

**AnCora:  
Argument Structure Guidelines  
for Catalan and Spanish**

Taulé, M., Aparicio, J., Martí, M.A. & Borrega, O.

CLIC – CENTRE DE LENGUATGE I COMPUTACIÓ  
UNIVERSITAT DE BARCELONA

July 2009

## Table of contents

<b>1. Introduction</b> .....	<b>2</b>
<b>2. Basic semantic event classes</b> .....	<b>2</b>
<b>3. Thematic roles and argument structure</b> .....	<b>3</b>
<b>4. Lexical Semantic Structures and diathesis alternations</b> .....	<b>5</b>
4.1. LSS1 (A): accomplishments ([+dynamic], [+telic]) .....	6
4.2. LSS2 (B): achievements ([-dynamic], [+telic]) .....	7
4.3. LSS3 (C): states ([-dynamic], [-telic]) .....	8
4.4. LSS4 (D): activities ([+dynamic], [-telic]) .....	9
4.5. Special diatheses: impersonal and causative alternations .....	10
<b>5. The lexicon and the corpora: Annotation criteria</b> .....	<b>11</b>
5.1. AnCora lexicon .....	11
5.2. AnCora corpora: annotation criteria .....	14
5.2.1. Verbs and verb phrases. LSS tag (<els>) .....	14
5.2.2. Syntactic complements of verbs .....	15
<b>6. References</b> .....	<b>17</b>

## 1. Introduction

This manual presents the guidelines for the annotation of argument structure of verbal predicates and their semantic class of the Spanish and Catalan AnCora corpora. The semantic annotation of verbal predicates implies the systematic mapping between syntax and semantics, basically expressed in the argument structure. This mapping ultimately motivates the semantic classes. In this proposal, each verbal predicate was assigned to a specific semantic class and every syntactic function was tagged with both arguments and thematic roles (Aparicio *et al.*, 2008 and Taulé *et al.*, 2008). The semantic properties used were defined assuming lexical decomposition (Levin & Rappaport Hovav, 1995; and Rappaport Hovav & Levin, 1998) from which the concept of Lexical Semantic Structure (LSS) was taken. The LSS as well as the kind of diatheses alternations in which the predicate can participate, determines the number of arguments that a verbal predicate requires and the thematic role of these arguments. In this line, we followed the guides laid down by Kipper *et al.* (2002) and Kingsbury *et al.* (2002) in the construction of *VerbNet*.

We consider the proposal presented by Levin and Rappaport Hovav to be appropriated for our work mainly for two reasons. First, because in their model converge lexical semantic, event and argument information and diathesis alternations. And, second, because similar works in corpus and computational linguistics have been carried out following this approach, such as *VerbNet*, a lexicon with lexical semantic, argument and diathesis information for English predicates. *VerbNet* follows Levin's semantic classification and adopts *PropBank* semantic annotation (Palmer *et al.*, 2005).

We characterize predicates by means of a limited number of LSS and Event Structure Patterns, according to the four basic event classes: states, activities, accomplishments, and achievements (Vendler, 1967 and Dowty, 1991). These general classes can be split into subclasses, as we will see in section 4. Semantic roles are determined by the event class the predicate belongs to and by the type of diathesis alternation the predicate presents. Thus, not only thematic roles are assigned, but also predicates are characterized both from the aspectual and from the argumental perspective. In fact, the semantic classes determine the mapping between syntactic functions and semantic roles.

This information is currently being stored in the lexicon **AnCora-Verb** (Aparicio *et al.*, 2008) for both languages, which is, in practice, our annotation guide. Section 2 will deal with the definition of the four basic event classes. In section 3, we will discuss the thematic roles adopted and in section 4 we will talk about the LSS adopted and the possible diathesis alternations they can participate in. Section 5 will explain in greater detail the annotation criteria we have followed, and will illustrate with examples the composition of AnCora-Verb lexicon and the annotation of AnCora corpora.

## 2. Basic semantic event classes

Four general semantic classes have been laid down which can be further subclassified depending on thematic roles and diatheses. In the definition of these main classes only Argument0 and Argument1 have been taken into consideration, because they are the basic arguments involved in the definition of predicate structures (they typically correspond to the syntactic Subject and Direct Object functions). These criteria give raise to a coarse grained classification which has been later on specified by splitting each general class into subclasses (LSS). This subclassification has not been developed in as much detail as the thematic role assignment, since, although it may be very useful, mapping thematic roles into syntactic functions is not the main goal of this methodology.

Following Vendler (1967) and Dowty (1991), we assume there are four ontological event classes: states, activities (or processes), accomplishments and achievements:

- (1) [x <STATE>]
- (2) [x ACT <MANNER/ INSTRUMENT> y]
- (3) [x CAUSE [BECOME [y <STATE/THING/PLACE>]]]
- (4) [BECOME [y <STATE>]]]

The general frame depicted by (1) corresponds to the ontological class *state*, with just one entity involved in the event, which focuses in the state. The frame in (2) corresponds to *activities* (or *processes*), and usually presents agentive subjects and passive objects, thus allowing passive constructions. The frame in (3) corresponds to *accomplishments* that refer to resulting states in external cause processes, and usually presents causative subjects and allows anti-causative constructions. Finally, the frame in (4) corresponds to *achievements* that refer to a resulting state in processes without external cause<sup>1</sup>.

The lexical decomposition of a predicate in the form of a LSS contains three basic components: the semantic primitives, the constants and the variables. The semantic primitives correspond to the components CAUSE, BECOME and ACT, which determine both the basic meaning of the verb and the event type. The constants (*MANNER*, *INSTRUMENT*, *STATE*, etc.) express the idiosyncratic aspect of the verb meaning and are represented in italics. The variables (*x* and *y*) represent the arguments that the verb needs to be syntactically expressed. As we will see, LSS determines the number of arguments a verbal predicate requires and the thematic role of these arguments, and restricts the set of all possible diatheses.

In order to finish this section, it must be pointed out that each one of these general semantic classes is defined also in terms of telicity and dynamicity. Telicity is the property of a verb or verb phrase that presents an event as being complete or completable in some sense. Dinamicity is the property of a verb or verb phrase that presents an event as implying activity in some sense. Thus, *states* are defined as [-dynamic] and [-telic]; *activities* are defined as [+dynamic] and [-telic]; *accomplishments* are defined as [+dynamic] and [+telic]; and, finally, *achievements* are defined as [-dynamic] and [+telic]. We have not brought under consideration the trait [±punctual].

### 3. Thematic roles and argument structure

The semantic relation that each argument maintains with the event denoted by the verb is defined by the thematic roles. We have adopted a set of 19 different thematic roles, each one of them able to mapping to several syntactic functions and argument positions. A complete list of thematic roles with their corresponding label is shown in (5).

