

# The IULA Spanish LSP Treebank

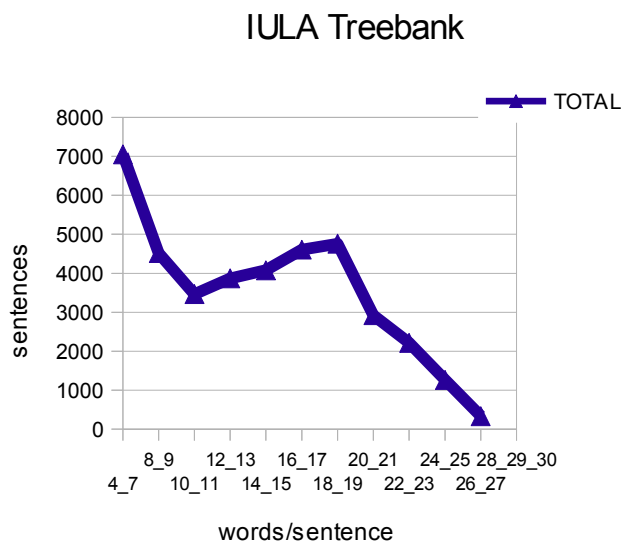
This document describes the linguistic annotations that the IULA Spanish LSP Treebank provides.

Contents:

1. The corpus
2. The annotation process
3. Representation of linguistic phenomena
  - 3.1. Complements and modifiers
  - 3.2. Clitics
    - 3.2.1. Cliticization
    - 3.2.2. Pronominal verbs
    - 3.2.3. Constructions with *se*
  - 3.3. Null subjects
  - 3.4. Elliptical NPs
  - 3.5. Elliptical finite verbs
  - 3.6. VP complements
  - 3.7. Coordination
4. References

## 1. The corpus

The IULA Spanish LSP Treebank contains 42,099 syntactically (dependencies) annotated sentences, distributed among different domains and sentence length, as an extension of the already existing IULA Technical Corpus (Vivaldi, 2009; Cabré et al. 2006), which is only PoS tagged.<sup>1</sup> Fig. 1 shows the ratio of number of sentences per sentence length in the treebank.



**Fig. 1** The IULA Spanish LSP Treebank, ratio of number of sentences per sentence length.

## 2. The annotation process

Following (Oepen et al, 2002), the corpus has been annotated with the publicly available corpus annotation environment of the *Deep Linguistic Processing with HPSG Initiative* (DELPH-IN),<sup>2</sup> also used in several treebank projects within this international initiative (Hashimoto et al, 2007; Kordoni and Zhang, 2009; Branco et al, 2010; Marimon 2010; Flickinger et al, 2012).

The corpus annotation environment in the DELPH-IN framework is based on the manual selection of the correct analysis among all the analyses that are produced by a hand-built symbolic grammar. The DELPH-IN framework also provides a Maximum Entropy (MaxEnt) based parse ranker that ranks the parses generated by the grammar, allowing the annotator to focus on the  $n$  most likely trees, and thus reducing the required annotation effort.

<sup>1</sup>The IULA Technical Corpus is a collection of written texts from the fields of Law, Economy, Genomics, Medicine, and Environment, and a contrastive corpus from the press. This corpus of 1,389 documents contains 31,436,451 words distributed among 412,707 sentences.

<sup>2</sup> <http://www.delph-in.net/>.

## 2.1. Parsing with HPSG

To parse the corpus the IULA Spanish LSP Treebank project uses the wide-coverage Spanish DELPH-IN grammar for deep processing: the *Spanish Resource Grammar* (SRG) (Marimon, 2012).

The SRG is grounded in the theoretical framework of *Head-driven Phrase Structure Grammar* (HPSG) (Pollard and Sag, 1987, 1994), a constraint-based lexicalist approach to grammatical theory, and it uses the *Minimal Recursion Semantics* (MRS) semantic representation (Copestake et al, 2006). The grammar is implemented in the *Linguistic Knowledge Builder* (LKB) system, an interactive grammar development environment for typed feature structure grammars (Copestake, 2002), based on an early version of the *LinGO Grammar Matrix* (Bender and Flickinger, 2005; Bender et al, 2010).

## 2.2. Disambiguation

The manual selection task has been performed using the `[incr tstb()]` profiling environment of the DELPH-IN framework (Oepen and Carroll, 2000).

Briefly, `[incr tstb()]` includes a tree comparison tool that allows the annotator to select the appropriate parse for each sentence directly, as it is displayed as a labeled phrase structure tree. When the grammar produces hundreds of analyses for a given sentence, the annotator can reduce the set of parses incrementally, through the choice of so-called discriminants (Carter, 1997); i.e., by selecting (or, alternatively, rejecting) the lexical or phrasal features that distinguish between the different parses, until the appropriate parse is left (or until the number of remaining choices allows the direct selection of the appropriate parse).

As it is always the case with symbolic grammars, the SRG produces several hundreds (or even thousands) of analyses for a corpus sentence. The DELPH-IN framework, however, provides a MaxEnt based stochastic ranker that sorts the parses produced by the grammar, thus allowing the annotator to reduce the forest to the  $n$ -best trees, typically to less than 500 top readings (Toutanova et al, 2005), and thus reducing the required annotation effort. Statistics are gathered from disambiguated parses and can be updated as the number of annotated sentences increases. In the IULA Spanish LSP Treebank, where the corpus was split into different files by sentence length, statistics are updated with each newly annotated file.

All the decisions made by the annotators are recorded in the database of the `[incr tsdb()]` profiling environment and will progressively enhance the stochastic system delivery of the requested  $n$ -best parses for a given sentence ranked as a prediction of the likelihood of being the right parse.

