Semantic Annotation of Deverbal Nominalizations in the Spanish AnCora Corpus

3 authors, including:

Mariona Taulé Delor
University of Barcelona
86 PUBLICATIONS **888** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

SOMEMBED-SLANG: Comprensión del Lenguaje en los medios de comunicación social. Representando contextos de forma continua-Lengua no estándar (TIN2015-71147-C2-2) View project

STERHEOTYPES: STudying European Racial Hoaxes and sterEOTYPES View project

Semantic Annotation of Deverbal Nominalizations in the Spanish AnCora corpus

Aina Peris
CLiC- UB
Gran Via,585
08007 Barcelona
aina.peris@ub.edu

Mariona Taulé
CLiC- UB
Gran Via,585
08007 Barcelona
mtaule@ub.edu

Horaci Rodríguez
TALP -UPC
Jordi Girona Salgado 1-3
08034 Barcelona
horacio@lsi.upc.edu

Abstract

This paper presents the methodology and the linguistic criteria followed to enrich the AnCora-Es corpus with the semantic annotation of deverbal nominalizations. The first step was to run two independent automated processes: one for the annotation of denotation types and another one for the annotation of argument structure. Secondly, we manually checked both types of information and measured inter-annotator agreement. The result is the Spanish AnCora-Es corpus enriched with the semantic annotation of deverbal nominalizations. As far as we know, this is the first Spanish corpus annotated with this type of information.

1. Introduction

In recent years semantically annotated corpora have been made available to the research community: PropBank (Palmer et al., 2005) for English, the TIGER corpus (Brants et al., 2002) for German, the PragueTreebank (Bohmova et al., 2001) for Czech, and the AnCora corpora (Taulé et al., 2008) for Catalan and Spanish. Most corpora focus on the argument structure of verbs, but some also annotate deverbal nominalizations because, like verbs, they also contain rich semantic information. However, they have until now been represented only in English corpora such as NomBank (Meyers, 2007)[1]. This paper presents the methodology and the linguistic criteria followed to annotate deverbal nominalizations in the Spanish AnCora-Es corpus[2]. The main goal achieved is the enrichment of AnCora-Es with the annotation of denotation types (i.e., result, event, and underspecified) and the argument structure of deverbal nominalizations. Identifying this information can be very useful for many tasks and applications of Natural Language Processing (NLP). In addition, such a corpus can provide real evidence for the linguistic study of nominalizations.

---

[1] Nominalizations are also represented in the different FrameNet projects, which are lexical databases supported by corpus evidence (Boas, 2009).

[2] AnCora-Es is a 500,000-word corpus annotated at different linguistic levels: from morphology to pragmatics (coreference): http://clic.ub.edu/corpus/ancora.

The remainder of the paper is organized as follows. Sections 2 and 3 describe the annotation of denotation types and of argument structures, respectively. Both sections include a description of the inter-annotator agreement tests. Finally, main conclusions are drawn in Section 4.

2.   Denotation Types

A relevant feature of deverbal nominalizations is their denotation, i.e., the semantic interpretation they are associated with. We focused on this information because it is one of the most controversial and studied topics in the literature, and it is not a straightforward distinction. In the linguistic literature (Grimshaw, 1990; Picallo, 1999; Alexiadou, 2001), nominalizations are said to have basically two semantic denotations: nominalizations denoting an event (1a)—i.e., they refer to an action,—or a result (1b)—i.e., they refer to the result of an action. However, in a previous study (Peris and Taulé, 2009), we observed that these two denotations do not allow us to account for the data in the corpus. First, it is not always possible to distinguish between event and result, since the linguistic context is sometimes not informative enough. We label such cases as underspecified types (1c), resulting finally in three possible denotation values. Second, we noticed that nominalizations can take part in a lexicalized construction, thus, we added the attribute <lexicalized>. One of the three above-mentioned denotation values is assigned to the whole lexicalized construction only in the case of nominal lexicalizations (1d)[3].

