

# AnCora-Net: Mapping the Spanish Ancora-Verb lexicon to VerbNet

Mariona Taulé, M. Antònia Martí, Oriol Borrega

Centre de Llenguatge i Computació (CLiC)

University of Barcelona

Gran Via, 585, 08007 Barcelona.

{mtaule,amarti,oriol.borrega}@ub.edu

## Abstract

We present a new multilingual resource, the mapping of AnCora-Verb lexicon to VerbNet, with the aim of enriching both resources with lexical information. A two-way case study is conducted linking first, the VerbNet 13.1 class to the corresponding AnCora-Verb senses, and then a sample of AnCora-Verb senses, belonging to the broader a32 class, to VerbNet. The aim is to analyze the compatibility of both resources. The results show a neat correspondence. Taking into account that VerbNet is related to other semantic resources (PropBank, FrameNet, WordNet, and OntoNotes), AnCora-Verb can be enriched with this information. In the same way, Verb-Net can benefit from the more general classes used in the Spanish resource, and will be related to equivalent verbs in another language that can be useful for comparative studies.

## 1 Introduction

This paper describes the mapping of AnCora-Verb, a Spanish verbal lexicon, to VerbNet in the framework of the AnCora-Net project. The aim of this project is the integration of the EPEC-ADI Basque lexicon (Aduriz et al., 2006) and the Catalan and Spanish AnCora-Verb lexicons (Aparicio et al. 2008) into the English *Unified Verb Index* (UVI)<sup>1</sup>, giving rise to a multilingual —English, Spanish, Catalan and Basque— verbal lexicon. This mapping will enrich the contents of the different lexical resources involved by incorporating the semantic information included in each of them.

UVI is constituted by several English resources— VerbNet (Kipper-Schuler, 2006), PropBank (Palmer et al., 2005), FrameNet (Baker et al., 2003), OntoNotes Sense Groupings (Hovy et al., 2006), and WordNet (Fell-

baum 1998). The mapping between our lexicons and UVI will be done through PropBank and VerbNet at sense level. This choice is determined by the fact that these resources are the closest to AnCora-Verb. The Spanish lexicon follows the same annotation scheme as PropBank for argument structure, and is organized in semantic classes like VerbNet.

We hypothesize that these resources, though different in the way they semantically classify and represent verbal predicates, can still be compatible and complementary. Moreover, having these resources connected can provide a ‘more complete picture of the meaning aspects of a verb’ (Čulo et al., 2008).

Given that VerbNet provides the semantic classes for PropBank, we have conducted a case study to analyse the compatibility between AnCora-Verb and VerbNet semantic classifications. We deal with the complex task of comparing and linking two different semantic classifications within two different languages. In particular, we address the following aspects: a) how compatible is the fine grained classification proposed in VerbNet with the coarser grained one adopted in AnCora-Verb; b) whether the VerbNet and AnCora-Verb class mapping is compatible and consistent; and, finally c) how do arguments and thematic roles map.

The remainder of the paper is organized as follows. Section 2 describes the AnCora-Verb lexicons. Section 3 presents the case study carried out, and the analysis of results obtained are discussed in section 4. Finally, in section 5 conclusions, final remarks, and future work are presented.

## 2 AnCora-Verb lexicons

Ancora-Verb consists of two verbal lexicons, one for Spanish and one for Catalan, used as

<sup>1</sup> <http://verbs.colorado.edu/verb-index/index.php>

the basis for the semantic annotation of AnCora corpora with arguments and thematic roles (Taulé et al., 2008).<sup>2</sup> In the AnCora-Verb lexicons<sup>3</sup>, the mapping between syntactic functions, arguments and thematic roles of each verbal predicate is established taking into account the verbal semantic class and the diathesis alternations in which the predicate can participate. This mapping was manually encoded, and tests of inter-annotator agreement were carried out in order to ensure consistency in the description of predicates. Each verb is divided into different senses ('01, 02... n') and each sense is related to one or more syntactic-semantic frames (transitive, ditransitive, unaccusative, inergative, etc.). Each frame is characterized according to the four ontological event classes (accomplishment, achievement, states, and activities), and to the diathesis alternations in which a sense can occur.

