Procesamiento del Lenguaje Natural, Revista nº 46 marzo de 2011, pp 11-18 recibido 17-12-10 revisado 15-02-11 aceptado 24-02-11

# AnCora-Nom: A Spanish Lexicon of Deverbal Nominalizations

## *AnCora-Nom: Un Léxico de Nominalizaciones Deverbales del Español*

**Aina Peris, Mariona Taulé**
CLiC-Centre de Llenguatge i Computació,
Universitat de Barcelona.
Gran Via de les Corts Catalanes, 585
08007, Barcelona
{aina.peris, mtaule}@ub.edu

**Resumen:** En este artículo se describe un nuevo recurso: AnCora-Nom, un léxico de nominalizaciones deverbales del español. Actualmente, contiene 1.655 entradas léxicas y 3.094 sentidos, donde cada sentido tiene asociado el tipo denotativo y la estructura argumental con los papeles temáticos correspondientes. Este léxico se ha extraído automáticamente a partir de la información anotada en el corpus AnCora-Es. AnCora-Nom se derivó teniendo en cuenta no sólo la información estrictamente relacionada con las nominalizaciones deverbales sino también con información morfológica y sintáctico-semántica previamente anotada en el corpus.
**Palabras clave:** nominalización deverbal, denotación, estructura argumental, recurso léxico, anotación de corpus.

**Abstract:** This paper describes a new lexical resource: Ancora-Nom, a Spanish lexicon of deverbal nominalizations. At present, it contains 1,655 lexical entries and 3,094 senses. Each sense has a denotation type associated, and the mapping of nominal complements with arguments and the corresponding theta roles is also annotated. A particular interest of this lexicon is that it has been automatically extracted from the annotated AnCora-Es corpus. AnCora-Nom was derived taking into account the information directly related to nominalizations, but also the morphological and syntactic-semantic information annotated in the corpus.
**Keywords:** deverbal nominalization, denotation, argument structure, lexical resource, corpus annotation.

## *1 Introduction*

This paper presents AnCora-Nom, a Spanish lexicon of deverbal nominalizations. At present, it contains 1,655 lexical entries corresponding to the different deverbal nominalizations appearing in the annotated corpus AnCora-Es (Taulé et al., 2008). In AnCora-Nom, each sense of a deverbal noun has a denotation type (i.e., event, result, and underspecified) associated to it, and the mapping of nominal complements with arguments and the corresponding theta roles are also annotated. The mapping between syntactic constituents and arguments is established taking into account the syntactic and semantic information of the verb from which the nominalization is derived. We use the AnCora-Verb lexicon (Aparicio et al., 2008) for obtaining the semantic verbal information related to the argument structure. Having a lexical resource that represents the semantics of nominalizations ─denotation type and argument structure─ and links these deverbal nouns with the corresponding verbs can be very helpful for information extraction tasks. It can also give more insight into the linguistic relation between verbs and their nominalizations, since it captures the similarity between sentences such as those in (1).

(1a) *La solución pasaba* ***por [abaratar el despido].***

ISSN 1135-5948

'The solution was to make dismissal cheaper'.

(1b) *La solución consistía en* ***[el abaratamiento del despido].***

'The solution consisted of the cost reduction of dismissal'.

Both sentences in (1) refer to the same event, expressed either by a verb (1a) or by a deverbal nominalization (1b). In fact, they are paraphrases.

The main goal of this paper is to present the content of the AnCora-Nom lexicon.

The remainder of the paper is organized as follows. Section 2 briefly discusses the related work. Section 3 focuses on the methodology developed to obtain the nominal lexicon. Section 4 describes the information coded in the nominal entries. Finally, main conclusions are drawn in Section 5.

## *2 Related work*

There exist similar lexical resources for other languages. NOMLEX[1] (Macleod et al., 1998) is a lexicon of English deverbal morphologically derived nominalizations consisting of 1,025 entries. This resource seeks not only to describe the allowed complements for a nominalization, but also to relate the nominal complements to the arguments of the corresponding verb. The NomBank project (Meyer, 2007), which carried out the semantic annotation of nominal argument structure in the PennTreeBank corpus (Palmer, 2005), used the same kind of representation for nominal complements as NOMLEX. In fact, NOMLEX-PLUS (Meyers et al., 2004), is an extension of NOMLEX, that automatically increases the number of lexical entries to 7,050. It also includes deverbal-cousin and deadjectival nominalizations, and argument-taking nouns such as relational nouns.