- (5) adverbial (ADV), agent (AGT), attribute (ATR), beneficiary (BEN), cause (CAU), destination (DES), final state (EFI), initial state (EIN), experiencer (EXP), extension (EXT), purpose (FIN), instrument (INS), location (LOC), manner (MNR), origin (ORI), patient (PAT), source (SRC), theme (TEM), time (TMP).

For the arguments, we have followed the proposal of PropBank (Palmer *et al.*, 2005), where the arguments required by the verb sense are incrementally numbered, expressing their degree of proximity in relation to its predicate. Thus, we distinguish between seven possible argumental slots: Arg0, Arg1, Arg2, Arg3, Arg4, ArgM and ArgL. The first five tags are numbered from less to more obliqueness with respect to the verb. ArgM corresponds to adjuncts, and ArgL codes lexicalized complements of light verbs<sup>2</sup>.

<sup>1</sup> In Catalan and Spanish there are two types of passive constructions: passives with auxiliary verb *ser* (to be) plus the main verb in participle form and passives with the pronoun *se* (*Esta mañana han sido vendidos cinco libros – Esta mañana se han vendido cinco libros* ‘Five books have been sold this morning’). The pronoun *se* may as well be used in anti-causative constructions (*La puerta se abrió*).

<sup>2</sup> *PropBank* uses one more tag (ArgA) to indicate the inductive agent. We have decided to leave it out

Argument structure is determined by LSS. Depending on event structure and diathesis alternations, arguments might appear in different syntactic positions, and thematic roles might appear in different argument slots. Table (6) shows all possible combinations of argument, thematic role and syntactic function, with examples in Spanish.

Argument	Θ-Role	Syntactic Function	Example
Arg0	AGT	Subject (SUJ)	<b>Juan</b> lee una novela
		Agent complement (CAG)	Clara es amada <b>por todos</b>
	CAU	Subject (SUJ)	<b>El viento</b> abrió la puerta
	EXP	Subject (SUJ)	<b>Juan</b> sueña
	SRC	Subject (SUJ)	<b>Juan</b> sudaba
Arg1	EXT	Direct object (CD)	Juan caminó <b>3 kilómetros</b>
	PAT	Direct object (CD)	Juan lee <b>una novela</b>
		Subject (SUJ)	<b>Clara</b> es amada por todos
	TEM	Direct object (CD)	El viento abrió <b>la puerta</b>
		Subject (SUJ)	<b>Juan</b> llegó tarde
	LOC	Prepositional object (CREG) <sup>3</sup>	Juan hurgaba <b>en la tierra</b>
Arg2	ATR	Attribute (ATR)	Juan es <b>listo</b>
		Prepositional object (CREG)	Una docena equivale <b>a doce unidades</b>
		Predicative (CPRED)	La película se titula <b>Casablanca</b>
		Adjunct (CC)	Prefiero permanecer <b>en el anonimato</b>
		Direct object (CD)	Juan tiene <b>un coche blanco</b>
	BEN	Indirect object (CI)	Juan aconsejó <b>a Clara</b> que lo hiciera
	LOC	Adjunct (CC)	<b>El periodista</b> aborda la noticia <b>en el diario</b>
		Subject (SUJ)	<b>El diario</b> abordó la noticia
		Prepositional object (CREG)	Juan entró <b>en la habitación</b>
	EFI	Adjunct (CC)	Juan entró <b>en coma</b>
		Predicative (CPRED)	Juan se nacionalizó <b>español</b>
		Prepositional object (CREG)	Juan convirtió su pasión <b>en profesión</b>
	EXP	Indirect object (CI)	<b>A Juan<sub>i</sub></b> , no <b>le<sub>i</sub></b> gusta correr
	INS	Adjunct (CC)	Juan cubrió su coche <b>con una lona</b>
		Subject (SUJ)	<b>Una lona</b> cubre el coche de Juan
Prepositional object (CREG)		Juan va equipado <b>con todo lo necesario</b>	
EXT	Direct object (CD)	El paro subió <b>15.891 personas</b>	
	Prepositional object (CREG)	El paro se situó <b>en 923.426 personas</b>	
	Adjunct (CC)	El paro subió <b>en 15.891 personas</b>	
Arg3	LOC	Adjunct (CC)	El presidente informó a la ciudadanía de su decisión <b>en una nota de prensa</b>
		Subject (SUJ)	<b>Una nota de prensa del presidente</b> informó de su decisión
	ORI	Adjunct (CC)	Juan viajó a Tokio <b>desde Barcelona</b>
		Prepositional object (CREG)	La luz provenía <b>de una bombilla</b>
	BEN	Indirect object (CI)	<b>A Juan<sub>i</sub></b> , se <b>le<sub>i</sub></b> ha quedado pequeña la casa
	EIN	Adjunct (CC)	Juan subió su oferta <b>de 3000 euros a 4000</b>
	FIN	Prepositional object (CREG)	El faro nos sirve <b>de guía</b>
	EXP	Indirect object (CI)	Su actitud no <b>me</b> parece adecuada
ATR	Predicative (CPRED)	Juan pasó varios días <b>enfadado</b>	
	Adjunct (CC)	Juan fue recibido <b>con aplausos</b>	
INS	Subject (SUJ)	<b>Los aplausos</b> recibieron su llegada	
Arg4	EFI	Adjunct (CC)	Juan subió su oferta de 3000 <b>a 4000 euros</b>
	DES	Adjunct (CC)	Juan volvió <b>a casa</b> desde el trabajo a pie
		Prepositional object (CREG)	Su curiosidad le llevó <b>a descubrirlo</b>
	ADV	Adjunct (CC)	Juan leía la carta <b>de nuevo</b>
	LOC	Adjunct (CC)	Juan leía la carta <b>en el jardín</b>
	ORI	Adjunct (CC)	Le inculcaron la disciplina <b>desde la cuna</b>
	TMP	Adjunct (CC)	Juan leía la carta <b>a las 3 de la tarde</b>

and, instead, resolve these cases syntactically.

<sup>3</sup> We have not specified the thematic role of most prepositional objects, due to the fuzziness of their semantic behavior. See section 5 for more information on this subject.