## 2.3. Linguistic annotations

The linguistic analysis produced by the LKB system for each parsed sentence combines the annotation of constituent structure in the form of a binary branching phrase structure tree, the annotation of structural semantics (predicate-argument relations) in the form of a MRS representation, and the annotation of dependency structure in the form of a derivation tree, extracted from a complete syntactico-semantic analysis represented in a parse tree with standard HPSG typed feature structures at each node.

The derivation tree is encoded in a nested, parenthesized structure whose elements correspond to identifiers of grammar rules and lexical items. Phrase structure rules --marked by the suffix `\_c` (for `construction')-- identify the daughter sequence, separated by a hyphen, and, in headed-phrase constructions, a basic dependency relation between them, namely: subject-head (sb-hd), head-complement (hd-cmp), head-adjunct (hd-ad), specifier-head (sp-hd), clitic-head (cl-hd), and filler-head (flr-hd). Lexical items are annotated with part-of-speech information according to the EAGLES tagset for Spanish<sup>3</sup> and their lexical entry identifier, and they optionally include an identifier of a lexical rule. Fig. 2 shows an example with sentence *El cuerpo humano irradia rayos de calor en todas las direcciones* ('The human body radiates heat beams in all directions.').

```
(sb-hd_c
  (sp-hd_c
    (da0ms0 (el_d "El"))
    (hd-ad_c
      (ncms000 (cuerpo_n "cuerpo"))
      (aq0ms0 (humano_a "humano"))))
  (hd-ad_c
    (hd-cmp_c
      (vmip3s0 (irradiar_v-np "irradiar"))
      (hd-nbar_c
        (hd-ad_c
          (ncmp000 (rayo_n "rayos"))
          (hd-comp_c
            (sps00 (de_p "de"))
            (hd-nbar_c
              (ncms000 (calor_n "calor"))))))))
    (hd-cmp_c
      (sps00 (de_p "en"))
      (sp-hd_c
        (sp-hd_c
          (di0fp0 (todo_d "todas"))
          (da0fp0 (el_d "las")))
        (hd-pt_c
          (ncfp000 (direccion_n "direcciones"))
          (fp (pt ".")))))))))
```

**Fig. 2** Derivation tree of *El cuerpo humano irradia rayos de calor en todas las direcciones* ('The human body radiates heat beams in all directions.').

From this derivation tree, we obtain the information for the dependency structures that the IULA Spanish LSP Treebank provides in two formats: (i) a theory-neutral column-based format, in the style of CoNLL-2006 shared task (Buchholz and Marsi, 2006), where sentence tokens are represented on one line, consisting of the seven fields that we describe in Table 1, and (ii) a graph dependency.

Dependencies are asymmetrical relations (except coordination) between single words: one word is always subordinated (dependent) to the other, called head. We have noted this relation using an oriented arrow, which goes from the dependent node to the head node which represents the governing element; e.g. the verb is considered the core of the sentence and the subject is taken to be dependent on the verb.

<sup>3</sup> See <http://www.ilc.cnr.it/EAGLES96/annotate/annotate.html>.

Fig. 3 shows the dependency structure that the treebank provides, both in the column-based format and as a graph dependency, for the sentence *El cuerpo humano irradia rayos de calor en todas las direcciones* ('The human body radiates heat beams in all directions.'). Table 2 and Table 3 show the complete set of dependencies labels and syntactic categories that are distinguished in the corpus, respectively.

Field number	Field name	Description
1	ID	Token counter, starting at 1 for each new sentence.
2	FORM	Word form
3	LEMMA	Lemma
4	CATEGORY	Syntactic category
5	PoS TAG	Part-of-speech Tag according to the EAGLES tagset
6	HEAD	Head of the current token
7	DEPENDENCY	Dependency relation to the HEAD

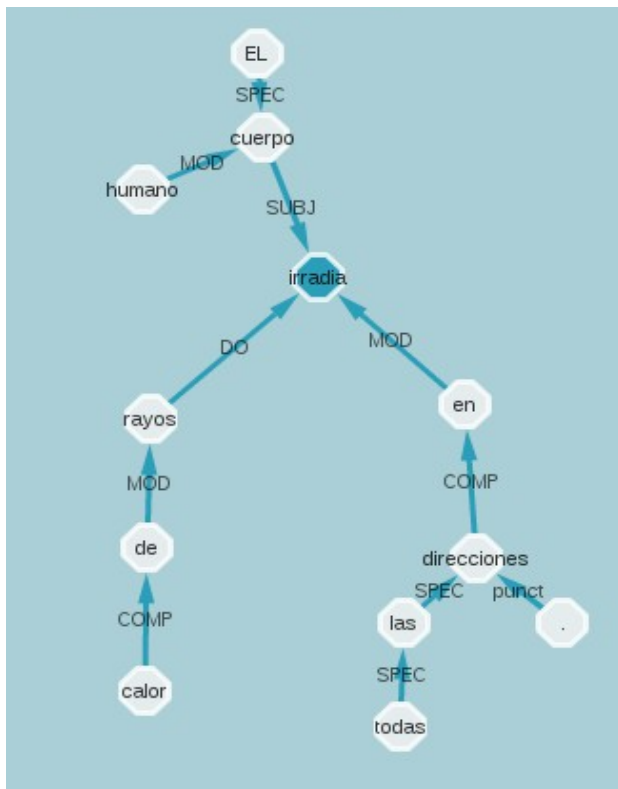
**Table 1**

Tag	Dependency
ROOT	Root
SUBJ	Subject
DO	Direct Object
IO	Indirect Object
OBLC	Oblique Object
BYAG	By agent complement
ATR	Attribute
PRD	Predicative complement
OPRD	Object predicative complement
PP-LOC	Locative prepositional complement
PP-DIR	Directional prepositional complement
SUBJ-GAP	Subject in a gapping construction
COMP-GAP	Complement in a gapping construction
MOD-GAP	Modifier in a gapping construction
VOC	Vocative
IMPM	Impersonal marker
PASSM	Passive marker
PRNM	Pronominal marker
COMP	Complement
MOD	Modifier
NEG	Negation
SPEC	Specifier
COORD	Coordination
CONJ	Conjunction
PUNCT	Punctuation