Another proposal to represent deverbal nominalizations is the Berkeley FrameNet Project (Ruppenhoffer et al., 2006). This project has created an on-line lexical resource[2] for English with 11,600 lexical units based on frame semantics and supported by corpus evidence. The aim is to document the range of semantic and syntactic combinatory possibilities -valences- of each predicative word, including verbs, nouns and adjectives, in each of its senses. The project produces semantic frames, annotating a set of examples for each predicate and describing the network of relations between the semantic frames. Concretely, deverbal nominalizations are annotated as events or entities, in a similar way to our event-result distinction (see Section 4.1). In the case of events, they are included in the semantic frame of the base verb while in the case of results they are considered to be a different semantic frame.

Although English is the language that has received the most attention, there are also resources devoted to nominalizations in other languages. In the line of the Berkeley FrameNet Project, there are proposals aiming at creating FrameNets for other languages such as German (Burchardt et al., 2009), Japanese (Ohara, 2009) and Spanish (Subirats, 2009). The Spanish FrameNet, for instance, contains 600 lexical units spread over 100 different semantic frames.

Besides the FrameNet framework, there are a number of proposals for languages other than English. An example is "The Essex Database of Russian Verbs and Their Nominalizations"[3] (Spencer and Zaretskaya, 1999). This database contains information about 800 Russian verbs and their corresponding nominalizations, distinguishing those verbal entries that nominalize the whole event and preserving the verbal argument structure from those that denote a concrete or abstract result of the verb. This database includes morphological and semantic information of this kind of nouns. Each nominal sense has one of the three semantic categories assigned to it (i.e., complex event, simple event and result) based on Grimshaw (1990) proposals, which are comparable to our denotation-type attribute (see Section 4.1).

There is also a French lexicon of nominalizations (Balvet et al., 2010) developed by the Nomage research group[4]. They aim at describing the aspectual properties of deverbal nouns. This group focuses on the development of two resources: a semantically annotated corpus of deverbal nouns, and an electronic lexicon. At present, the lexicon contains 815 lexical entries. For each one, they provide a definition, information about the source verb,

[1] http://nlp.cs.nyu.edu/nomlex/index.html

[2] http://framenet.icsi.berkeley.edu/

[3] http://privatewww.essex.ac.uk/~spena/res_interests.htm

[4] http://stl.recherche.univ-lille3.fr/programmesetcontrats/NOMAGE/NOMAGEenglish.html

the argument structure and the aspectual class of the nominalization.

Most of these lexicons, with the exception of NOMLEX-PLUS, were built manually. This is in contrast to the AnCora-Nom lexicon, which is automatically obtained from the data annotated in AnCora-Es.

## *3 Methodology*

As mentioned above, the methodology followed for building the AnCora-Nom lexicon takes advantage of the information annotated in AnCora-Es. It is a 500,000-word Spanish corpus consisting of newspaper texts annotated at different linguistic levels: morphology (PoS and lemmas), syntax (constituents and functions), semantics (verbal argument structure, thematic roles, semantic verb classes, named entities, and WordNet nominal senses), and pragmatics (coreference (Recasens and Martí, 2010))[5]. Recently, 23,000 deverbal nominalization tokens were also semantically annotated in the corpus (Peris et al., 2010). By *deverbal noun* we mean a noun morphologically derived from a verb or a so-called *cousin* noun (Meyers, 2007). Cousin nouns are nouns that give rise to verbs (e.g., *relación* 'relation' > *relacionar* 'to relate'), or nouns semantically related to verbs (e.g., *escarnio* 'mocking' is related to *mofarse* 'to make fun'). Specifically, we annotated the argument structure of these deverbal nominalizations, their semantic interpretation (i.e., result, event, and underspecified)[6], and whether or not they appear in a lexicalized construction. To this end, the first step was to run two independent automated processes: one for the annotation of denotation types and lexicalizations, using the ADN-classifier (Peris et al., 2010), and another one for the annotation of argument structures (Peris and Taule, forthcoming). In the latter case, we designed a set of heuristic rules in a decision-list format that take into account the argument structures specified in the AnCora-Verb lexicon. This is because we assume that deverbal nominalizations inherit the argument structure of the base verb. In fact, in the corpus each deverbal noun is linked to its corresponding verbal lexical entry sense via the attribute <originlexicalid>. This attribute allows for the extraction of nominalization senses in the lexicon (see Section 3.1). The second step was to check both types of information manually and to measure inter-annotator agreement to ensure the quality of the final corpus annotation (Peris et al., 2010) and thus, of the extracted lexicon presented here.