ArgM	FIN	Adjunct (CC)	Juan leía la carta <b>para entenderla</b>
	MNR	Adjunct (CC)	Juan leía la carta <b>en voz alta</b>
	ATR	Predicative (CPRED)	Juan leía la carta <b>ansioso</b>
	INS	Adjunct (CC)	Juan leía la carta <b>con una lupa</b>
	EXT	Adjunct (CC)	Juan leyó la carta <b>siete veces</b>
	CAU	Adjunct (CC)	Juan leía la carta <b>porque le gustaba</b>
ArgL	No $\theta$ -Role	Attribute (ATR)	La suerte está <b>echada</b>
		Adjunct (CC)	Se puso <b>en evidencia</b> él solo
		Direct object (CD)	Puso <b>punto final</b> a la discusión
		Prepositional object (CREG)	Se lo dio <b>a entender</b> sutilmente
		Predicative (CPRED)	Se puso la camisa <b>perdida de aceite</b>
		Subject (SUJ)	<b>El cadáver</b> fue levantado a la 1 de la tarde

#### 4. Lexical Semantic Structures and diathesis alternations

LSS determines the number of arguments that a verbal predicate requires and the thematic role of these arguments, and describes the syntactic function of said arguments. In our model, each LSS restricts the set of all possible diatheses<sup>4</sup> it can incur in, and each verb sense is associated to one LSS. Diathesis alternations are the result of focusing different components of the LSS they belong to. That is to say, diatheses are surface structures that result from focusing different components of the predicate LSS. Diatheses must be understood as the syntactic expression of a semantic opposition.

Furthermore, the expression of most alternations entails an aspectual change, which necessarily implies a change of semantic class. As an example, let us consider the following sentences:

- (7) Juan **corre** (*Agentive-ingerative – activity*)  
 [Juan runs] – 1 argument
- Juan **corre** los 100 metros (*Transitive-agentive – accomplishment*)  
 [Juan runs the 100-meters] – 2 arguments
- Se **corren** los 100 metros (*Unnacusative-state – achievement*)  
 [The 100-meters are run] – 2 arguments (1 is elliptic)
- Se **corre** los 100 metros (*Impersonal-state – achievement*)  
 [The 100-meters are run] – 1 argument
- Hice **correr** a Juan los 100 metros (*Transitive-causative – accomplishment*)  
 [I made Juan run the 100-meters] – 3 arguments
- Se le **hizo correr** los 100 metros (*Unnacusative-impersonal – achievement*)  
 [He was forced to run the 100-meters] – 2 arguments
- Corridos** los 100 metros, Juan descansó (*Resultative-attributive – state*)  
 [Once the 100-meters were over, Juan relaxed] – 1 argument

All sentences in (7) denote the same event (Someone –Juan– ran the 100-meters, maybe forced by someone else –me, in the example- or by something). But each one of the sentences focuses in a different component of the basic event, thus changing both aspectual changes (from ingergativity to passivity, resultativity, etc.) and semantic class (we can see at least one sentence belonging to each one of the four main semantic event classes described in section 2). This example is very illustrative about the degree in which a verb belonging to a determined semantic class may move towards other semantic classes under certain syntactic conditions.

As we have already said, our proposal of classification is coarse grained. We have only considered productive diatheses. Specific alternations shared by few verbs have been left out because they do not define general classes. This improves general robustness and coherence.

Next, we will present the specific LSS derived from the general semantic event classes

<sup>4</sup> We follow in essence the diathesis classification of Vázquez *et. al* (2000). It must be said, nevertheless, that the causative and the impersonal alternations are possible for all LSS.

discussed above. These LSS are the result of combining the general class with the argument structure and the thematic roles that can fulfill each argument slot. Each verbal class is also characterized for admitting certain diatheses alternations. For each verbal sense, its semantic class is established, and the mapping between syntactic functions<sup>5</sup>, argument structure and thematic roles is declared.

There are 13 LSS compiled and described, grouped around the 4 general event classes. These 13 LSS derive from the analysis of 1462<sup>6</sup> verbs in AnCora corpora. On the basis of a draft of the annotation guide, annotator agreement tests have been carried out. In a first step, 70 verbs have been studied and tagged by five annotators in parallel, and in three phases (10, 30 and 30 verbs in each phase). After annotating each group an agreement discussion was carried out in order to revise and update the annotators guide. Once the guidelines were established, in a second step, 400 verbs were annotated by two pairs of annotators, each pair working in parallel with the same set of verbs. For these pairs of annotators the agreement rate was of 95% and 96%, respectively. This agreement rate has been obtained by confronting the results of the mapping between functions and thematic roles of one member of the pair against the other. The remaining 4% and 5% of disagreement has been discussed and the annotator guide modified when necessary. Almost all cases of disagreement are related to sense discrimination (assignment of LSS) and the identification of verbal forms, for instance, when it is necessary to decide if a given structure corresponds to a verb and its complements or to an idiom (*dar + un susto* vs. *dar\_un\_susto*, ‘to fright’). The analysis of the remaining 1000 verbs has been done by the annotators independently.

#### 4.1 LSS1 (A): accomplishments ([+dynamic], [+telic])

This general event structure is subdivided into three subclasses: transitive-causative (A1), transitive-agentive (A2) and ditransitive-agentive (A3), which is further subdivided into ditransitive-agentive-locative (A31) and ditransitive-agentive-benefactive (A32). All four LSS share the resultative alternation and the passive alternation.

LSS corresponding to accomplishments are composed by the combination of the semantic predicates CAUSE, DO and BECOME in a complex event structure involving a causing subevent and a change of state or location subevent.

##### A1: transitive-causative class

[x CAUSE [BECOME [y <STATE>]]]

Arg0=CAU

Arg1=TEM

Diatheses<sup>7</sup>: [+anticausative(B2)<sup>8</sup>][+impersonal(=)][+resultative(C2)][+/-passive(B2)]

[+benefactive(=)]

Spanish verbs: *romper* ‘to break’, *abrir* ‘to open’, *cerrar* ‘to close’, *hundir* ‘to sink’...

Catalan verbs: *afectar* ‘to affect’, *convertir* ‘to turn into’, *omplir* ‘to fill’...

Transitive-causative verbs (A1 class), such as *romper* ‘to break’, associate the causer argument (*x*) with the semantic predicate CAUSE and the participant that undergoes the change with the argument (*y*). Since these verbs participate in the inchoative and resultative alternations, *x* is referred to as *Arg0-Cause* and *y* as *Arg1-Theme*. Arg0 is syntactically the subject, while Arg1 is syntactically the direct object.

<sup>5</sup> We extracted the verbal syntactic frames from the corpus as it has been described in Taulé *et al.* (2005) and Civit *et al.* (2005).

<sup>6</sup> The total amount of verbs in the lexicon is: 2807 (Spanish) and 2144 (Catalan).

<sup>7</sup> We indicate the new class the verb will belong to when it undergoes the alternation between round brackets. The equal sign means no change in class.

<sup>8</sup> Anticausative alternation is also known as ergative or inchoative alternation.