We only considered very productive diatheses, such as causative/inchoative, active/passive, resultative, oblique subject, transitive/intransitive, object extension, cognate object, and beneficiary alternation (Vázquez et al., 2000).<sup>4</sup>

The semantic relation that each argument maintains with the event denoted by the verbal predicate is defined by thematic roles. We adopted a set of 20 different thematic roles, each of which can be mapped to several syntactic functions and argument positions.<sup>5</sup> The set of thematic roles is a subset of the 30 different thematic roles used in VerbNet. Following PropBank, the arguments required by the verb sense are incrementally numbered (arg0, arg1, arg2, etc.), expressing their degree of proximity in relation to their predicate. These

<sup>2</sup> AnCora is a multilingual corpus consisting of 500,000 words in Catalan (AnCora-CA) and Spanish (AnCora-ES) annotated with morphological (POS), syntactic (constituents and functions), semantic (argument structure, thematic roles, named entities, and nominal synsets) and coreference information (Recasens & Martí, 2010). The corpora are freely available at <http://clic.ub.edu/corpus/ancora>.

<sup>3</sup> <http://clic.ub.edu/corpus/ancora-lexics>

<sup>4</sup> The specific alternations shared by few verbs were not considered because they do not define general verb classes.

<sup>5</sup> The list of the different thematic labels is the following: "agt" (Agent), "cau" (Cause), "exp" (Experiencer), "scr" (Source), "pat" (Patient), "tem" (Theme), "cot" (Cotheme), "atr" (Attribute), "ben" (Beneficiary), "ext" (Extension), "ins" (Instrument), "loc" (Locative), "tmp" (Time), "mnr" (Manner), "ori" (Origin), "des" (Goal), "goal" (Purpose), "ein" (Initial State), "efi" (Final State) and "adv" (Adverbial).

criteria give rise to a coarse grained verbal classification with a total of 22 different classes compiled.

The Spanish AnCora-Verb lexicon consists of 2,821 lemmas corresponding to 3,934 different senses and 5,481 syntactic-semantic frames.

```
<lexentry lemma="prestar" lng="es" type="verb">
<sense id="1">
<frame default="yes" lss="c21" type="state-at-
tributive">
<argument argument="arg1" function="suj"
thematicrole="tem"/>
<argument argument="arg2" function="creg"
thematicrole="atr"/>
<constituent type="sp" preposition="a"/>
</argument>
<examples>
<example>Comparar la piel de los mamíferos y
anfibios se presta a una mayor polémica</example>
</examples>
</frame>
</sense <sense id="2">
<frame default="yes" lss="a32" type="ditransi-
tive">
<argument argument="arg0" function="suj"
thematicrole="agt"/>
<argument argument="arg1" function="cd"
thematicrole="pat"/>
<argument argument="arg2" function="ci"
thematicrole="ben"/>
<constituent type="sp" preposition="a"/>
</argument>
<examples>
<example>El Banco Interamericano
de Desarrollo prestará a Perú 120 millones de
dólares<sup>7</sup> como apoyo a un programa de mejora de la
calidad y cobertura de la educación secundaria en el
país andino</example>
</examples>
</frame>
<frame default="passive" lss="b12" type="unaccus-
ative-passive-ditransitive">
<argument argument="arg1" function="suj"
thematicrole="pat"/>
<argument argument="arg2" function="ci"
thematicrole="ben"/>
...</sense>...
</lexentry>
```

Figure 1: AnCora-Verb lexical entry of *prestar*

Figure 1 shows the information associated with the entry *prestar* ‘to lend’ in the AnCora-Verb-Es lexicon: the lemma (“prestar”); the category (“verb”); the different senses associated to their corresponding semantic classes—in this case, the first sense with the state-tribu-

<sup>6</sup> ‘To compare the skin of mammals and amphibians lends itself to a more controversial’.

<sup>7</sup> ‘The Inter-American Development Bank will lend 120 million U.S. dollars to Peru’.

tive semantic class (lss="c21"), and the second with the ditransitive class (lss="a32")<sup>8</sup>; the mapping between syntactic function, argument and thematic role (for instance, in the second sense the subject "suj" correspond to the first argument "arg0" with the thematic role of agent "agt"); and, the diatheses alternations in which the verb occurs (in the second sense, *prestar* can appear in passive "unaccusative-passive-ditransitive"). As we can observe, the expression of the passive alternation entails an argument crossing: the affected object, appears as direct object ("cd") in the active structure and as subject ("suj") in the passive structure, being in both cases the argument ("arg1") with the thematic role of patient ("pat"). Furthermore, the expression of this alternation also involves an aspectual change, since the active reading corresponds with an accomplishment (lss="a32") and the passive reading with an achievement (lss="b12"). Finally, examples are also included.