**Table 2** List of dependency labels of the IULA Spanish LSP Treebank.

Tag	Syntactic category
v	verb
n	noun
p	pronoun
a	adjective
r	adverb
s	preposition
d	determiner
c	conjunction
z	number
f	punctuation mark

**Table 3** List of syntactic categories of the IULA Spanish LSP Treebank.



ID	Forma	Lema	CPostag	FEATS	HEAD	DEPREL
1	EL	el	d	DA0MS0	2	SPEC
2	cuerpo	cuerpo	n	NCMS000	4	SUBJ
3	humano	humano	a	AQ0MS0	2	MOD
4	irradia	irradiar	v	VMIP3S0	0	ROOT
5	rayos	rayo	n	NCMP000	4	DO
6	de	de	s	SPS00	5	MOD
7	calor	calor	n	NCMS000	6	COMP
8	en	en	s	SPS00	4	MOD
9	todas	uno	d	DI0FP0	10	SPEC
10	las	el	d	DA0FP0	11	SPEC
11	direcciones	dirección	n	NCFP000	8	COMP
12	.	.	f	Fp	11	punct

**Fig. 3** *El cuerpo humano irradia rayos de calor en todas las direcciones* ('The human body radiates heat beams in all directions.').

### 3. Representation of linguistic phenomena

#### 3.1. Complements and modifiers

Dependency labels in the IULA Spanish LSP Treebank distinguish between syntactic complements and modifiers of the verb or verb phrase, and they also categorize the different types of verbal complements. The dependency labels for the verbal complements are shown in Table 3.

The IULA Spanish LSP Treebank also makes the distinction between complements and modifiers inside NPs, APs, PPs, and ADVPs, by labeling them COMP and MOD, respectively.

---

Tags	Grammatical functions
SUBJ	Subject
DO	Direct Object
IO	Indirect Object
OBLC	Oblique Object
BYAG	By agent complement
ATR	Attribute
PRD	Predicative complement
OPRD	Object predicative complement
PP-LOC	Locative prepositional complement
PP-DIR	Directional prepositional complement

---

Table 3 Dependency labels for the verbal complements.

#### 3.2. Clitics

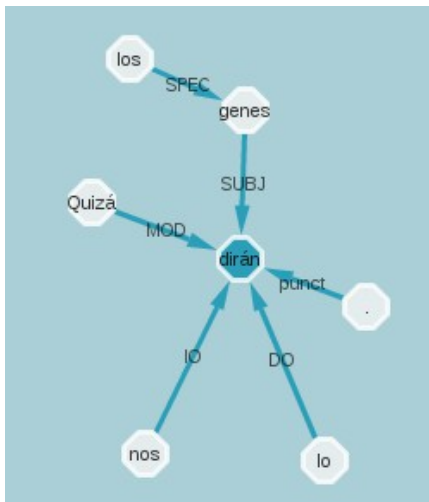
##### 3.2.1. Cliticization

Spanish clitic pronouns are unstressed object pronouns that appear adjacent to a host verb, either attached to its right, the so-called *enclitics*, or as independent lexical units in front of it, known as *proclitics*. Infinitives, gerunds, and non-negated imperatives have enclitic pronouns, verbs in personal forms always require proclitics, and past participles cannot have clitics.<sup>4</sup>

In the IULA Spanish LSP Treebank only proclitics are annotated. Here, the treebank distinguishes two different grammatical functions –direct object and indirect object– for proclitics which substitute verbal complements. Examples of proclitics and enclitics in the treebank are given in Fig. 4 (proclitics) and Fig. 5 (enclitics).

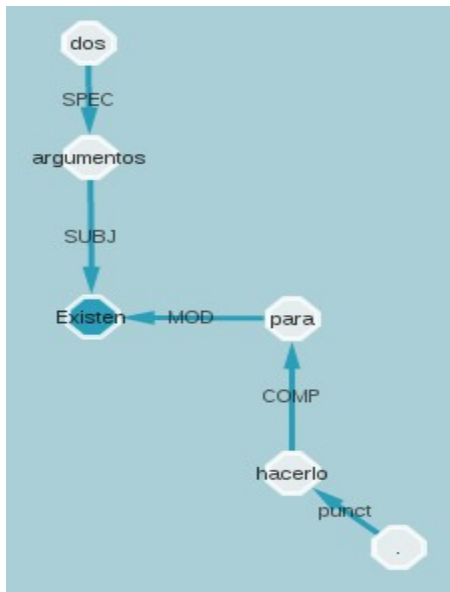
---

<sup>4</sup> In compound tenses, Spanish clitics must “climb” in the syntactic structure and they must appear as proclitics in front of the auxiliary verb *haber* (‘to have’). This phenomenon is referred to as clitic climbing. Clitic climbing can also occur with modal and aspectual verbs, subject-control verbs, causative verbs, and perception verbs. Thus, if one of these verb classes appears, the clitic may attach to the main verb or it may stay within the embedded verb.



