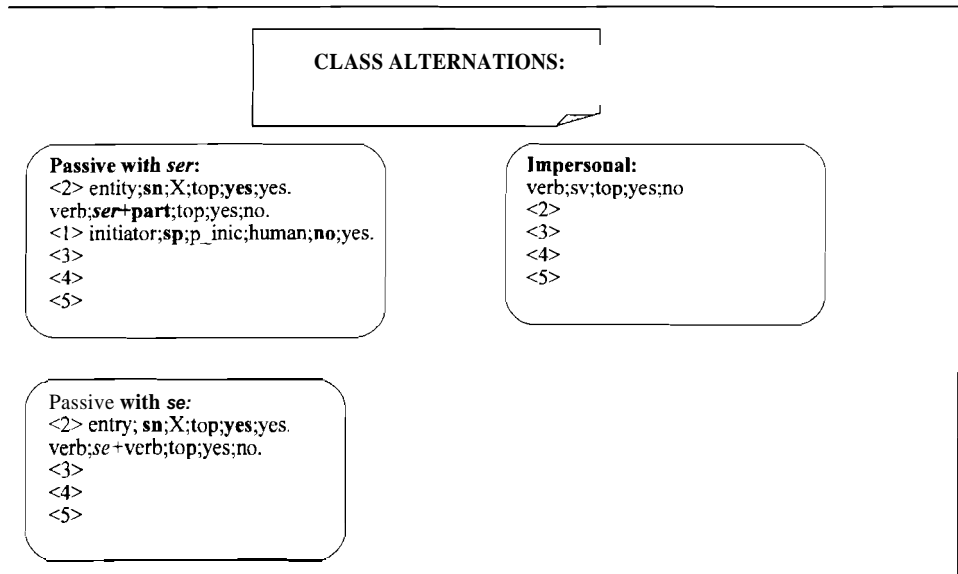


Figure 9: Representation of the information in the hierarchy: diathesis alternations of the class

Further on the trajectory class example. the alternations of the verb *trasladar* would be as follows:

(1) passive with *ser*<sup>1</sup>:

algo <2> es trasladado por alguien <1> por X <3> de Y <4> a Z <5>

*something <2> is moved by somebody <1> through X <3> from Y <4> to Z <5>*

(2) passive with *se*<sup>2</sup>:

algo <2> se traslada por X <3> de Y <4> a Z <5>

*something <2> is moved through X <3> from Y <4> to Z <5>*

(3) impersonal.<sup>3</sup>:

algo <2> se traslada por X <3> de Y <4> a Z <5>

*something <2> moves through X <3> from Y <4> to Z <5>*

<sup>1</sup> Passive construction with verb *ser* (to be).

<sup>2</sup> Passive costruction with reflexive pronoun *se*

<sup>3</sup> Impersonal costruction.

The subclasses comprised in the trajectory class are these four: *non-autonomous movement*, *autonomous movement*, *communication* and *transfer*. The non-autonomous movement subclass is characterised by explicitly presenting the five components:

alguien <1> desplaza algo <2> por X <3> desde Y <4> a Z <5>  
*somebody <1> moves something <2> through X <3> from Y <4> to Z <5>*

The autonomous movement subclass presents a co-indexing<sup>4</sup> of the initiator and entity components, the latter always occurring in subject position:

alguien <1-2> va por X <3> de Y <4> a Z <5>  
*somebody <1> goes through X <3> from Y <4> to Z <5>*

As it can be observed in the example, (1) is both the initiator of the action *ir* and the entity being moved.

Finally, only three components occur explicitly in the verbs of communication and transfer (initiator—which is simultaneously source-, entity and goal):

a. alguien <1,4> **dice** algo <2> a alguien <5>  
*somebody <1,4> says something <2> to someone <5>*

or:

b. alguien <1,4> **da** algo <2> a alguien <5>  
*somebody <1,4> gives something <2> to someone <5>*

The absence of components is marked with “0”. However, should the information regarding any of them be the same as that provided by the class, it will be then marked with the identification number. Moreover, the prepositions that can occur in component (5) are only specified in the subclasses and can be divided into two groups: “p\_dest1”, which includes “a/para”<sup>5</sup>, and “p\_dest2”, which covers the rest. Accordingly, the basic diagram which corresponds to the communication subclass would be the following (cf. Figure 10):

---

<sup>4</sup> Two components are referred to as co-indexed when both of them are syntactically realised in the same constituent.  
<sup>5</sup> To/for.