### 3 A case study

The case study consists in comparing and analyzing the compatibility of AnCora-Verb and VerbNet. It is conducted in two phases: 1) First, a VerbNet class is selected and all the members of this class are mapped to the corresponding verb senses in AnCora-Verb. 2) Secondly, the AnCora-Verb classes involved in the mapping process are obtained and then the 'inverse' mapping is conducted. That is, the remaining verb senses belonging to the selected AnCora-Verb classes are mapped to VerbNet. Notice that the mapping is established between PropBank senses, which are linked to VerbNet semantic classes, and AnCora-Verb senses.

We assume the following hypotheses: a) the VerbNet class will map to only one AnCora-Verb class; and b) the broader AnCora-Verb class will map into a restricted set of VerbNet classes. This set will share some characteristics being able to constitute a more general class in VerbNet.

<sup>8</sup> The first sense (*prestar* a) corresponds to a pronominal verb with the meaning of 'to lend itself; to accommodate or offer (itself) to: your words lend themselves to confusion', while the second sense refers to 'to permit the use (of something) with the expectation of return to the same or an equivalent: she lent me the book'.

### 3.1 Mapping VerbNet to AnCora-Verb

The linking of VerbNet<sup>9</sup> to AnCora-Verb is carried out automatically mapping the two resources through WordNet 3.0 and PropBank, and then manually validating the mapping obtained, or connecting those verbal senses without automatic correspondences. WordNet 3.0 is used as a bilingual lexicon and PropBank allows for the argument mapping. The mappings between VerbNet and WordNet 3.0 and PropBank are already established in UVI.

For this case study we selected the verbs belonging to VerbNet class 'give 13.1' (i.e., lend, loan, pass, peddle, refund, render) and to the subclass 'give 13.1-1' (i.e., give, hock, lease, pawn, rent, sell). They are defined as verbs of change of possession most of which display the dative alternation (Levin, 1993). This kind of alternation does not occur in Spanish. The order of constituents can vary (*El Banco prestará 120 millones a Perú* vs. *El Banco prestará a Perú 120 millones*)<sup>10</sup>, but Spanish does not allow for the double object construction (*\*El Banco prestará Perú 120 millones*) as in English. All verbs belonging to these classes map to the *a32* AnCora-Verb class (lss="a32" in figure 1), that is, the class including ditransitive agentive benefactive verbs. Figure 2, summarizes the syntactic and semantic properties of the *a32* class.

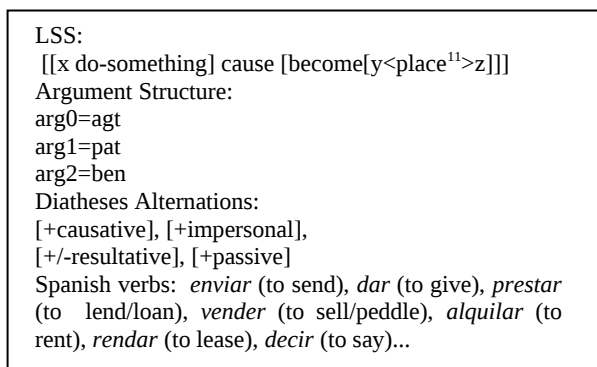


Figure 2: *a32* AnCora-Verb class

*a32* class associates a causer argument (x) with the semantic predicate *do*, and a third argument (z) with a location in space with semantic traits such as [+animate] or [+human].

<sup>9</sup> We use VerbNet 3.0.

<sup>10</sup> 'The Bank will lend 120 millions to Perú' vs. 'The Bank will lend (to) Perú 120 millions'

<sup>11</sup> 'place' has to be interpreted in a very general way, including the goal of physical as well as verbal transfer.

Since all the verbs belonging to this class allow for passive alternation, argument  $x$  is referred to as  $arg0$ -agent and argument  $y$  as  $arg1$ -patient. Argument  $z$  is referred to as  $arg2$ -beneficiary.  $arg0$  is syntactically the subject (*suj*), while  $arg1$  is the direct object (*cd*), and  $arg2$  the indirect object (*ci*) in the default frame. This class accounts approximately for 3% of the verbs represented in AnCora-Verb.