ID	Forma	Lema	CPostag	FEATS	HEAD	DEPREL
1	Quizá	quizá	r	RG	6	MOD
2	los	el	d	DA0MP0	3	SPEC
3	genes	gen	n	NCMP000	6	SUBJ
4	nos	nos	p	PP1CP000	6	IO
5	lo	lo	p	PP3MSA00	6	DO
6	dirán	decir	v	VMIF3P0	0	ROOT
7	.	.	f	Fp	6	punct

**Fig. 4** *Quizá los genes nos lo dirán*' ('Perhaps genes will tell us').



ID	Forma	Lema	CPostag	FEATS	HEAD	DEPREL
1	Existen	existir	v	VMIP3P0	0	ROOT
2	dos	Z	z	Z	3	SPEC
3	argumentos	argumento	n	NCMP000	1	SUBJ
4	para	para	s	SPS00	1	MOD
5	hacerlo	hacer	v	VMN0000	4	COMP
6	.	.	f	Fp	5	punct

**Fig. 5** *Existen dos argumentos para hacerlo* (There are two reasons for doing it).

Unlike French and Italian, where clitics and full phrases are considered to be in strict complementary distribution within the clause, Spanish clitic pronouns may also appear together with the complement they refer to, in what is known as *clitic doubling* constructions. For clitic doubling, enclitics are assigned the same grammatical function as the complement they refer to.



### 3.2.2. Pronominal verbs

The clitic pronouns *me*, *nos*, *te*, *os*, and *se* can also appear with so-called *inherent reflexive* verbs (or *pronominal verbs*); i.e., verbs which require a clitic pronoun co-indexed with the subject and which lack the corresponding non-reflexive form

In the IULA Spanish LSP Treebank these clitics are marked as MPRON (i.e., pronominal marker) as illustrated in Fig. 6 with the sentence *A ello me referiré en la parte final de mi exposición* (I will refer to it in the last part of my presentation).

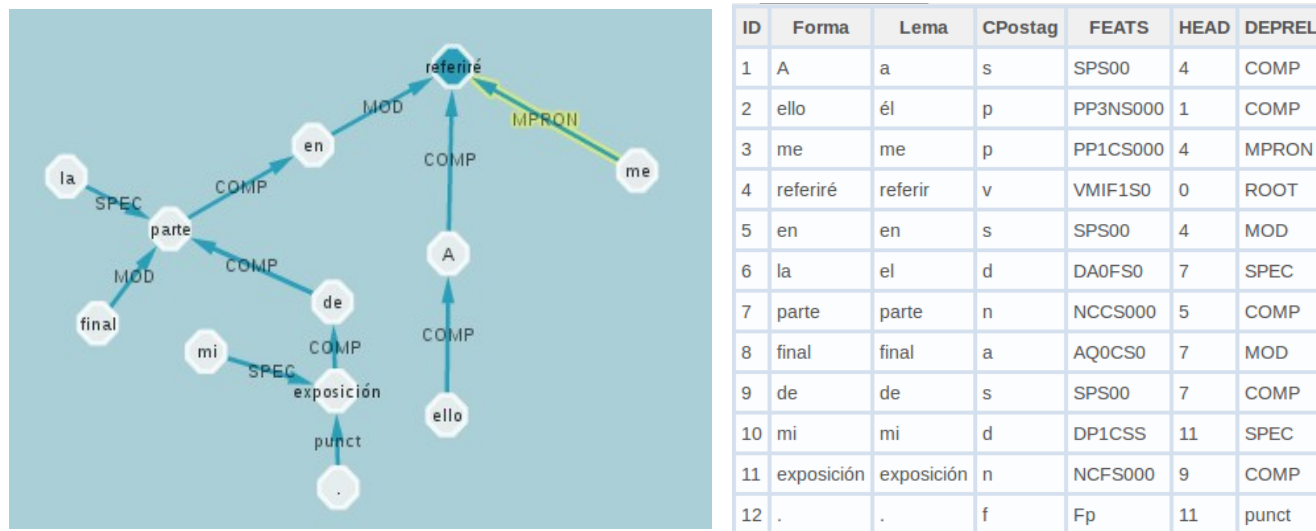


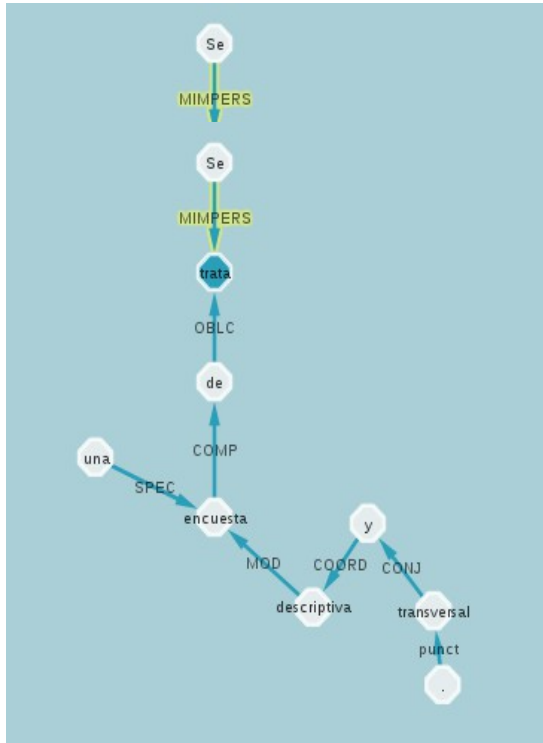
Fig. 6 *A ello me referiré en la parte final de mi exposición* (I will refer to it in the last part of my presentation).

### 3.2.3. Constructions with *se*

In Spanish, the form *se* can also appear in the so-called *impersonal* and *passive se-constructions*. In these constructions, a verb concurs with the clitic *se* which is not a verbal argument, but a grammatical marker.