The lexicon was derived taking into account the information directly related to nominalizations, as well as the morphological and syntactic-semantic information annotated in the corpus.

### 3.1 Extraction process

The 23,000 tokens annotated correspond to 1,655 different lemmas. For each of these lemmas we created a lexical entry using the information annotated in the corpus. The extraction process consists of consulting all occurrences of each lemma and determining the different nominal senses a nominalization can have. In this process, we also extract the different features associated with each nominalization sense: argument structure, type of specifier, number, the verb from which it is derived, the verbal frame, etc.

To establish the senses of the nominalizations we take into account (1) the denotation type, (2) the sense of the base verb, and (3) whether or not the nominalization is part of a lexicalized construction. The criteria followed are:

- if a nominalization is part of a lexicalized construction, it constitutes a sense in itself; and
- those nominalizations with the same denotation and that are derived from the same verb sense are grouped under the same nominal sense.

Once the senses are established, the different features extracted from the AnCora corpus are associated to them (detailed in Section 4).

Each sense can also contain one or more nominal frames, depending on the verbal frame from which the nominalization is derived. There are a total of 3,204 different nominal frames. At the frame level, three important features are coded: the argument structure with the corresponding theta roles, the type of specifier, and a feature indicating whether or not the nominalization appears in plural (See Section 4). For the first two features, we also mark the frequency, i.e. the number of times that an argumental constituent or a type of

[5] http://clic.ub.edu/corpus/ancora

[6] For the linguistic criteria to distinguish between these three denotation types see Peris et al., (2010).

specifier occurs in the corpus. The plurality feature is not coded in the lexicon until all the nominalization occurrences in the same frame are explored; if one appears in plural, the value is positive. This extraction process has not produced errors in the generation of the lexicon. In the manual revision process, a 5% of errors was found, which corresponds to errors from the corpus annotation. In the next section we detail all the information contained in the lexical entry.

Once the 1,655 entries were created, we erased from the corpus the information related to the denotation type and whether or not the nominalization is lexicalized. We leave only a pointer to the corresponding nominal entry, where this information is declared. Furthermore, each nominal lexical entry is related to its corresponding verbal one, so AnCora-Nom and AnCora-Verb are two resources that are completely linked. Both lexicons are in XML format.

AnCora-Nom Lexicon

In this section we detail the structure and the information specified in the nominal entries. Figure 1 shows the full information associated with the lexical entry aceptación 'acceptance'.

We can observe that the lemma aceptación has two nominal senses. In this case, they are established because of the different denotation ─the first one denotes an event and, the second a result─ the verbal sense being the same for both (originlink="verb.aceptar.1").