## **A2: transitive-agentive class**

[[x DO-SOMETHING]CAUSE[BECOME[y<STATE>]]]

Arg0=AGT

Arg1=PAT

Diatheses: [+causative(A1)][+impersonal(=)][+/-resultative(C2)][+passive(B2)]

[+/-intransitive(D1)][+/-oblique subject(=)][+/-benefactive(=)]

Spanish verbs: *escribir* ‘to write’, *barrer* ‘to sweep’, *leer* ‘to read’, *visitar* ‘to visit’...

Catalan verbs: *afirmar* ‘to affirm’, *decidir* ‘to decide’, *guanyar* ‘to win’...

Transitive-agentive verbs (A2 class), such as *escribir* ‘to write’, associate the causer argument (*x*) with the semantic predicate DO, and since they allow the passive alternation, the argument *x* is referred to as *Arg0-Agent* and the argument *y* as *Arg1-Patient*. Arg0 is syntactically the subject, while Arg1 is syntactically the direct object.

## **A31: ditransitive-agentive-locative class**

[[x DO-SOMETHING]CAUSE[BECOME[y<PLACE>z]]]

Arg0=AGT

Arg1=PAT

Arg2=LOC

Diatheses: [+causative(A1)][+impersonal(=)][+/-resultative(C2)][+passive(B1)][+/-oblique subject(=)]

Spanish verbs: *poner* ‘to put’, *almacenar* ‘to store’, *publicar* ‘to issue’...

Catalan verbs: *incorporar* ‘to include’, *moure* ‘to move’, *traslladar* ‘to transfer’...

Ditransitive-agentive-locative verbs (A31 class), such as *poner* ‘to put’, associate the causer argument (*x*) with the semantic predicate DO, and the third argument (*z*) with a location in space. Since they allow passive alternation, the argument *x* is referred to as *Arg0-Agent* and the argument *y* as *Arg1-Patient*. Argument *z* is referred to as *Arg2-Location*. Arg0 is syntactically the subject, while Arg1 is syntactically the direct object. Arg2 may syntactically be an adjunct or a prepositional object.

## **A32: ditransitive-agentive-benefactive class**

[[x DO-SOMETHING]CAUSE[BECOME[y<PLACE>z]]]

Arg0=AGT

Arg1=PAT

Arg2=BEN

Diatheses: [+causative(A1)][+impersonal(=)][+/-resultative(C2)][+passive(B1)]

Spanish verbs: *enviar* ‘to send’, *dar* ‘to give’, *decir* ‘to say’, *robar* ‘to rob’...

Catalan verbs: *explicar* ‘to explain’, *permetre* ‘to allow’, *vendre* ‘to sell’...

Ditransitive-agentive-benefactive verbs (A32 class), such as *enviar* ‘to send’, associate the causer argument (*x*) with the semantic predicate DO, and the third argument (*z*) with a location in space with semantic traits such as [+animate] or [+human]. Since they allow passive alternation, the argument *x* is referred to as *Arg0-Agent* and the argument *y* as *Arg1-Patient*. Argument *z* is referred to as *Arg2-Beneficiary*. Arg0 is syntactically the subject, while Arg1 is syntactically the direct object and Arg2 is syntactically the indirect object.

## **4.2 LSS2 (B): achievements ([−dynamic], [+telic])**

This general event structure is subdivided into two subclasses: unaccusative-motion (B1) and unaccusative-state (B2), depending on the constant they associate with (either PLACE or STATE). Unaccusative verbs are basically monadic in terms of their LSS and in terms of their argument structure, take a single internal argument (Arg1).

Achievements are associated with a simple event structure which lacks the causing subevent that characterizes accomplishments.

The representation of the argument structure allows some distinctions to be made between the internal and the external argument of a verb. Internal arguments are expressed in the syntax projected inside the verb phrase (VP), whereas, external arguments are expressed external to the VP headed by the verb selecting those arguments. Unaccusativity is related to the fact that the grammatical subject of an unaccusative verb behaves as the direct object of a transitive verb, consequently, the subject of an accusative verb and the object of a transitive verb bear the same semantic role: *Theme*, and occasionally *Patient*.

### **B1: unaccusative-motion class**

[BECOME[y<PLACE>]]

Arg1=TEM/PAT

Arg2=LOC/BEN

Diatheses: [+causative(A1)][+impersonal(=)][+/-object extension(A2)][+/-resultative(C2)]

Spanish verbs: *llegar* ‘to arrive’, *ir* ‘to go’ *salir* ‘to go\_out’, *venir* ‘to come’...

Catalan verbs: *baixar* ‘to go\_down’, *caure* ‘to fall’, *entrar* ‘to go\_in’, *pujar* ‘to go\_up’...

Arg1 is syntactically the subject, while Arg2 may syntactically be an adjunct or a prepositional object (when fulfilling the role *location*) or the indirect object (when fulfilling the role *beneficiary*). Also, when being a passive alternation of a verb from LSS A31 or A32, an Arg0 may be present with the thematic role of *agent* and the syntactic function of agent complement.

### **B2: unaccusative-state class**

[BECOME[y<STATE>]]

Arg1=TEM/PAT

Arg2=EFI/BEN

Diatheses: [+causative(A1)][+impersonal(=)]

Spanish verbs: *crecer* ‘to grow’, *floreecer* ‘to bloom’...

Catalan verbs: *créixer* ‘to grow’, *trencar-se* ‘to get broken’, *enfonsar-se* ‘to collapse’...

Arg1 is syntactically the subject, while Arg2 may syntactically be an adjunct, a prepositional object or a predicative (when fulfilling the role *final state*) or the indirect object (when fulfilling the role *beneficiary*). Also, when being a passive alternation of a verb from LSS A1 or A2, an Arg0 may be present with the thematic role of *agent* and the syntactic function of agent complement.

## **4.3 LSS3 (C): states ([-dynamic], [-telic])**

This general event structure is subdivided into four subclasses: existence (C1), attributive (C2), scalar (C3) and benefactive (C4). In terms of their argument structure, stative verbs take two arguments. On the one hand, they take an internal argument (*Arg1*), which appears as syntactic subject bearing the semantic role *Theme*. On the other hand, they take an *Arg2*, whose thematic role gives rise to the four verbal classes.

### **C1: existence class**

[x<STATE>y]

Arg1=TEM

Arg2=LOC

Diatheses: [+causative(A1)][+impersonal(=)][+benefactive(=)]

Spanish verbs: *estar* ‘to be’, *existir* ‘to exist’...

Catalan verbs: *haver-hi* ‘there\_is/are’, *existir* ‘to exist’...

Arg1 is syntactically the subject. Arg2 maps to a *locative* and is syntactically an adjunct or a prepositional complement, if expressed at all. It is very common for verbs belonging to this class to show just one explicit argument. Auxiliary verb *haver* (existential ‘to be’ or ‘to have’)

in its use as main verb belongs to this class.