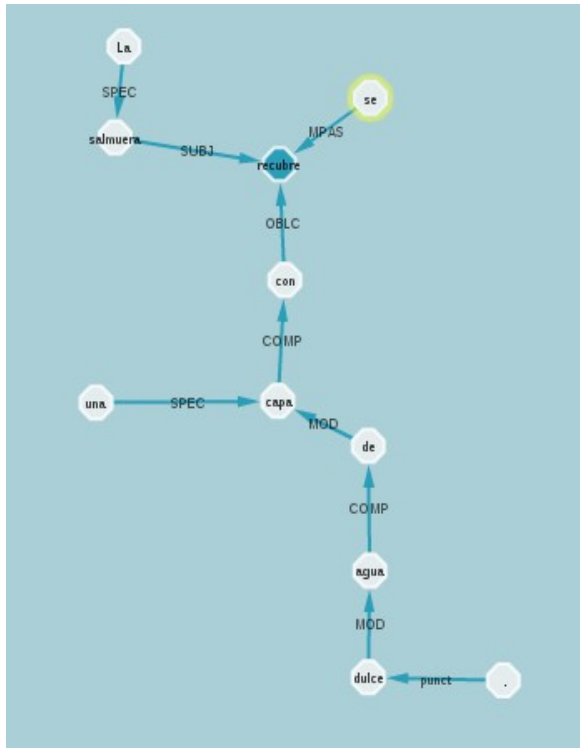
In *passive constructions* the verb has a unique argument which is the syntactic subject. This construction can only appear with transitive verbs. Unlike passives, *impersonal constructions* do not have an overt subject and the verb appears in third singular person. Another difference is that this construction can appear not only with transitive verbs, but also with intransitive verbs, unaccusative verbs, and verbs taking sentential complements.

The IULA Spanish LSP Treebank makes the distinction between these two usages of the grammatical marker *se*, which is labeled as MIMPERS (i.e., impersonal marker) in impersonal constructions (Fig. 7), and MPAS (i.e., passive marker) in passive constructions (Fig. 8).



ID	Forma	Lema	CPostag	FEATS	HEAD	DEPREL
1	Se	se	p	P03CN000	2	MIMPERS
2	trata	tratar	v	VMIP3S0	0	ROOT
3	de	de	s	SPS00	2	OBLC
4	una	una	z	Z	5	SPEC
5	encuesta	encuesta	n	NCFS000	3	COMP
6	descriptiva	descriptivo	a	AQ0FS0	5	MOD
7	y	y	c	CC	6	COORD
8	transversal	transversal	a	AQ0CS0	7	CONJ
9	.	.	f	Fp	8	punct

Fig. 7 *Se trata de una encuesta descriptiva y transversal* (It's a descriptive and transversal survey).



ID	Forma	Lema	CPostag	FEATS	HEAD	DEPREL
1	La	el	d	DA0FS0	2	SPEC
2	salmuera	salmuera	n	NCFS000	4	SUBJ
3	se	se	p	P03CN000	4	MPAS
4	recubre	recubrir	v	VMIP3S0	0	ROOT
5	con	con	s	SPS00	4	OBLC
6	una	una	z	Z	7	SPEC
7	capa	capa	n	NCFS000	5	COMP
8	de	de	s	SPS00	7	MOD
9	agua	agua	n	NCCS000	8	COMP
10	dulce	dulce	a	AQ0CS0	9	MOD
11	.	.	f	Fp	10	punct

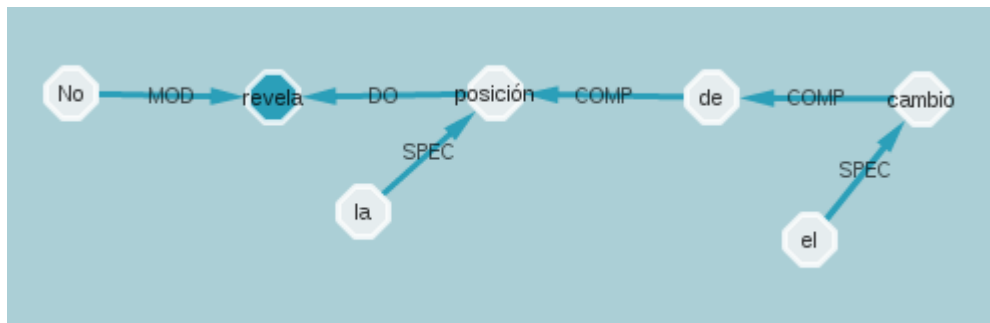
Fig 8. *La salmuera se recubre con una capa de agua dulce* (Brine is covered with a layer of freshwater).

### 3.3. Null subjects

Being a pro-drop language, Spanish frequently omits explicit subjects in finite clauses where the information about the person and number of the subject is encoded in the affix of the verb.

Fig. 9 illustrates the dependency structure that the treebank provides for null subjects with the sentence *No revela la posición del cambio* (It does not reveal the change position). As it can be observed, no elliptical element with the syntactic function subject is inserted, since only dependencies between actual words in the sentence are marked.