(1a) [La **reconstrucción**$_{<event>}$ de la ciudad por los chinos]$_{NP}$ tiene lugar en estos momentos.[4]
'[The **reconstruction** of the city by the Chinese] is being carried out at the moment.'
(1b) No espere [una **definición**$_{<result>}$ de la palabra cultura de María]$_{NP}$.
'Do not expect [a **definition** of the word culture from María].'
(1c) Se espera [la **llegada**$_{<underspecified>}$ de 450 observadores extranjeros]$_{NP}$.
'[The **arrival** of 450 foreign observers] is expected.'
(1d) Se habla de [un **golpe de Estado**] $_{<lexicalized="yes"><result>}$ de manera irresponsable.
'[A **coup d'état**] is being talked about in an irresponsible way.'

2.1 Methodology

The annotation of denotation types consists in associating a type to each deverbal noun. By *deverbal noun* we mean a noun morphologically derived

---

[3] We distinguish six types of lexicalization containing nominalizations according to their similarity to different word classes: nominal, prepositional, verbal, adjectival, adverbial and conjunctive lexicalizations.
[4] All the examples are extracted from the AnCora-Es corpus.

from a verb or a so-called *cousin* noun (Meyers, 2007)[5]. Cousin nouns are nouns that give rise to a verb (e.g., *revolución* 'revolution'> *revolucionar* 'to revolutionize'), or nouns semantically related to a verb (e.g., *genocidio* 'genocide' can be related to *exterminar* 'to exterminate'). Deverbal nouns were previously selected from a list automatically obtained using a predefined set of ten suffixes[6] that take verbal stems and have an action-result meaning (Santiago and Bustos, 1999).

The annotation of denotation types was carried out in two steps. First, the annotation was automated by means of the ADN classifier (Peris et al., 2010), which uses a machine-learning approach taking as features most of the information that appears in the current guidelines (see Section 2.3), obtaining 80, 6% F-measure. In a second step, the results were manually validated in order to ensure the accuracy of the annotation. Precisely, we annotated a total of 23,000 tokens belonging to 1,655 types of deverbal nominalizations.

2.2 Annotation Scheme

The attribute to represent the denotation value in the corpus is <denotationtype> and its possible values are: event, result and underspecified. This information is assigned to deverbal nouns together with the attribute <originlexicalid>, whose value is the base verb; thus, ensuring the connection with the corresponding verbal lexical entry of AnCora-Verb-Es (Aparicio et al., 2008). Lexicalized constructions are marked with the attribute <lexicalized> and the value 'yes'.

2.3 Linguistic Criteria for the Classification of Denotation Types

This section details the morphological, syntactic and semantic criteria for the classification of deverbal nominalizations into denotation types. Before applying these criteria, the annotators check whether the nominalization is part of a lexicalized construction. If so, they choose a lexicalization type. In (1d), *golpe de estado* 'coup d'état' is considered to be a lexicalized construction for three reasons. First, its reference changes in relation to the simple nominalization (*golpe* 'hit'); second, the second element (*estado* ' état') cannot take its own complements (e.g., *democrático* 'democratic') (1d'); and finally, the insertion of an element into the lexicalized construction is infelicitous (1d'').

---

[5] We understand that a deverbal nominalization is semantically related to a verb, regardless of whether it is morphologically derived from a verb (marked with the <cousin> attribute with a negative value) or not (marked with the <cousin> attribute with a positive value).

[6] The suffixes are: -a, -aje, -ión/-ción/-sión/-ón, -da/-do, -dura/-ura, -e, -ido, -miento/-mento, -ncia/-nza, -o/-eo.

(1d') *Se habla de [un **golpe de Estado** democrático] de manera irresponsable
'[A democratic **coup d'état**] is being talked about in an irresponsible way.'
(1d'')*Se habla de [un **golpe de** gran **Estado**] de manera irresponsable
'[A **coup de** large **état**] is being talked about in an irresponsible way.'

Once it has been decided whether the nominalization is part of a lexicalized construction, the denotation value is assigned according to the criteria that we present next. These criteria are not deterministic, they must be understood as indicators, the combination of which help us to decide the nominalization denotation.

a) Incorporated Argument: In Spanish the denotation type is result when the deverbal nominalization incorporates an internal argument from the corresponding base verb. For instance, *invento* 'invention' denotes the object resulting from the verbal action as well as the verbal action (2).