### **C2: attributive class**

[x<STATE>y]

Arg1=TEM

Arg2=ATR

Diatheses: [+causative(A1)][+impersonal(=)][+/-ablative(=)][+benefactive(=)]

Spanish verbs: *ser* 'to be', *parecer* 'to seem', *tener* 'to have' ...

Catalan verbs: *estar* 'to be', *tenir* 'to have' ...

Arg1 is syntactically the subject, while Arg2 is syntactically an attribute, typically, or a direct object (for such verbs as *tener* –to have). As for thematic roles, Arg2 maps to an *attribute*. Semiauxiliary verbs such as *ser* or *estar* in their most common sense belong to this class.

### **C3: scalar class**

[x<STATE>y]

Arg1=TEM

Arg2=EXT

Diatheses: [+causative(A1)][+impersonal(=)][+benefactive(=)]

Spanish verbs: *medir* 'to measure', *pesar* 'to weight', *valer* 'to cost' ...

Catalan verbs: *costar* 'to cost', *durar* 'to last', *pesar* 'to weight' ...

Arg1 is syntactically the subject while Arg2 may either be a direct object, an adjunct or a prepositional object. Arg2 maps with *extension* thematic role, an argument referring to some sizable and measurable magnitude such as length, weight, time, price, etc. Notice that, even if verbs belonging to this class appear to be transitive –they may accept a direct object complement–, passive alternation is not possible.

### **C4: benefactive class**

[x<STATE>y]

Arg1=TEM

Arg2=BEN/EXP

Diatheses: [+causative(A1)][+impersonal(=)]

Spanish verbs: *gustar* 'to like', *doler* 'to hurt' ...

Catalan verbs: *agradar* 'to like', *preocupar* 'to worry' ...

Arg1 is syntactically the subject while Arg2 is syntactically an indirect object and maps to a beneficiary or an experiencer thematic role. Verbs belonging to this class are few in number, but some of them present a very high use frequency.

## **4.4 LSS4 (D): activities ([+dynamic], [–telic])**

This general event structure is subdivided into three subclasses: agentive-ingerative (D1), experiencer-ingerative (D2) and source-ingerative (D3). As can be seen, activities are related to ingerative verbs, a set of verbs that in terms of their LSS are basically monadic and in terms of their argument structure take a single external argument, whose thematic role determines verb class. Ingerative verbs' only semantic predicate (ACT) denotes an acting entity (x) that does something.

### **D1: agentive-ingerative class**

[x ACT <MANNER/INSTRUMENT>]

Arg0=AGT

Diatheses: [+causative(A1)][+impersonal(=)][+/-object extension(A2)]

Spanish verbs: *correr* 'to run', *caminar* 'to walk', *nadar* 'to swim' ...

Catalan verbs: *jugar* 'to play', *navegar* 'to sail', *treballar* 'to work' ...

Arg0 is syntactically the subject, and its thematic role is *agent*. Verbs belonging to this class may be made transitive by adding a direct object, as the diatheses alternations point out.

### D2: experiencer-ingerative class

[x ACT <MANNER/INSTRUMENT>]

Arg0=EXP

Diatheses: [+causative(A1)][+impersonal(=)][+cognate object(A2)]

Spanish verbs: *dormir* ‘to sleep’, *soñar* ‘to dream’...

Catalan verbs: *dormir* ‘to sleep’, *respirar* ‘to breathe’...

Arg0 is syntactically the subject, and its thematic role is *experiencer*. Verbs belonging to this class may be made transitive by adding a direct object whose semantic content is normally already present in the verb (like in *dream a dream*).

### D3: source-ingerative class

[x ACT <MANNER/INSTRUMENT>]

Arg0=SRC

Diatheses: [+causative(A1)][+impersonal(=)][+cognate object(A2)]

Spanish verbs: *roncar* ‘to snore’, *sudar* ‘to sweat’...

Catalan verbs: *cridar* ‘to shout’, *plorar* ‘to cry’...

Arg0 is syntactically the subject, and its thematic role is source. Verbs belonging to this class may be made transitive by adding a direct object whose semantic content is somehow already present in the verb (like in *cry tears*).

## 4.5 Special diatheses: impersonal and causative alternations

To end up this section, a word on these two special diathesis alternations must be made. It has already been said above that causative and impersonal alternations are possible for all LSS. Let us further explain this fact.

All verb classes (and almost all verbs) admit impersonal alternation by means of the addition of the pronoun *se*<sup>9</sup>. What this pronoun normally does is to reduce the valence of the verb in one argument. Thus, a verb having three arguments in its argument structure will have only two if accompanied by *se*. That is to say, the presence of this pronoun blocks the apparition either of the subject or of the direct object (typically). When blocking the apparition of the direct object, passive or anticausative alternations are found, and it only affects transitive verbs (LSS1), changing their class. When blocking the apparition of the subject (be it an internal or an external argument), any LSS may be affected, but there is no change in verb class. Let us present some examples.

(8) Cuando **se llega**<sub>(B2)</sub> tarde, no **se pide**<sub>(A32)</sub> explicaciones ni **se pone**<sub>(A2)</sub> excusas  
[When one is late, one does not ask for explanations nor give excuses]

En el cine sólo **se llora**<sub>(D3)</sub> si **se es**<sub>(C2)</sub> sensible  
[One only cries in the cinema if one is sensitive]

On the other hand, all verbs in Spanish and Catalan accept causative alternation by means of the addition of the light (semiauxiliary) verbs *hacer/fer* or *dejar/deixar* plus the main verb in infinitive form. Even causative verbs (LSS A1) may participate in this kind of constructions. Verb class is then changed to transitive-causative (A1), with a causer argument external to the verb (Arg0-CAU, the syntactic subject), a theme argument internal to the verb (Arg1-TEM, the

<sup>9</sup> This pronoun has a wide variety of uses both in Spanish and Catalan. For further information on this subject, see the syntactic annotation guide to AnCora corpora: <http://clic.ub.edu/ancora/>.

syntactic direct object) and two more optional arguments, a beneficiary one (Arg2-BEN, the syntactic indirect object) and an instrumental one (Arg3-INS, an adjunct). In these cases, the semantic subject of the main verb (the participant in the event which actually carries out the action expressed by the verb) is the one expressed in the Arg2-BEN argument, when present, or in the Arg1-TEM. The causer is sometimes called ‘inducer agent’. Let us present an example.