Furthermore, each sense contains one or more nominal frames. By nominal frame, we mean the link to the corresponding verbal frame (which was established depending on diathesis alternations). It is important to preserve this link because the argument structure of the nominalization is inherited from the

```
<?xml version="1.0" encoding="UTF-8"?>
<lexentry lemma="aceptación" lng="es" origin="deverbal" type="noun">
<sense cousin="no" denotation="result" id="1" lexicalized="no" originlemma="aceptar" originlink="verb.aceptar.1" wordnetsynset="16:00117820+16:10039397">
<frame appearsinplural="no" type="default">
<argument argument="arg0" thematicrole="agt">
<constituent frequency="1" preposition="de" type="sp"/>
<constituent frequency="1" type="s.a"/>
</argument>
<specifiers>
<constituent frequency="1" postype="article" type="determiner"/>
<constituent frequency="1" type="void"/>
</specifiers>
<examples>
<example file="CESS-CAST-P/141_19981202.tbf.xml" nodepath="4.5.3.2.1.0" sentencenodepath="4">Para el realizador y guionista , el protagonista masculino , Stéphane , " es muy interesante porque Ø encarna la tolerancia , aceptación de los demás . </example>
…  </examples>
</frame>
</sense>
<sense cousin="no" denotation="event" id="2" lexicalized="no" originlemma="aceptar" originlink="verb.aceptar.1" wordnetsynset="16:00117820">
<frame appearsinplural="no" type="default">
<argument argument="arg1" thematicrole="pat">
<constituent frequency="2" preposition="de" type="sp"/>
<constituent frequency="1" postype="possessive" type="determiner"/>
</argument>
<specifiers>
<constituent frequency="2" postype="article" type="determiner"/>
</specifiers>
<examples>
<example file="CESS-CAST-A/11714_20000314.tbf.xml" nodepath="7.4.1.1.1.1.3.2.1.2.0" sentencenodepath="7"> El PP esperará los movimientos del consejero de Economía, desde la determinación de que cualquier apoyo dependerá de la " capacidad de diálogo y de llegar a acuerdos " y de la aceptación de nuestra capacidad de influencia </example>
</examples>
</frame>
</sense>
</lexentry>
```

Figure 1: *aceptación* lexical entry

corresponding verb, and this information is coded in the verbal frames. In Figure 1, each sense contains only a frame with the "default" type. This is because the occurrences of *aceptación* in the corpus link correspond to the default verbal frame, that is, the less marked frame of *aceptar* 'to accept' ('A21.transitive-agent-patient'). The different nominal senses and frames contain a series of attributes extracted from the morphosyntactic and semantic information specified in the AnCora-Es corpus. We present them in two different groups ─sense attributes and frame attributes─ in the following subsections. In addition to the features extracted from the corpus, the following features are also included:

a) **Lemma**. In Figure 1, the value for this attribute is the noun *aceptación* (lemma="aceptación").

b) The attribute **language** ("lng") codifies the language represented in the lexical entry. AnCora resources work with Spanish and Catalan, so the values of this attribute are "es" for Spanish (lng="es") and "ca" for Catalan (lng="ca"). At present, AnCora-Nom only deals with Spanish nominalizations, but in the future Catalan ones will also be taken into account.

c) The attribute **origin** indicates the type of word from which the nouns are derived. In Figure 1, the value for this attribute is "deverbal", meaning that this lexical entry concerns a noun derived from a verb. At present, AnCora-Nom only contains deverbal nouns but in the future it will include other type of nominalizations such as deadjectivals.

d) The attribute **type** refers to the word class, "noun" in Figure 1.

```
<?xml version="1.0" encoding="UTF-8"?>
<lexentry lemma="golpe" lng="es" origin="deverbal" type="noun">
  <sense alternativelemma="golpe_de_estado" cousin="no" denotation="result" id="4" lexicalizationtype="nominal"
lexicalized="yes" originlemma="golpear" originlink="verb.golpear.1" wordnetsynset="16:00629246">
    <frame appearsinplural="no" type="default">
      <argument argument="arg0" thematicrole="agt">
        <constituent frequency="1" type="s.a"/>
        <constituent frequency="1" preposition="de" type="sp"/>
      </argument>
      <argument argument="argL">
        <constituent frequency="6" preposition="de" type="sp"/>
      </argument>
      <argument argument="argM" thematicrole="fin">
        <constituent frequency="1" preposition="en_favor_de" type="sp"/>
      </argument>
      <specifiers>
        <constituent frequency="3" postype="indefinite" type="determiner"/>
        <constituent frequency="2" type="void"/>
        <constituent frequency="1" postype="demonstrative" type="determiner"/>
      </specifiers>
      <reference_modifiers>
        <constituent frequency="1" type="s.a"/>
      </reference_modifiers>
      <examples>
        <example file="3LB-CAST/111_C-6.tbf.xml" nodepath="1.2.1.1" sentencenodepath="1">El empresario fiyiano
George_Speight encabeza este golpe de Estado en_favor_de la comunidad de nativos fiyianos, como ha definido su
acción . </example>
      </examples>
    </frame>
  </sense>
</lexentry>
```