(2) [El **invento**Arg1$_{<denotationtype= "result">}$ de Juan] tuvo mucho éxito.
' [John's **invention**] had a lot of success.'

b) Plurality: An identifying criterion for result nominalizations proposed in the literature is their ability to take the plural inflection (3a), unlike event nominalizations (3b)[7].

(3a) Para compensar [las **pérdidas**$_{<denotationtype="result">}$ ante sus depredadores], los titíes traen al mundo gemelos.
'To compensate [the **losses** before their predators], monkeys bring twins to the world.'
(3b) […] aunque [la **pérdida**$_{<denotationtype="event">}$ del pívot Rodney_Dent] puede condenar a los de Rick_Pitino.
'[…] although [the **loss** of the pivot Rodney_Dent] can condemn those of Rick_Pitino.'

c) Determiners: It is widely accepted that event nominalizations can only be introduced by the definite article and the possessive, and they can also appear without a determiner (4a); whereas, result nominalizations can also be introduced by other types of determiners such as demonstratives, indefinite articles and numerals (4b).

(4a) No fue un hecho aislado, sino [*la* **culminación**$_{<event>}$ de [una dinámica de deterioro de las instituciones por_parte_del PP]].

---

[7] It is also stated that some events can be pluralized. We are aware of this possibility, but our annotation experience has revealed that most plural nominalizations denote results.

'It was not an isolated fact, but [*the* **culmination** of [a process of deterioration of the institutions on the part of the PP]].

(4b) Las exportaciones totales pasaron de 12,3 millones de dólares a 14,8 millones, lo que supone [*una* **subida**_{\<result\>} del 20,47_por_ciento].

'The total exports increased from 12.3 million dollars to 14.8 millions, which means [*an* **increase** of 20.47 per cent].'

d) Complementation: From the literature we learned that there are two types of nominalization complements that characterize a specific denotation type. First, relational adjectives are arguments of result nominalizations (5a), but not of event nominalizations (5b). Indeed, (5b) is ungrammatical because *producción* 'production' cannot be understood as an event: interpreting the relational adjective *quesera* 'cheese' as an Arg1 blocks this reading.

(5a) El tema de conversación era [la **actuación**_{\<result\>} *policial*_{AP-arg0-agt}].
'The topic of discussion was [the police **acting**].'
(5b) *[La **producción**_{\<event\>} quesera_{AP-arg1-pat} por los holandeses][8].
'[The cheese **production** by the Dutch].'

Second, temporal adjuncts of result nominalizations must be introduced by the preposition *de* 'of' (6a), whereas this preposition is not needed for temporal adjuncts of event nominalizations (6b).

(6a) Hoy, tras [una **negociación**_{\< result\>} *de trece horas* _{PP-argM-tmp}], se ha aprobado un nuevo texto sobre la reforma del seguro de desempleo.'
'Today, after [a **negotiation** of thirteen hours], a new text has been approved on the reform of the unemployment insurance.'
(6b) La compañía presentó una auditoría por primera vez desde [su **constitución** _{\< event\>} *en 1989* _{PP-argM-tmp}]
'The company submitted a clean audit for the first time since [its **constitution** in 1989].'

e) Verbal class: Following Alexiadou (2001) and Picallo (1999) we have taken into account the semantic class of the verb from which the nominalization derives. This is very useful in order to decide the denotation type. Nominalizations are annotated with the verbal classes declared in the verbal lexicon AnCora-Verb. There are four general classes that are defined according to Vendler's (1967) event classes: accomplishments, achievements, states and activities. Next, we briefly detail how they influence the annotation of denotation types.

---

[8] This example is not from the AnCora-Es corpus.

Accomplishments: Verbs belonging to this class can give rise to result, event and underspecified nominalizations. The reading of the nominalization depends on which verbal arguments are syntactically realized in the nominalized NP and by which constituents (Table 1).