(9) [unmarked ditransitive sense]:

Pedro<sub>(Arg0-AGT)</sub> **dijo**<sub>(A32)</sub> algunas cosas feas<sub>(Arg1-PAT)</sub> a causa de la actitud de Juan<sub>(ArgM-CAU)</sub>.  
Pedro said some ugly things due to Juan’s attitude.

[+causative alternation]:

Con su actitud<sub>(Arg3-INS)</sub>, Juan<sub>(Arg0-CAU)</sub> **hizo decir**<sub>(A1)</sub> a Pedro<sub>(Arg2-BEN)</sub> algunas cosas feas<sub>(Arg1-TEM)</sub>.  
[With his attitude, Juan made Pedro say some ugly things].

Apart from this basic causative alternation, a causative-reflexive alternation is also possible for most verb classes, constructed with pronoun *se*, plus a semiauxiliary causative verb (*hacer/fer* or *dejar/deixar*), plus the main verb in infinitive. The most remarkable feature of this causative-reflexive diathesis is that causer and theme coincide in the same argument (the subject –inducer agent). In these cases, we have decided to annotate the subject as Arg0-CAU, considering it an external argument, and not as Arg1-TEM, because we do not want to raise the ambiguity between reflexivity and anticausativity. Therefore, causative-reflexive verb phrases may only have two arguments: Arg0-CAU and Arg2-INS. Let us illustrate this with an example.

(9) [unmarked transitive sense]:

**Engañó**<sub>(A2)</sub> al árbitro<sub>(Arg1-PAT)</sub> con palabras bonitas<sub>(ArgM-INS)</sub>.  
[(He) deceived the referee with fair words].

[+causative-reflexive alternation]

El arbitro<sub>(Arg0-CAU)</sub> **se dejó engañar**<sub>(A1)</sub> con palabras bonitas<sub>(Arg2-INS)</sub>.  
[lit.: The referee let himself be deceived with fair words].

## 5. The lexicon and the corpora: Annotation criteria

Data in the lexicon are stored in UTF-8 encoded plain format files. In the corpora, data are stored in UTF-8 encoded XML format. LSS are annotated as an attribute of the node <verb>. Arguments, thematic roles and functions are annotated as an attribute each of the corresponding nodes of the complements. Each attribute describes only one feature of the node and may have only one value. Each attribute affects only the node it accompanies, and not its descendant/ancestor/sibling nodes.

For the file names in the lexicon, special characters have been avoided in order to prevent format or encoding conflicts. Thus,  $\tilde{n}$  is written down as *n* (or *nn* when ambiguity is possible), *l·l* is written down as *l-l*,  $\zeta$  is written down as *c*, and accents have been left out. Reflexive pronoun *se* and other clitics are separated from the infinitive they accompany by -, while \_ is used to write down multi-words. For instance:

(10) *banar.txt* stands for *bañar*

*sonnar.txt* stands for *sonar* (due to the possible ambiguity with *sonar*)

*al-lucinar.txt* stands for *al-lucinar*

*amenacar.txt* stands for *amenazar*

*abstenir-se.txt* (with reflexive pronoun *se*)

*echar\_en\_falta.txt* (multi-word)

*anar-se-n\_a\_fer\_punyetes.txt* (multi-word with several clitics)

### 5.1 AnCora lexicon

The guide for annotating the corpus is AnCora lexicon (<http://clic.ub.edu/ancora/>), which has been evolving and growing throughout all the process (stretching over two years long). Nowadays, AnCora lexicon is composed of 2807 files for Spanish and 2144 files for Catalan.

Each file contains the information relating to one verb in all its senses and alternations in plain text format. The name of the file corresponds to the verb in its infinitive form (the enunciation form).

Each verb sense is assigned a number. For each sense, its LSS is declared, followed by argument structure and a list of examples ranging from one to dozens, and finally a list of diathesis alternations (when possible) which share the format (LSS declaration, argument structure, examples) with the main sense. All information present in the lexicon is documented in the corpora unless stated otherwise. There is no limit to the senses a file can contain, or to the number of diathesis alternations a sense may present. Just to illustrate, the verb with the most senses, *fer* ('to do' in Catalan), has 101 senses which add up to 123 diatheses.

Argument structure is declared as follows: first, the syntactic function is specified, followed by a tabulator. Next, a row of three fields separated by the pound symbol (#) contains: first, argument type and number; second, preposition or prepositions governed; third, thematic role of the argument. The preposition and thematic role spaces may be empty. More than one option for prepositions and thematic roles may be declared, separated by a slash (/). Let us see an example of an actual complete file of the lexicon, corresponding to the Catalan verb *veure* ('to see').

```
(11) veure.txt
veure - 01
ELS:a2
SUJ   Arg0##AGT
CD    Arg1##PAT
CPRED ArgM##ATR
CC    ArgM#en/sobre#TMP/amb/per#MNR/a/en#LOC

EX:   "que poden arribar a veure afectada en un 100_% la_seva plantilla"
EX:   "**0* va veure indicis que feien pensar que la nena podia haver estat víctima d'abusos
sexuals"

+PASSIVA
ELS:b2
SUJ   Arg1##PAT
CAG   Arg0#per#AGT
CPRED ArgM##ATR
CC    ArgM#a/de/en#TMP/a#LOC

EX:   "dels 40 més vistos"
EX:   "que es poden veure"
EX:   "El recital es podrà veure a les vuit_de_la_tarda a la Biblioteca_Pública de la capital del
Pallars_Jussà"

+IMPERSONAL
ELS:a2
CD    Arg1##PAT
CPRED ArgM##ATR
CC    ArgM##LOC

EX:   "s' hi veu reflectida la_seva sensibilitat"

+BENEFACTIVA
ELS:a2
SUJ   Arg0##AGT
CD    Arg1##PAT
CI    Arg2#a#BEN

EX:   "cap productor li ha vist cara de fer gràcia"
```

veure - 02  
ELS:c2  
SUJ Arg1##TEM  
CPRED Arg2##ATR  
CC ArgM##LOC/#TMP

EX: "la comarca es pot veure greument desprotegida"  
EX: "s' hi han vist implicats un jaguar amb matrícula provisional S-5881-BBG , conduït per Fèlix\_S.\_G , i un Opel\_Astra amb matrícula GI-2909-AV"

veure - 03  
ELS:a2  
SUJ Arg0##AGT  
CD Arg1##PAT

+PASSIVA  
ELS:b2  
SUJ Arg1##PAT  
CC ArgL##

EX: "el cas Banesto va quedar vist per a sentència"

veure - 04  
ELS:a1  
SUJ Arg0##CAU  
CD Arg1##TEM  
CI Arg2#a#BEN

EX: "El setge a què em va sotmetre Rafael\_Ribó em va fer veure que potser sí\_que \*0\* només pensava en mi"

veure - 05  
ELS:a2  
SUJ Arg0##AGT  
CD ArgL##  
CREG Arg1#amb#

EJ: "els corruptes que tinguin un se les hauran de veure amb el fiscal general"

This file is composed of five different senses. Notice that diathesis alternations affect only the sense they follow. Notice as well that thematic role slots may be empty, and prepositions are only declared for prepositional objects, indirect objects and other oblique complements when deemed necessary. Notice also that alternations may change the verb's semantic class, although it is not compulsory. Finally, notice that when there are no examples of a particular complement or sense in the corpus, they are not included in the file, although an empty sense may serve for creating an alternation which would be impossible otherwise (in the example, sense 03).