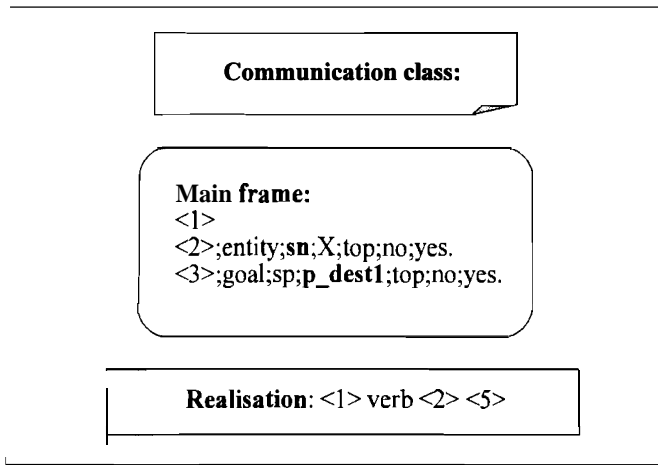


Figure 10: Representation of the information in the hierarchy: the communication subclass

In order to obtain the alternations that a particular subclass can display, the unification of the general alternation information is performed, together with the particularities for each class. This takes place by giving priority to the information provided by the subclass, should it contradict that given by the class. Bearing this in mind, the "impersonal" alternation for the communication subclass would be organised as it is shown in figure 11, for the sentence *se charla de política*<sup>1</sup>

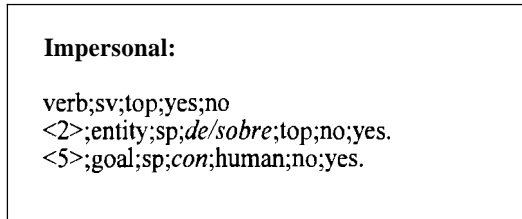


Figure 11: Impersonal alternation for the communication subclass

---

<sup>1</sup> They/we (are) talk(ing) politics

Say.

Chat/Talk.

<sup>2</sup> About.

<sup>3</sup> About.

<sup>4</sup> To/for.

<sup>5</sup> With.

*Specific verb forms can impose their own restrictions. Contrary to the verb "decir", that follows strictly the subclass model, there occurs "charlar"<sup>3</sup>, which does not admit a noun phrase in the structure of the 'entity' component, and also demands a prepositional phrase headed by the prepositions "de" or "sobre", in opposition to the definition of subclass.*

*In addition, it does not accept prepositions "a/para"<sup>6</sup> either in order to express goal. It only accepts "con"<sup>7</sup> due to which its entry in the lexicon would be as shown in figure 12:*

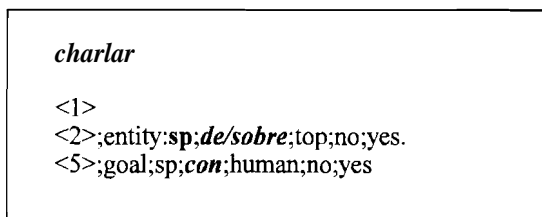


Figure 12: Basic diagram for verb "charlar"

*Last but not least, so as to obtain the diagram of specific alternations that can be manifested by a verb, the task in question is one of gathering both the information given by the verb and that provided by the alternations of the subclass. This would allow us to achieve the diagram in figure 13 for the impersonal alternation of "charlar":*

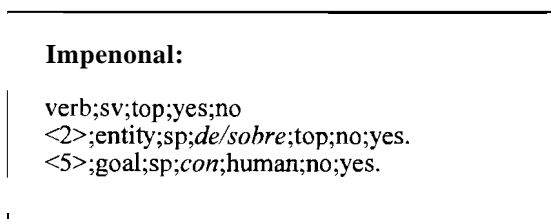


Figure 13: Impersonal alternation for the verb "charlar"

#### IV. RESULTS FROM THE EXTRACTION PROCESS

LEXPIR provides a semantic interpretation that is based on the information contained in it. This information is the output of the previously applied processes: the analysis module and the EWN consultation.

As a result of the final extraction process, the different diathesis interpretations are obtained, where each of them is also assigned the semantic tagging of the text in terms of meaning components. This can be seen in the following table:



LEXPIR: a Verb Lexicon

Basic model:

<b>LEXPIR</b>		<b>EWN</b>
initiator	<i>el invitado</i>	<i>Human</i>
event	<i>Trajo</i>	<i>Location Agentive</i>
entity	<i>una botella de vino</i>	<i>Artifact (Physical Comestible)</i>

Reflexive passive model:

<b>LEXPIR</b>		<b>EWN</b>
event	<i>se trasladan</i>	<i>Location Agentive</i>
entity	<i>los muebles</i>	<i>Artifact (Physical)</i>
goal	<i>a la casa</i>	<i>Building Object Group Occupation Part Place</i>

## V. CONCLUSIONS AND FURTHER WORK

The construction of a computational lexicon based on *Pirápidas*, i.e., on a theoretically founded model of lexical entry, allows a predicate analysis that can be used in an IE system: each argument holds an associated meaning component that plays a key role during the extraction process. The knowledge handled is purely linguistic and domain independent, which constitutes one of the main contributions of our system. The notion of diathesis as semantic oppositions that are linked to their syntactic structures establishes the connection between the phrase form (TACAT) and its semantic interpretation, within the extraction module.