Figure 2: lexicalized sense *golpe de estado*

## 3.2 Sense attributes

The features related to sense attributes are the following:

a) The attribute **cousin** marks whether the nominalization is morphologically derived from a verb ("cousin=no", in Figure 1) or is a cousin noun ("cousin=yes").

b) The **denotation** attribute indicates the semantic interpretation of the deverbal noun. The possible values are: "event", if the nominalization expresses an action; "result", if

the nominalization denotes the result of an action, and "underspecified" if the contextual features of the nominalization do not permit the distinction between the two previous readings. In Figure 1, there are two senses, the first one being "result" (denotation="result") and the second one "event" (denotation="event").

c) Each sense contains an **identifier** ("id") to indicate the sense number. In Figure 1, the first sense is "id=1" and the second "id =2".

d) The **lexicalized** attribute indicates whether the nominalization is part of a lexicalized construction ─lexicalized="yes" (Figure 2) or not 'lexicalized="no" (Figure 1). In the former case, two additional attributes are added: (a) the **alternativelemma**, specifying the whole lexicalized construction of which the nominalization is part ("golpe_de_estado" in Figure 2)**,** and (b) **lexicalization_type**, to distinguish between six types of lexicalizations according to their similarity to different word classes: "nominal" (e.g., *golpe_de_estado* 'coup d'état'), "verbal" (e.g., *estar de acuerdo* 'to agree'), "adjectival" (e.g., *al alza* 'rising'), "adverbial" (e.g., *con cuidado* 'with care'), "prepositional" (e.g., *en busca de* 'in search of'), or "conjunctive" (e.g., *en la medida que* 'as far as'). In the case of nominal lexicalizations one of the three above-mentioned denotation values is assigned to the whole lexicalized construction. In Figure 2, the lexicalized construction "golpe_de_estado" is a nominal lexicalization (lexcalization_type= "nominal"), therefore, it has a denotation value ("denotation= result").

e) The attribute **originlemma** specifies the verb lemma from which the noun is derived. In Figure 1, the value for this attribute is "aceptar" in both senses (originlemma="aceptar").

f) Since verbs can have different senses, the attribute **originlink** points to the concrete verbal sense of the base verb. Recall that different verbal senses linked to a nominalization would imply different nominal senses. In Figure 1, however, the "originlink" attribute takes the same value in both senses "verb.aceptar.1" ("oringinlink=verb.aceptar.1") the one used for establish the relation between the nominal and the verbal entries.

g) Since nouns in the AnCora corpus are annotated with WordNet synsets[6], we incorporate this information into the attribute "**wordnetsynset**". In Figure 1, the first sense of *aceptación* corresponds to two synsets (wordnetsynset="1600117820+16:10039397"), while the second has only one (wordnetsynset="16:00117820").

## 3.3 Frame attributes

As mentioned above, in each sense we distinguish different nominal frames, as well as the arguments and thematic roles of the nominalization, the constituents that realize these arguments and how frequently they appear. Other attributes contained at the frame level are the specifier type of the nominalization, and if the deverbal noun nominalization appears in plural. The latter two attributes are important for establishing the nominalization denotation. Next, we detail the information at the frame level:

a) The attribute **type** indicates the verbal frame from which the nominalization is derived. In the AnCora-Verb, each verbal sense can be realized in more than one frame: default, passive, anticausative, locative, etc. In the nominal entries, we mark the corresponding verbal frames, which are the possible values for this attribute. In most cases, its value is "default" as in Figure 1 ("type=default").

b) **Argument (Structure)**. In this complex attribute, the different arguments ("argument") and the corresponding thematic roles ("thematicrole") are specified. To represent the arguments we follow the same annotation scheme used in AnCora-Verb (Aparicio et al., 2008). In Figure 1, the result sense has one argument ("arg0") with the *agent* thematic role ("agt"). This argument is realized once ("frequency=1") by a prepositional phrase ("constituent type =sp") introduced by *de* 'of' ("preposition=de"), and once by an adjective phrase ("type=s.a"). The event sense has one argument ("arg1") with the *patient* thematic role ("pat"). This argument is realized twice ("frequency=2") by a prepositional phrase ("constituent type =sp") introduced by the preposition *de* 'of' ("preposition=de") and once by a possessive determiner ("type=determiner", "postype=possessive").