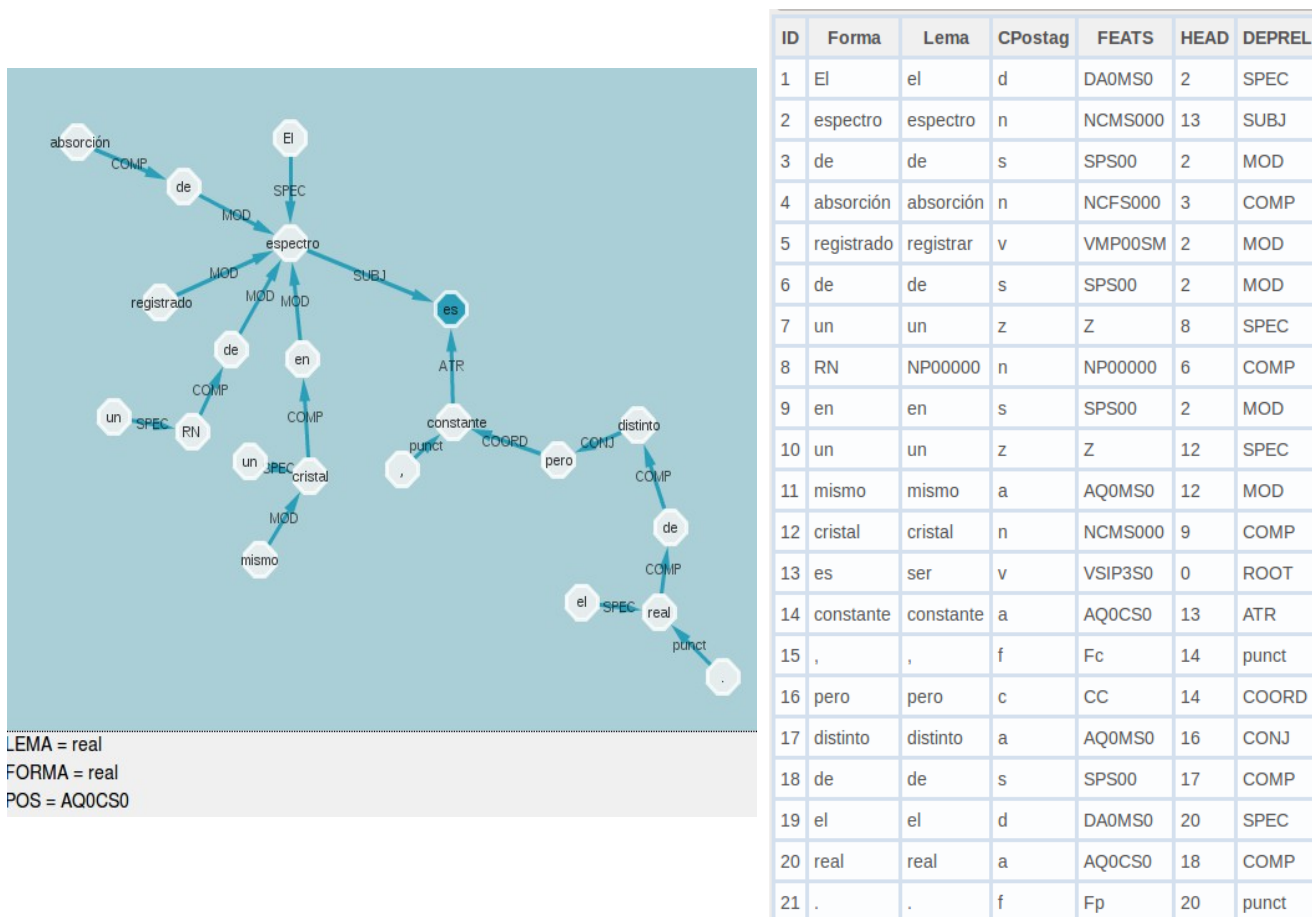
ID	Forma	Lema	CPostag	FEATS	HEAD	DEPREL
1	No	no	r	RN	2	MOD
2	revela	revelar	v	VMIP3S0	0	ROOT
3	la	el	d	DA0FS0	4	SPEC
4	posición	posición	n	NCFS000	2	DO
5	de	de	s	SPS00	4	COMP
6	el	el	d	DA0MS0	7	SPEC
7	cambio	cambio	n	NCMS000	5	COMP



**Fig. 9** *No revela la posición del cambio* (It does not reveal the change position).

### 3.4. Elliptical NPs

As can be observed in Fig. 10, no elliptical element is inserted for marking elided nominal heads, and the IULA Spanish LSP Treebank follows the standard strategy used to deal with empty heads in dependency corpora: the modifier of the elided head is chosen to become the head and it is labeled with the syntactic function of the elided head. So, in the example, the adjective in the elliptical NP (i.e. *real*) is labeled as COMP of the preposition.