Arg0	Arg1	Arg2	Denotation
Not realized	Not realized	Not realized	Result
Not realized	Not realized	PP/GRP	Underspecified
Not realized	PP/GRP/Poss[9]	PP/GRP/Poss	Event
PP-*por*-Agent	PP/GRP/Poss	Not realized or PP/GRP/Poss	Event
PP-*de* Agent	PP/GRP/Poss	Not realized or PP/GRP/Poss	Result
PP-*de/por* Agent	PP/GRP/Poss	Not realized	Underspecified
PP-*de/por* Agent	PP/GRP/Poss	PP/GRP/Poss	Event
PP-*de/por* Cause	PP/GRP/Poss	Not realized	Underspecified
PP-*de/por* Cause	PP/GRP/Poss	PP/GRP/Poss	Underspecified
Any constituent	patient possessive	Any constituent	Event

Table 1: Denotation types according to the argument realization for nominalizations derived from an accomplishment.

Achievements: Verbs belonging to the achievement class are realized in unaccusative structures (i.e., with no Arg0). Therefore, the denotation types of the corresponding nominalizations depend on the syntactic realization of Arg1 and Arg2 (see Table 2).

Arg1	Arg2	Denotation
Not realized	Not realized	Result
Realized	Not realized	Underspecified
Not realized	Realized	Underspecified
Realized	Realized	Event

Table 2: Denotation types according to the argument realization for nominalizations derived from an achievement.

States: Verbs included in this semantic class denote states and their corresponding nominalizations are always result nominalizations.

Activities: Predicates belonging to this class are unergative and their corresponding nominalizations can only have a result interpretation. Their subject can be explicit, but it is always introduced by the preposition *de* 'of'.

e) Selectors: When the above criteria do not lead to a clear denotation type, we found other indicators that can help select one, the so-called

---

[9] PP stands for prepositional phrase, GRP stands for genitive relative pronoun, and Poss stands for possessive determiner.

selectors. We distinguish two types of selectors: (i) external selectors, i.e., those elements that point to a specific denotation from outside the nominalized NP (7a); and (ii) internal selectors, i.e., prefixes within the nominalized NP that indicate a specific denotation type (7b).

(7a) *Durante* [la **presentación**$_{<event>}$ del libro $_{CN-arg1-pat}$], él abogó por la formación de los investigadores en innovación tecnológica.
'During [the **presentation** of the book], he advocated for the training of researchers in technological innovation.'
(7b) Hoy [la **reubicación**$_{<event>}$ del ex ministro $_{CN-arg1-tem}$] no resulta fácil.
 'Today, [the **relocation** of the ex minister] does not seem easy.'

The preposition *durante* 'during' in (7a) gives a clue to interpret *presentación* 'presentation' as an event. In (7b), the nominalization *reubicación* 'reubication' with the prefix *re-* having a reiterative meaning must be of event type. This is due to the fact that the repetitive meaning only applies to bases that denote actions.

2.4 Inter-Annotator Agreement

In order to ensure the quality and reliability of the manual validation of denotation types, an inter-annotator agreement test was carried out. Five Linguistic graduate students participated in the test. Since none of them had experience in distinguishing between denotation types and this is not an easy semantic distinction, we built a training data set consisting of one hundred sentences. Each sentence contained a deverbal nominalization. For the real test we used a data set of two hundred sentences (i.e., two hundred deverbal nominalizations). For the purpose of annotation, the five annotators took into account the criteria presented in the previous section. They were required to work individually. We measured agreement using observed agreement (Scott, 1955) and the Kappa coefficient (Siegel and Castellan, 1988). Table 3 shows the average results of the inter-annotator agreement test.

Average Pairwise Results	Training Set	Test Set
Observed agreement	68%	75%
Kappa	44%	60%

Table 3: Inter-annotator agreement results for denotation types.

As expected, there is an improvement between the training and test data sets that is even more noticeable in the kappa measure (16% improvement). With regard to the result for the test data set, it can be said that the agreement level is not bad (60% Kappa, 75% observed agreement) given that the

semantic distinction we are dealing with is very difficult[10]. Furthermore, this agreement score makes it possible to ensure the quality of the annotation since the annotators were allowed to seek advice from each other during the manual annotation process.

3. Argument Structure and Thematic Roles

The annotation of argument structures was based on the initial assumption that deverbal nominalizations inherit the argument structure of the base verb. Thus, the assignment of arguments and thematic roles was conducted taking into account the argument structures specified in the AnCora-Verb-Es lexicon. As we explained above, each noun is linked to the corresponding verbal lexical entry via the attribute <originlexicalid>. The annotation is limited to constituents inside the nominalized NP (8a) and also to arguments incorporated into the noun (8b)[11]. In the latter case, the noun was annotated with the attributes <arg> and <tem> and their values (see Section 3.2) during the manual validation process.