c) The attribute **referencemodifier** represents the nominal complements that are not arguments but which modify the reference of the nominalization. Frequency is also taken into account. In Figure 2, the lexicalized construction *golpe de estado* has an adjectival

[6] We used WordNet 1.6 for Spanish.

phrase as reference modifier ("constituent type =s.a") appearing once ("frequency=1").

d) The type of specifier is a widely accepted criterion to distinguish between result and event readings, so we include this information in the attribute **specifier**. The possible values are: "article", "indefinite", "demonstrative", "exclamative", "numeral", "interrogative" "possessive", "ordinal" and "void", when there is no specifier. In this attribute, we also take into account the frequency the specifiers are realized. In Figure 1, the result sense is specified once by an article determiner ("type=determiner", "postype=article").

e) The attribute **appearsinplural** indicates if an occurrence of a nominalization in a particular frame appears in plural. It is a boolean attribute. In Figure 1, neither of the senses appear in plural, thus, the value is "no".

Finally, each lexical entry also contains all the examples from which the information has been extracted, specifying the corpus file, the node path and the sentence where it is.

## *4 Quantitative Analysis of Data*

AnCora-Nom consists of 1,655 deverbal lexical entries corresponding to 3,094 senses and 3,204 frames. Table 1 shows the distribution of nominal senses in denotation types and lexicalized (or non-lexicalized).

| ***Denotation*** | ***Lexicalized*** | ***Non-lexicalized*** | ***Total*** |
|---|---|---|---|
| Event | 0 | 631 | 631 |
| Result | 115 | 1,771 | 1,886 |
| Underspecified | 2 | 490 | 492 |
| None | 85 | 0 | 85 |
| Total | 202 | 2,892 | **3,094** |

Table 1: Distribution of nominal senses

The most frequent denotation among nominalizations is the result class (61%) followed by those nominalizations belonging to the event (20%) and the underspecified class (16%). It is not surprising that result nominalizations are the most frequent since events tend to be realized mostly by verbal clauses and nominalizations are more frequently used for the result-concrete meaning, more typical for nouns. The remaining 3% correspond to nominalizations without a denotation value because they are part of a lexicalized construction that is not nominal. In fact, these lexicalized constructions represent a 42% of the total number of lexicalized constructions. The remaining 68% correspond to nominal lexicalizations, which mostly belong to the result class.

Table 2 presents the distribution of senses taking into account the denotation type and the number of arguments. It shows that result senses are the most likely to appear without any argument (26%) in comparison to event (8%) and underspecified (4%) senses. In fact, most of the times event (53%) and underspecified (63%) senses appear with at least one argument in contrast to result senses (32%). When nominalizations have two arguments, event and underspecified senses are also more common (27% and 23% respectively) compared to the result senses (18%). However, result senses appear more frequently with more than two arguments (24%) than event (12%) and underspecified (10%) senses. This is explained by the fact that result nominalizations appear with more adjunct arguments (argMs), which can be associated with several theta roles (manner, time, location, etc.) in the same NP. However, event and underspecified nominalizations appear with real arguments (arg0, arg1…) that can only appear once in the same NP.

| ***Arguments*** | ***Event*** | ***Result*** | ***Underspecified*** |
|---|---|---|---|
| 0 | 48 | 499 | 22 |
| 1 | 336 | 603 | 312 |
| 2 | 168 | 340 | 111 |
| More than 2 | 79 | 444 | 47 |
| Total | 631 | 1,886 | 492 |

Table 2: Distribution of senses taking into account argument realization

## *5 Conclusions*

In this paper, we present a new lexical resource: AnCora-Nom, a Spanish lexicon containing 1,655 lexical entries of deverbal nominalizations. This lexicon was developed from the information encoded in the AnCora-Es corpus. It includes all the nominalizations found in the corpus with their possible denotations and argument structure combinations. AnCora-Nom is linked to the AnCora-Es corpus and to the AnCora-Verb Spanish lexicon, constituting an excellent resource for studying the argument realization of both nouns and verbs.