To present, three classes have been defined, which comprise a total of 1.500 verbs, and the trajectory class has already been implemented in LEXPIR. Work is currently focusing on the implementation of the remaining classes in the LEXPIR DB. It seems, though, that the trajectory class is the most complex one, due to both the number of components constituting it and the differences between the groups that build it up. The design of the trajectory class has required a considerable amount of work, which will be already available when LEXPIR is extended to the other verb classes.

## Acknowledgements:

This research has been partially funded by the Spanish Research Department (CICYT's BASURDE project TIC98-0423-C06 and XTRACT project PB98-1226), by the MEC's grant AP98 39864555 and by the Catalan Research Department (CIRIT's consolidated research group 1999SGR-150).

## REFERENCES

- Alonge A., Bertagna. F., Bloksma. L., Climent. S., Peters. W., Rodríguez. H., Roventini. A., & Vossen. P. (1998): "The Top-Down Strategy for Building EuroWordNet: Vocabulary Coverage, Base Concepts and Top Ontology". *Computers and the Humanities*. Brussels.
- Arévalo. M., Martí. M.A. & Taulé. M. (2000): *MACO: un analizador morfológico del español*. Clic-W.P. 01/2000.
- Atserias. J., Carmona. J., Castellón. I., Cervell. S., Civit. M., Marquez. L., Martí. M.A., Padró. L., Placer. R., Rodríguez. H., Taulé M. & Turmo J. (1998): "Morphosyntactic Analysis and Parsing of Unrestricted Spanish Text". *Proceedings of the First International Conference on Language Resources and Evaluation (LREC '98)* Granada, Spain.
- Atserias. J. & Rodriguez. H. (1998): "Tacet: TAgged Corpus Analyzer Tool". LSI Technical Report (LSI-98-2-T).
- Basili. R., Pazienza. M.T. & Zanzotto. F.M. (1998): "Efficient Parsing for Information Extraction". *Proceedings of the 13<sup>th</sup> European Conference on Artificial Intelligence (ECAI '98)*.
- Briscoe. T. & Carroll. J., (1997): "Automatic Extraction of Subcategorization from Corpora". *Proceedings of ACL '97*.
- Castellón I., Civit M. & Atserias J., (1998): "Syntactic Parsing of Unrestricted Spanish Text". *Proceedings of LREC '98*. Granada, Spain.
- Civit. M., & Castellón. I. (1998): "Gramesp: Una gramática de corpus para el español". *RESLA*. La Rioja, Spain.

- Escudero. G., Márquez L. & Rigau. G. (3000): "Boosting Applied to Word Sense Disambiguation". *Proceedings of the European Conference on Machine Learning (ECML'00)*.
- Fellbaum, C. (Ed.): *WordNet: An Electronic Lexical Database*. MIT Press. Canibridge. M-A.
- Fernández. A. & Martí. M.A (1996): "Classification of psychological verbs". *Revista de la SEPLN*, no. 70.
- Fernández. A., Martí. M.A., Vázquez. G., & Castellón. I. (1999): "Establishing Semantic Oppositions for the Typification of Predicates". *Language Design*, 2.
- Lapata. M. (1999): "Acquiring Lexical Generalizations from Corpora: A Case Study for Diathesis Alternations". *Proceedings of the ACL'99*.
- Levin. B. (1993): *English Verb Classes and Alternations. Towards a Lexical Organization of English verbs*. Chicago University Press. Chicago.
- Fernández. A., Vázquez. G., Martí. M.A., Castellón. I (1999): "Los predicados de cambio y su representación en una BCL". revista de la SEPLN. n. 34.
- Morante. R., Castellón. I. & Vázquez. G. (1998): "Los verbos de trayectoria". *Revista de la SEPLN*, no. 73.
- Morante. R.(1999): "Adaptación de EuroWordNet para la enseñanza del español como lengua extranjera". *Actas del XIX Congreso de AESLA*. Universidad de Alcalá de Henares.
- Padró L. (1998): *A Hybrid Environment for Syntax-Semantic Tagging*. PhD Thesis. Departament de Llenguatges i Sistemes Informatics. Universitat Politècnica de Catalunya. Barcelona.
- Poznansky. V. & Sanfilippo. A.. (1991): "Detecting Dependencies between Semantic-Verb Subclasses and Subcategorisation Frames in Text Corpora". In Uri Zernik (Ed.): *Lexical Acquisition: Exploiting On-line Resources to Build a Lexicon*. LEA. Hilldale.
- Pustejovsky. J. (1995): *The Generative Lexicon*. Cambridge MIT Press. Massachusetts
- Turmo. J. (1997): "TURBIO: sistema de EI a partir de textos estructurados". *Revista de la SEPLN*, no. 21.
- Vazquez. G., Fernández. A., Castellón. I. & Martí. M.A. (7000): *Clasificación Verbal: alternancias de diátesis*. Edicions de la Universitat de Lleida.

Vossen, P. (Ed.) (1998): “. Special Issue on EuroWordNet”. *Computers and Humanities*, no. 32, Kluwer Academic Press.

Zernik, Uri (Ed.) (1991): *Lexical Acquisition: Exploiting On-line Resources to Build a Lexicon*. I.E.A. Hildale.