(8a) [El **impulso** *de la investigación*] es un punto clave para España.
'[The **promotion** *of research*] is a key point for Spain.'
 (8b) [La ***propuesta*** de María] es muy buena.
'[María's ***proposal***] is very good.'

   In Spanish, the constituents that can be arguments inside the NP are: PPs, relational adjective phrases (APs), GRPs, and possessive specifiers (spec-dp). Other types of constituents such as the rest of APs, NPs, adverbial phrases (ADVPs), or sentences (S) cannot be arguments inside the NP and receive the semantic label "RefMod" ("Reference Modifier"), which indicates that they modify the reference of the noun they are complementing.

3.1 Methodology

The annotation of argument structures consists in assigning the argument position (attribute <arg>) and the corresponding thematic role (attribute <tem>) to each argument in a nominalized NP. As in the case of the denotation type, an automated process was first applied, followed by a manual validation. A set of heuristic rules in a decision-list format was used for the automated annotation, obtaining 74% F-measure (Peris and Taulé, forthcoming).

---

[10] Standard guidelines describe a kappa over 75% as excellent, 40% to 75% as fair to good, and below 40% as poor (Fleiss, 1981). According to this, a 60% is a quite good agreement.
   [11] The *proposal* is the thing proposed, so the Arg1-patient is incorporated into the noun.

3.2 Annotation Scheme

We use the same annotation scheme as the one followed to annotate the argument structure of verbs in AnCora-Es, which was in turn based on PropBank (Palmer et al., 2005) and VerbNet (Kingsbury, et al., 2002). In this way, we ensure the consistency of the annotation of arguments between the different predicates (nouns and verbs). The selected arguments are incrementally numbered expressing their degree of proximity in relation to their predicate ('arg0', 'arg1', 'arg2', 'arg3', 'arg4') and the adjuncts are labeled as ArgM. The list of thematic roles includes 19 different labels widely used in linguistics[12]. In a nutshell, we assign the attribute <arg> and <tem> to each nominal argument. The combination of the six arguments tags and the thematic roles tags results in a final tagset of 36 semantic tags.

3.3 Criteria

The annotators look up the verbal entries of the AnCora-Verb lexicon to assign arguments and thematic roles to the constituents of nominalized NPs. If a constituent can be interpreted according to an argument declared in the verbal entry, then it is annotated with this argument position <arg> and thematic role <tem>. If not, then there are two possibilities: (a) the constituent is interpreted as an ArgM that is not represented in the verbal entry, or (b) it has no argumental interpretation but is a noun modifier, which is labelled as "RefMod".

3.4 Inter-Annotator Agreement

We also wanted to ensure the quality and reliability of the manual annotation of arguments and thematic roles. To this end, the inter-annotator agreement test was conducted on a data set of one hundred sentences, each sentence containing a deverbal nominalization with at least one candidate to be an argumental constituent. A total of 131 constituents were included. Three Linguistic graduate students with previous experience in annotating verbal argument structure participated in the test. Due to their experience no training was needed. The test consisted in deciding, for each constituent, (a) whether it was an argument, and, if so, (b) which argument position and thematic role it should be assigned (out of a total of 36 possible tags). To this end, they had to take into account the information specified in the AnCora-Verb lexicon about the verbal sense from which they decided the deverbal nominalization

---

[12] These labels are: 'agt' (agent), 'cau' (cause), 'exp' (experiencer), 'scr' (source), 'pat' (patient), 'tem' (theme), 'atr' (attribute), 'ben' (beneficiary), 'ext' (extension), 'ins' (instrument), 'loc' (locative), 'tmp' (time), 'mnr' (manner), 'ori' (origin), 'des' (goal), 'fin' (purpose), 'ein' (initial state), 'efi' (final state) and 'adv' (adverbial).

came. This was important since we measured inter-annotator agreement taking into account whether the annotators agreed on the verbal class from which the nominalization derived. We measured inter-annotator agreement using observed agreement (Scott, 1955) and the Kappa coefficient (Siegel and Castellan, 1988). Disagreements on the argument and thematic role are expected when the verbal class taken as starting point is not the same since the arguments and thematic roles to be mapped vary. For this reason, we calculate observed and kappa measures increasing the penalization rate for disagreements when the annotators picked the same verbal class, and reducing the penalization rate for disagreements when the annotators picked different verbal classes. The weighting schema for measuring agreement was empirically set to 40% for the former, and to 60% for the latter.