In the future, we intend to enlarge this lexicon with deadjectival nominalizations and relational nouns since we consider that they can also have an argument structure. We also intend to build a similar lexicon for Catalan nominalizations.

## *Akcnowledgments*

This research has received support from the projects Text-Knowledge 2.0 (TIN2009-13391-C04-04) and AnCora-Net (FFI2009-06497-E/FILO) from the Spanish Ministry of Science and Innovation, and a FPU grant (AP2007-01028) from the Spanish Ministry of Education. We also thank three anonymous reviewers for their comments and suggestions.

## *References*

Aparicio, J., Taulé, M. & Martí, M.A. (2008). 'AnCora-Verb: A Lexical Resource for the Semantic Annotation of Corpora'. *Proceedings of Language, Resources and Evaluation*. Marrakech, Morocco.

Balvet, A., Barque, L. & Marín, R. (2010). 'Building a lexicon of French deverbal nouns from a semantically annotated corpus'. *Proceedings of the 7th International Conference on Language Resources and Evaluation*, Valleta, Malta.

Burchardt, A. et al. (2009). 'FrameNet for the semantic analysis of German: Annotation, representation and automation'. In Hans C. Boas (ed.): *Multilingual FrameNets in Computational Lexicography: Methods and Applications*. Mouton de Guyter, 2009.

Grimshaw, J. (1990). *Argument Structure*. Cambridge, Massachussets: The MIT Press.

Macleod C., Grishman, R., Meyers, A., Barrett, L. & Reeves, R. (1998). 'NOMLEX: A Lexicon of Nominalizations'. *Proceedings of EURALEX'98*, Liege, Belgium.

Meyers, A. (2007). 'Anotation Guidelines for NomBank-Noun Argument Structure for PropBank'. Online Publication.

Meyers, A., Reeves, R., Macleod, C., Szekely, R., Zielinska V., Young, B., & Grishman, R. (2004). 'The Cross-Breeding of Dictionaries'. *Proceedings of the 4th International Conference on Language Resources and Evaluation*, Lisbon, Portugal.

Ohara, K.H, (2009). 'Frame-based contrastive lexical semantics in Japanese FrameNet: The case of risk and kakeru' In Boas, Hans, C. (ed.), *Mulitilingual FrameNets in Computational Lexicography: Methods and Applications*. Mouton de Gruyter.

Palmer, M., Kingsbury, P. and Gildea, D. (2005): The Proposition Bank: An Annotated Corpus of Semantic Roles, *Computational Linguistics*, 21 (1). USA: MIT Press.

Peris. A., Taulé, M., Boleda, G. & Rodríguez, H. (2010). 'AND-Classifier: Automatically assigning denotation types to nominalizations'. *Proceedings of the 7th International Conference on Language Resources and Evaluation*, Valleta, Malta.

Peris A., Taulé, M. & Rodríguez, H.(2010). 'Semantic Annotation of Deverbal Nominalizations in the Spanish AnCora corpus'. *Proceedings of The Ninth International Workshop on Treebanks and Linguistic Theories*, University of Tartu, Estonia.

Peris, A and Taulé, M. (forthcoming). 'Annotating the Argument Structure of Deverbal Nominalizations in Spanish'.

Recasens, M., Martí, M.A. (2010). 'AnCora-CO: Coreferentially annotated corpora for Spanish and Catalan'. *Language Resources and Evaluation,* Springer Science.

Ruppenhofer, J., Ellsworth, M., Petruck, M.L.R., Johnson, C.R. & Scheffczyk, J. (2006). 'FrameNet II: Extended Theory and Practice'. Online publication.

Spencer, A. and Zaretskaya, M. (1999), 'The Essex Database of Russian Verbs and their Nominalizations'. *Essex Research Reports in Linguistics*. University of Essex, Colchester, UK, 25.

Subirats, C. 2009 'Spanish FrameNet: A frame-semantic analysis of the Spanish lexicon'. In Hans Boas, ed. *Multilingual FrameNets in Computational Lexicography. Methods and Applications*. Berlin/New York: Mouton de Gruyter.

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).