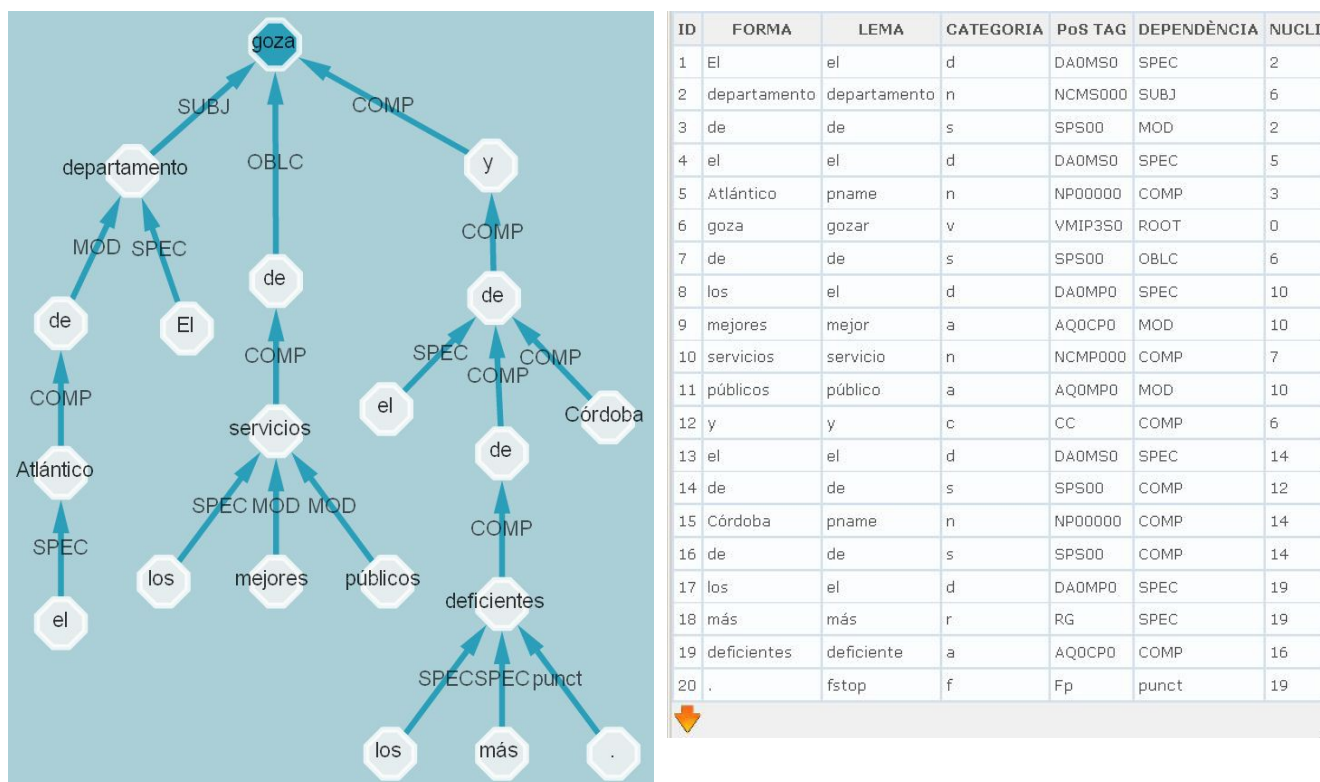


**Fig 10** *El espectro de absorción registrado de un RN en un mismo cristal es constante, pero distinto de el real.*

### 3.5. Elliptical finite verbs

This section describes the annotations that the IULA Spanish LSP Treebank offers for two types of coordinated constructions where the verb is missing from the second conjunct: sentence gapping and conjunction reduction (or argument cluster coordination).

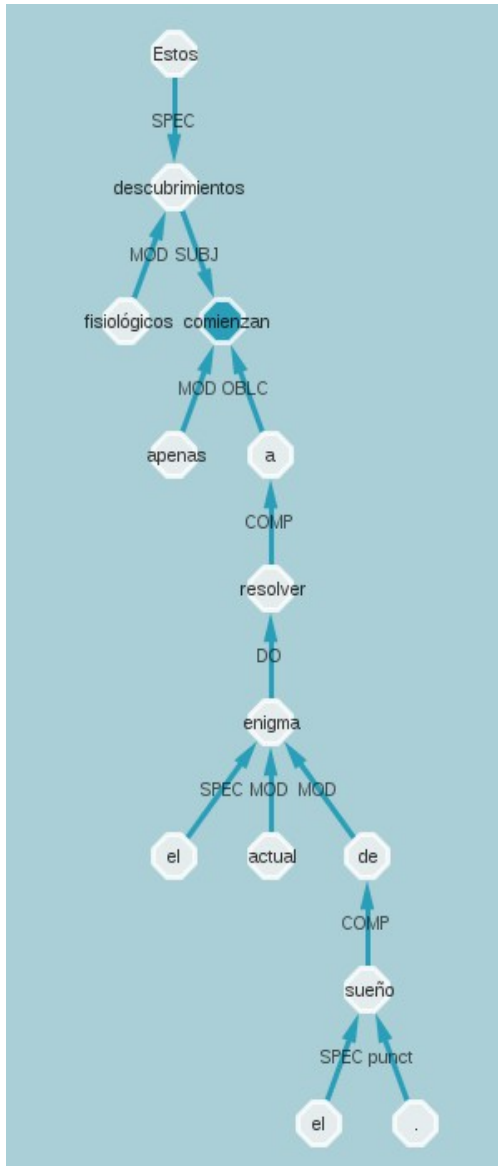
In these constructions, the parts of the second conjunct are attached to the conjunction, and the subject, complement, and modifier dependents carry a SUBJ\_GAP, COMP\_GAP, and MOD\_GAP label. An example is given in Fig. 11 with the sentence *El departamento del Atlántico goza de los mejores servicios públicos y el de Córdoba de los más deficientes* (The Atlantic department enjoys the best public services and the Cordoba department the most deficient).



**Fig. 11** *El departamento del Atlántico goza de los mejores servicios públicos y el de Córdoba de los más deficientes* (The Atlantic department enjoys the best public services and the Cordoba department the most deficient).

### 3.6. VP complements

For VP complements, no elliptical element is inserted to identify the subject of the infinitive, as can be observed in Fig.12.

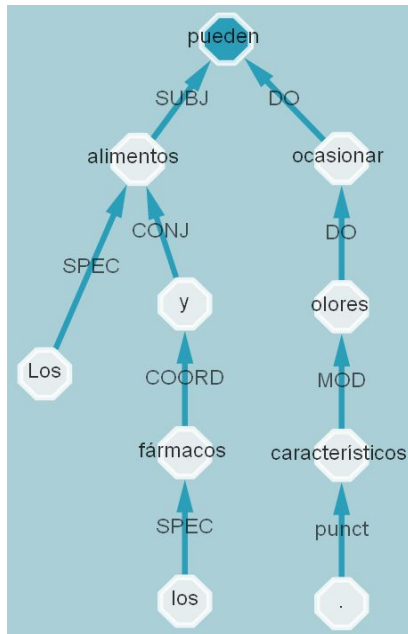


ID	Forma	Lema	CPostag	FEATS	HEAD	DEPREL
1	Estos	este	d	DD0MP0	2	SPEC
2	descubrimientos	descubrimiento	n	NCMP000	5	SUBJ
3	fisiológicos	fisiológico	a	AQ0MP0	2	MOD
4	apenas	apenas	r	RG	5	MOD
5	comienzan	comenzar	v	VMIP3P0	0	ROOT
6	a	a	s	SPS00	5	OBLC
7	resolver	resolver	v	VMN0000	6	COMP
8	el	el	d	DA0MS0	9	SPEC
9	enigma	enigma	n	NCMS000	7	DO
10	actual	actual	a	AQ0CS0	9	MOD
11	de	de	s	SPS00	9	MOD
12	el	el	d	DA0MS0	13	SPEC
13	sueño	sueño	n	NCMS000	11	COMP
14	.	.	f	Fp	13	punct

**Fig. 12** *Estos descubrimientos fisiológicos apenas comienzan a resolver el enigma actual del sueño* (These physiological discoveries are scarcely beginning to solve the actual sleep enigma).