File names in the lexicon may be simple words (infinitives) or multi-words (an infinitive followed by a complement or set of complements), such as the Catalan verbal expression *treure foc pels queixals* ('to be extremely angry'; literally: 'to spit fire out of one's teeth'). The similitude between these multi-words and lexicalized arguments (labelled as ArgL) is great. It has been difficult to decide when to annotate one way or the other. The determining factor is the degree of lexicalization of the expression (the more lexicalized, the more it is likely to become a multi-word). That means that normally lexicalized arguments (ArgL) allow for some degree of variation, while multi-words do not. Multi-words are lemmatized and treated as a single word (complements included in the expression have no function, they are not considered arguments, they do not receive thematic role and the LSS is assigned to the whole multi-word).

As for the annotation procedure, the lexicon has been the annotation guide, as has already been said. For every verb instance in the corpus, the corresponding lexicon file was consulted. If

the sense and diathesis was already present in the file, the information contained was mapped to the corpus, thus encoding verbal LSS and argument type and number of the complements, along with their thematic role. If the sense or diathesis was not present in the file, it was included, respecting the format we have explained, and the annotation was carried out as normal. If considered informative enough, the sentence containing the particular verb instance was also included as an example (all verb senses and alternations in the lexicon which are present in the corpora must include at least one example).

## 5.2 AnCora corpora: annotation criteria

Next, we summarize the guides for the annotation of semantic elements (LSS, arguments and thematic roles) in the corpora. We will specify which nodes were to be annotated, which tags they should bear, and which annotation exceptions are to be observed. We will begin by describing the annotation of verbs and verb phrases and, next, we will move on to the annotation of syntactic complements of verbs.

### 5.2.1 Verbs and verb phrases. LSS tag (<els>)

There is a set of verbal elements which receive LSS annotation, including most verbs and some adjectives. Here is complete list of such elements with examples. In the examples, the whole verb phrase is underlined, but only the word in bold letters gets the tag:

- a) inflected verbs in their synthetic tense forms (*no acabarem fins que no arribi el material* – we will not finish until the material arrives)
- b) non personal forms in periphrases (the inflected auxiliary does not get its LSS):
  - a. periphrastic tense forms (*no vàrem acabar fins que el material no va arribar* – we did not finish until the material arrived)
  - b. passive voice with *ser* (*el treball fou acabat quan els materials van ser enviats* – the work was finished when the materials were sent)
  - c. aspectual periphrases<sup>10</sup> (*acabàvem d'acabar el treball quan el material va començar a arribar* – we had just finished the work when the material started arriving)
  - d. causative alternation with *hacer/fer* or *dejar/deixar* (*els vaig fer acabar el treball després de deixar arribar el material* – I made them finish the work after letting the material arrive)
- c) infinitives when appearing as a main verb (in non personal completive clauses) (*acabar el treball va ser complicat sense el material* – finishing the work was difficult without the material)
- d) gerunds when appearing as a main verb (in non personal adverbial clauses) (*enviant el material, ens van permetre d'acabar el treball* – by sending the material, they let us finish the work)
- e) participles when appearing as a main verb:
  - a. in absolute clauses (past participle is then labeled as verb-participle) (*acabat el treball, ja no calia esperar el material* – once the work was finished, there was no need to wait for the material)
  - b. in non personal adjective clauses with explicit complements (past participle is then labeled as adjective-participle) (*el material, enviat a darrera hora, va arribar just a temps* – the material, sent in the last minute, arrived just in time)

On the other hand, there is still a set of verbal elements which are not labeled with LSS tags:

- a) auxiliary verbs in periphrastic constructions (see examples above)
- b) verbs participating in verbless sentences or clauses (*el material va ser enviat des del distribuïdor i rebut a la fàbrica sense problemes* – the material was sent from the dealer and received at the factory without problems)

<sup>10</sup> A complete list of periphrases can be found in the syntactic annotation guide for AnCora corpora.

- c) participles in non personal adjective clauses without explicit complements (*el material enviat va arribar just a temps* – the material sent arrived just in time)

There is a total of 105,112 elements with the <els> tag, 49,051 in the Catalan part of the corpora and 56,061 in the Spanish part.

## 5.2.2 Syntactic complements of verbs

As specified in the syntactic annotation guide for AnCora corpora, all constituents children of a non verbless clause node (<S>) or non verbless sentence node (<sentence>) bear function tags except for the verb phrase itself (labeled as <grup.verb>, <participi>, <infinitiu> or <gerundi>), interjections, nodes labeled as <INC> (inserted elements), nodes labeled as <morf.pron> (verbal pronominal morphemes), conjunctions (be it <coord> -coordinative, or <conj> -subordinative), and adjoined constituents. This is important because argument and thematic role tags are bounded to the function tag. No node can receive an argument tag if it does not have already a function tag, and no node can receive a thematic role tag if it does not have an argument tag already. That is to say, there is a hierarchy between these three features.

### Argument tag (<arg>)

Argument tags specify the type of argument each constituent fulfills, and its closeness to the predicate. All nodes with a function tag (<func>) must be annotated with an argument tag (<arg>) except in the following cases:

- passive markers (<func="pass">)
- impersonal markers (<func="impers">)
- non-argumental verb modifiers (<func="mod">)
- textual elements (<func="et">)
- orational adjuncts (<func="ao">)
- vocatives (<func="voc">)

As has already been hinted, there are seven possible values for the attribute <arg>: Arg0, Arg1, Arg2, Arg3, Arg4, ArgM and ArgL. The first five tags are numbered from less to more obliqueness with respect to the verb. ArgM corresponds to adjuncts, and ArgL codes lexicalized complements of light verbs. Arg0 corresponds to the external argument of the verb.