### 3.2 Mapping AnCora-Verb to VerbNet

In order to carry out the inverse mapping, the 30 most frequent  $a32$  AnCora-Verb class senses are selected (20% of the total) and mapped to VerbNet. In this case study, this mapping is manually done. In the AnCoraNet project it is carried out automatically. In figure 3 there is an example of the mapping between the involved resources. The tool used throughout all the process is AnCora-Pipe (Bertran et al., forthcoming).



Figure 3: a sample of the mapping

As Figure 3 shows, the mapping is established at two levels: sense and arguments. Sense mapping can be many-to-many (one AnCora-Verb sense may map to more than one VerbNet senses and vice-versa). In the bottom row of the sample, the link is illustrated: from *vender.1* ( $a32$ ) to *sell.01* ( $13.1-1$ ). As for argument mapping, the link is established taking into account argument position ( $arg0$ ,  $arg1$ ,...)

as well as, primarily, thematic role (*agent, patient*,...). That is, the anchor of the mapping is argument position, but what determines the final link is thematic role. As highlighted in the sample, a non-argumental complement ( $argM$ ) in AnCora-Verb maps to an argumental complement ( $arg3$ ) in VerbNet because they share the same thematic role (*adv~price paid*).

The members of AnCora-Verb  $a32$  class map to 20 different VerbNet classes. Half of them map to 'say 37.7' (6 links), 'reflexive\_appearance 48.1.2' (6 links) and 'order 60' (4 links) classes. The remaining ones map to 17 different classes (one or two links each) as shown in table 1.

Matches	VerbNet class	AnCora-Verb
6	37.7	Responder, confesar, declarar, afirmar, proponer, manifestar
6	48.1.2	Proponer, presentar, mostrar, manifestar, expresar, afirmar
4	60	Pedir, solicitar, reclamar, exigir
2	11.1	Entregar, enviar
2	13.1	Vender, dar
2	13.3	Ofrecer, conceder
2	78	Indicar, mostrar
1	10.5	Quitar
1	13.2	Devolver
1	22.2	Presentar
1	37.1.1	Explicar
1	37.1.2	Preguntar
1	37.10	Confesar
1	37.11	Comentar
1	37.13	Asegurar
1	37.9	Asegurar
1	58.2	Pedir
1	63	Imponer
1	64	Permitir
1	68	Pagar
0	No match	Atribuir

Table 1: AnCora-Verb to VerbNet mapping

The first column in Table 1 displays the number of matches. The second one, VerbNet class codes, which correspond to: steal (10.5), send (11.1), give (13.1), contribute (13.2), future\_having (13.3), amalgamate (22.2), transfer\_mesg (37.1.1), inquire (37.1.2), say (37.7), advise (37.9), confess (37.10), lecture (37.11), promise (37.13), reflexive\_appearance (48.1.2), beg (58.2), order (60), enforce (63), allow (64), pay (68) and indicate (78). Finally, the third column displays AnCora-Verb lemmas involved in the mapping. There are some

senses not represented in VerbNet, so the mapping of certain AnCora-Verb senses is not possible. As seen in the last row of Table 1, no satisfactory match was found for *atribuir* ('to attribute' is not represented in VerbNet, though it is in PropBank). In next section we analyze and discuss these results.

#### 4 Analysis of results

The results confirm that, in principle, verbs belonging to one VerbNet class ('give 13.1' and 'give 13.1-1') correspond to a unique AnCora-Verb class (*a32*). In this case we have a neat correspondence with no dispersion in the class members. The inverse mapping presents a completely different scenario: the 30 verbs of *a32* AnCora-Verb class map into 20 different VerbNet classes.

At first sight, this result could be interpreted as a lack of consistency between both resources, but in a deeper analysis some interesting generalisations arise. First, the VerbNet verbs mapped to *a32*, in spite of belonging to 20 different classes, are all ditransitive. This information is not explicitly declared in VerbNet, but it is captured thanks to the mapping with AnCoraVerb. The different criteria applied to the creation of these resources make them complementary: VerbNet is a semantic classification while AnCora-Verb is more syntax-oriented. The mapping between these resources makes information that is not explicitly declared evident, as for instance the ditransitivity shared by verbs coming from different classes in VerbNet. In a similar way, AnCora-Verb is enriched with more fine grained semantic information.