### 3.7. Coordination

The IULA Spanish LSP Treebank follows the standard approach used to deal with coordination in dependency corpora: the first conjunct is treated as the head of the coordinated structure, the coordinating conjunction is the head of the second conjunct using the COORD label, and the second conjunct is linked to the conjunction via a CONJ dependency label.



ID	FORMA	LEMA	CATEGORIA	PoS TAG	DEPENDÈNCIA	NUCLI
1	Los	el	d	DA0MP0	SPEC	2
2	alimentos	alimento	n	NCMP000	SUBJ	6
3	y	y	c	CC	CONJ	2
4	los	el	d	DA0MP0	SPEC	5
5	fármacos	fármaco	n	NCMP000	COORD	3
6	pueden	poder	v	VMIP3P0	ROOT	0
7	ocasionar	ocasionar	v	VMN0000	DO	6
8	olores	olor	n	NCMP000	DO	7
9	característicos	característico	a	AQ0MP0	MOD	8
10	.	fstop	f	Fp	punct	9

**Fig. 13** *Los alimentos y los fármacos pueden ocasionar olores característicos* (Food and drugs can produce characteristic odours).

#### 4. References

- Bender EM, Flickinger D (2005) Rapid prototyping of scalable grammars: towards modularity in extensions to a language-independent core. In: *Proceedings of IJCNLP'05 (Posters / Demos)*, Jeju Island, Korea, pp 203–208
- Bender EM, Drellishak S, Fokkens A, Poulson L, Saleem S (2010) Grammar Customization. *Research on Language & Computation* 8(1):23–72.
- Branco, António, Francisco Costa, João Silva, Sara Silveira, Sérgio Castro, Mariana Avelãs, Clara Pinto, and João Graca. 2010. Developing a Deep Linguistic Databank Supporting a Collection of Treebanks: the CINTIL DeepGramBank. In *Proceedings of LREC-2010*, La Valletta, Malta.
- Buchholz, Sabine and Erwin Marsi. 2006. CoNLL-X Shared Task on Multilingual Dependency Parsing. In *Proceedings of CoNLL-X*, New York City, USA.
- Cabré, M. T., C. Bach, & J. Vivaldi. (2006). 10 anys del Corpus de l'IULA. Barcelona: Institut Universitari de Lingüística Aplicada. Universitat Pompeu Fabra.
- Carter, David. 1997. The TreeBanker: A tool for supervised training of parsed corpora. In *Proceedings of the Fourteenth National Conference on Artificial Intelligence*, pages 598–603, Providence, Rhode Island.
- Copestake, Ann. 2002. *Implementing Typed Feature Structure Grammars*. CSLI Publications, Stanford.
- Copestake, Ann, Dan Flickinger, Carl Pollard, and Ivan A. Sag. 2006. Minimal Recursion Semantics: an Introduction. *Research on Language and Computation*, 3(4):281–332.
- Flickinger, Dan, Valia Kordoni, Yi Zhang, António Branco, Kiril Simov, Petya Osenova, Catarina Carvalheiro, Francisco Costa, and Sérgio Castro. 2012. ParDeepBank: Multiple Parallel Deep Treebanking. In *Proceedings of The 11th International Workshop on Treebanks and Linguistic Theories*, pages 97–108, Lisbon, Portugal.
- Hashimoto, Chikara, Francis Bond, and Melanie Siegel. 2007. Semi-automatic documentation of an implemented linguistic grammar augmented with a treebank. *Language Resources and Evaluation*. (Special issue on Asian language technology), 42(2):117–126.
- Kordoni, Valia and Yi Zhang. 2009. Annotating Wall Street Journal Texts Using a Hand-Crafted Deep Linguistic Grammar. In *Proceedings of LAW III*, Singapore.
- Marimon, Montserrat. 2010. The Tibidabo Treebank. *Procesamiento del Lenguaje Natural*, 45:113–119.
- Marimon, Montserrat. 2012. The Spanish DELPH-IN Grammar. *Language Resources and Evaluation*, in press.
- Oepen, Stephan and John Carroll. 2000. Performance Profiling for Parser Engineering. In Flickinger, Dan and Stephan Oepen and Junichi Tsujii and Hans Uszkoreit, editor, *Natural Language Engineering (6)1 —Special Issue: Efficiency Processing with HPSG: Methods, Systems, Evaluation*. Cambridge University Press, pages 81–97.
- Pollard, Carl and Ivan A. Sag. 1987. *Information-based Syntax and Semantics. Volume I: Fundamentals*. CSLI Lecture Notes, Stanford.



Pollard, Carl and Ivan A. Sag. 1994. *Head-driven Phrase Structure Grammar*. The University of Chicago Press and CSLI Publications, Chicago.

Toutanova, Kristina, Christopher D. Manning, Dan Flickinger, and Stephan Oepen. 2005. Stochastic HPSG parse disambiguation using the Redwoods corpus. *Research on Language and Computation*, 3(1):83–105.

Vivaldi, Jorge. 2009. Corpus and exploitation tool: IULACT and bwanaNet. In *A survey on corpus-based research (CICL-09)*, *Asociación Española de Lingüística del Corpus*, pages 224–239.