### Thematic role tag (<tem>)

Thematic role tags specify the thematic role each argument carries out. All nodes with an argument tag (<arg>) must be annotated with a thematic role tag (<tem>) except in the following cases:

- lexicalized complements of light verbs (<arg="ArgL">)
- argumental adjuncts with fuzzy semantic value (<func="cc"> and <arg="Arg1|Arg2|Arg3|Arg4">)
- prepositional objects with fuzzy semantic value (<func="creg">)

The possible values of the <tem> tag are the 19 different thematic roles we defined: adverbial (adv), agent (agt), attribute (atr), beneficiary (ben), cause (cau), destination (des), final state (efi), initial state (ein), experiencer (exp), extension (ext), purpose (fin), instrument (ins), location (loc), manner (mnr), origin (ori), patient (pat), source (src), theme (tem) and time (tmp). The information on which value to choose is specified in each lexicon file.

Finally, let us present an instance taken from the corpora, to illustrate the annotation format. Semantic tags dealt with in this guide are highlighted within a red dotted frame. The translation of the sentence is *Finally, (he) declared himself partial to not exaggerating the issues relating to this center.*

```

(12) <sentence>
  <sp func="et">
    <prep>
      <s gen="c" lem="per" num="c" pos="sps00" postype="preposition" wd="Per"/>
    </prep>
    <sadv>
      <grup.adv>
        <r lem="últim" pos="rg" wd="últim"/>
      </grup.adv>
    </sadv>
    <f lem="," pos="fc" punct="comma" wd=","/>
  </sp>
  <sn arg="arg1" elliptic="yes" func="subj" tem="tem">
    <morfema.pronominal>
      <p gen="c" lem="ell" num="c" person="3" pos="p0300000" wd="s"/>
    </morfema.pronominal>
    <grup.verb>
      <v gen="c" lem="haver" mood="indicative" num="s" person="3" pos="vaip3s0"
postype="auxiliary" tense="present" wd="ha"/>
      <v els="c2" gen="m" lem="mostrar" mood="pastparticiple" num="s" pos="vmp00sm"
postype="main" wd="mostrat"/>
    </grup.verb>
    <sa arg="arg2" func="cpred" predicate="subj" tem="atr">
      <grup.a>
        <a gen="m" lem="partidari" num="s" pos="aq0ms0" postype="qualificative"
wd="partidari"/>
      </grup.a>
    </sa>
    <sp>
      <prep>
        <s gen="c" lem="de" num="c" pos="sps00" postype="preposition" wd="de"/>
      </prep>
      <S clausetype="completive" impersonal="yes">
        <neg func="mod">
          <r lem="no" pos="rn" postype="negative" wd="no"/>
        </neg>
        <infinitiu>
          <v els="a2" gen="c" lem="sobredimensional" mood="infinitive" num="c"
pos="vmn0000" postype="main" wd="sobredimensional"/>
        </infinitiu>
        <sn arg="arg1" func="cd" tem="pat">
          <spec gen="m" num="p">
            <d gen="m" lem="el" num="p" pos="da0mp0" postype="article" wd="els"/>
          </spec>
          <grup.nom gen="m" num="p">
            <n gen="m" lem="assumpte" num="p" pos="ncmp000" postype="common"
sense="16:cs1" wd="assumptes"/>
          </grup.nom>
          <S clausetype="relative" impersonal="no">
            <relatiu arg="arg1" func="subj" tem="tem">
              <p gen="c" lem="que" num="c" pos="pr0cn000" postype="relative" wd="que"/>
            </relatiu>
            <grup.verb>
              <v els="c1" gen="c" lem="tenir_a_veure" mood="subjunctive" num="p"
person="3" pos="vmsp3p0" postype="main" tense="present" wd="tinguin_a_veure"/>
            </grup.verb>
            <sp arg="arg2" func="creg">
              <prep>
                <s gen="c" lem="amb" num="c" pos="sps00" postype="preposition" wd="amb"/>
              </prep>
              <sn>
                <spec gen="m" num="s">
                  <d gen="m" lem="aquest" num="s" pos="dd0ms0" postype="demonstrative"

```

```

wd="aquest"/>
      </spec>
      <grup.nom gen="m" num="s">
        <n gen="m" lem="centre" num="s" pos="ncms000" postype="common"
sense="16:02412359" wd="centre"/>
      </grup.nom>
    </sn>
  </sp>
</S>
</grup.nom>
</sn>
</S>
</sp>
</grup.a>
</sa>
<f lem="." pos="fp" punct="period" wd="."/>
</sentence>

```

## 6. References

- Aparicio, J., Taulé, M. & Martí M.A, (2008). AnCorra-Verb: A Lexical Resource for the Semantic Annotation of Corpora. *Proceedings of 6th International Conference on Language Resources and Evaluation*. Marrakesh (Morocco).
- Civit, M. & Martí, M.A. (2005). GramCat and GramEsp : two Grammars for Chunking. *Intelligent Information processing and Word Mining*, Gdansk, Poland. Springer Verlag.
- Dowty, D. (1991). Thematic proto-roles and argument selection. *Language*, 67.
- Kingsbury, P., Palmer, M. & Marcus, M. (2002). Adding semantic annotation to Penn TreeBank. *Proceedings of the 2002 Conference on Human Language Technology*, San Diego, CA.
- Kipper, K., Palmer, M. & Rambow, O. (2002). Extending PropBank with VerbNet Semantic Predicates. *Workshop on Applied Interlinguas*, held in conjunction with AMTA-2002. Tiburon, CA.
- Levin, B. & Rappaport-Hovav, M. (1995). *Unaccusativity. At the Syntax-Lexical Semantics Interface*. Cambridge, MA: MIT Press.
- Palmer, M., Kingsbury, P. & Gildea, D. (2005). The Proposition Bank: An Annotated Corpus of Semantic Roles. *Computational Linguistics*, 21 (1). USA: MIT Press.
- Rappaport Hovav, M. & Levin, B. (1998). Building Verb Meanings, in M. Butt and W. Geuder, eds., *The Projection of Arguments: Lexical and Compositional Factors*, CSLI (Soriano et al., 2008).
- Syntactic annotation guide to AnCorra corpora: <http://elie.ub.edu/ancora/>.
- Taulé, M., Aparicio, J., Castellví, J. & Martí, M.A. (2005). Mapping syntactic functions into semantic roles. *Proceedings of the Fourth Workshop on Treebanks and Linguistic Theories (TLT05)*. Barcelona: Universitat de Barcelona, 2005: 185-196.
- Taulé, M, Martí, M.A & Recasens, M. (2008). Ancora: Multilevel Annotated Corpora for Catalan and Spanish. *Proceedings of 6th International Conference on Language Resources and Evaluation*. Marrakesh (Morocco).
- Vázquez, G., Fernández, A. & Martí, M.A. (2000). *Clasificación verbal. Alternancias de diátesis*. Edicions de la Universitat de Lleida: Lleida.
- Vendler, Z. (1967). *Linguistics in Philosophy*. Cornell University Press: New York.