Second, all the verbs involved in the mapping process have three arguments: an *arg0* agent, an *arg1* theme and a third argument, that can be recipient, beneficiary, destination or patient (all of them with the feature [+animate]) depending on the meaning component represented in the corresponding VerbNet classes. For instance, in 'give 13.1' class the basic meaning component is 'transfer' and the third argument is a recipient, while in 'send 11.1' class the basic meaning components are motion and location, and the third argument is a destination. Despite the fact that these VerbNet classes grouped in the *a32* AnCora-Verb class have different semantic components, and that the verbs display different alternations, it is

also true that they share a general syntactic-semantic behaviour and, probably the common and more general meaning component of 'transfer' (Vázquez et al., 2000).

#### 5 Conclusions and remarks

The case study presented in this paper shows that the mapping between AnCora-Verb and VerbNet, two verbal semantic classifications within two different languages and based on different criteria, is a way to enrich and complement the contents of the involved resources, showing that the information included is compatible, consistent and complementary. VerbNet can benefit from the more general classes used in AnCora-Verb to create a coarser class classification capturing generalizations currently scattered throughout different VerbNet classes, for instance ditransitivity. AnCora-Verb classes can be grouped in a more fine grained subclasses considering VerbNet classification. Moreover, taking into account that VerbNet is also mapped to PropBank, FrameNet, OntoNotes, and WordNet in UVI, AnCora-Verb will be enriched with semantic information provided by these resources (VerbNet class, PropBank role-sets, FrameNet conceptual frames, WordNet synsets, OntoNotes Sense Groupings). Furthermore, all resources commented will also be enriched with the linking of Catalan AnCora-Verb lexicon, and the Basque EPEC-ADI lexicon, which are already being built. Finally, the resources in UVI will be related to equivalent verbs in three more languages, giving rise a multilingual verbal classification.

#### Acknowledgments

This research has been supported by the projects TEXT-MESS 2 (TIN2009-13391-C04-04) and AnCora-Net (FFI2009-06497-E/FILO) from the Spanish Ministry of Education and Science.

#### References

Aduriz I., Aranzabe M., Arriola J., Atutxa A., Díaz de Ilarraza A., Ezeiza N., Gojenola K., Oronoz M., Soroa A., Urizar R., 2006. Methodology and steps towards the construction of EPEC, a corpus of written Basque tagged at morphological and syntactic levels for the automatic processing. *Language and Computers*. Vol 56.

- Baker, C., Collin F., Charles J. Fillmore, B. Cronin  
2003. The Structure of the Framenet Database.  
*International Journal of Lexicography*, Volume  
16.3: 281-296.
- Bertran, M., M.A. Martí, M. Taulé, O. Borrega.  
2010. AnCoraPipe: a new tool for corpora anno-  
tation. Forthcoming.
- Čulo, O., K. Erk, S. Padó, S. Schulte im  
Walde. 2008. Comparing and combining se-  
mantic verb classifications. *Language Resources  
and Evaluation*, 42: 265-291, Springer.
- Fellbaum, C., ed. 1998. WordNet: An Eletronic  
Lexical Database. *Language, Speech and Com-  
munications*. MIT Press, Cambridge, Massachu-  
setts.
- Hovy, E., et al., 2006. OntoNotes: the 90% solu-  
tion. Proceedings of the Joint Human Language  
Tecnhology and North American Association of  
Computational Linguistics Conference. New  
York.
- Kipper, K., M. Palmer, O. Rambow. 2002. Extend-  
ing PropBank with VerbNet Semantic Predi-  
cates. *Workshop on Applied Interlinguas*, held in  
conjunction with AMTA-2002. Tiburon, CA.
- Levin, B. 1993. *English Verb Classes and Alterna-  
tions: A preliminary investigation*. University of  
Chicago Press, Chicago.
- Palmer, M., P. Kingsbury, D. Gildea. 2005. The  
Proposition Bank: An Annotated Corpus of Se-  
mantic Roles. *Computational Linguistics*, vol-  
ume 21, MIT Press, USA.
- Recasens, M., Martí, M.A. (2010) [AnCora-CO:  
Coreferentially annotated corpora for Spanish  
and Catalan](#). *Language Resources and Evalua-  
tion*. Springer Science.
- Vàzquez, G., A. Fernández, M.A. Martí . 2000.  
Clasificación verbal. Alternancias de diátesis.  
Ed. Universitat de Lleida , Spain.