Total agreement = (0.4*same VerbalClass) + (0.6*different VerbalClass)

Table 4 presents the results of the inter-annotator agreement test. The columns show the results for each pair of annotators and the average result. The rows present observed agreement (OA) and kappa coefficients according to the above formula.

ANNOTATOR PAIRS	A and B	A and C	B and C	AVERAGE RESULT
Same verbal class	119	125	125	
OA	86%	96%	90%	90.6%
Kappa	84%	94%	88%	88.6%
Different verbal class	12	6	6	
OA	66%	66%	83%	71.6%
Kappa	60%	58%	80%	66%
Total	131	131	131	
OA	74%	78%	85.8 %	79.2%
Kappa	69.6%	72.4%	83.2 %	75%

Table 4: Inter-annotator agreement results for argument structure
and thematic roles.

We focus on the average result (last column). As expected, when the annotators did not agree on the verbal class, the agreement decreases approximately 20% both in observed (71.6%) and kappa (66%) agreement scores with respect to when the annotators agreed on the verbal class (90.6% and 88.6%, respectively). According to the above-presented measure, the mean of inter-annotator agreement reaches 75% kappa, which translates to 79.2% observed agreement. This is a satisfactory level of agreement given

that there are 36 possible tag combinations, which largely increases the opportunities for disagreement. Furthermore, this agreement score allows us to provide for a reliable manual annotation process.

4. Conclusions and Future work

In this paper, we presented the methodology followed to annotate deverbal nominalizations in the AnCora-Es corpus. The first step was to run two independent automated processes: one for the annotation of denotation types and another one for the annotation of argument structures. Secondly, and the focus of this paper, we manually checked both types of information and measured inter-annotator agreement. The result is the Spanish AnCora-Es corpus enriched with the semantic annotation of deverbal nominalizations. As far as we know, this is the first Spanish corpus annotated with this type of information. Future work will focus on applying the presented methodology to annotate the nominalizations in the Catalan AnCora-Ca corpus.

References

Alexiadou, A. (2001). *The Functional Structure in Nominals. Nominalization and Ergativity*. Amsterdam / Philadelphia: John Benjamins.

Aparicio, J. et al. (2008). 'AnCora-Verb: A Lexical Resource for the Semantic Annotation of Corpora'. In *Proceedings of Language, Resources and Evaluation*, LREC 2008. Marrakech, Morocco.

Boas, H (2009). *Multilingual FrameNets in Computational Lexicography. Methods and Applications*. Mouton de Gruyter.

Bohmova, A. et al. (2001). 'The Prague Dependency Treebank: Three-Level Annotation Scenario' In *Treebanks: Building and Using Syntactically Annotated Corpora, ed. Anne Abeille. Kluwer Academic Publishers*.

Brants, S et al., (2002). 'The TIGER Treebank'. In *Proceedings of the Workshop on Treebanks and Linguistic Theories*. Sozopol.

Fleiss, J.L. (1981). Statistical methods for rates and proportions. New York: John Wiley.

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

Kingsbury, P. et al. (2002). 'Adding semantic annotation to PennTreeBank'. In *Proceedings of the 2002 Conference on Human Language Technology*. San Diego, CA.

Meyers, A. (2007). 'Anotation Guidelines for NomBank-Noun Argument Structure for PropBank'. Online Publication: http://nlp.cs.nyu.edu/meyers/nombank/nombank-specs-2007.pdf

Palmer, M. et al. (2005). 'The Proposition Bank: An Annotated Corpus of Semantic Roles', *Computational Linguistics*, 21 (1). MIT Press.

Peris, A. and M.Taulé (2009). 'Evaluación de los criterios lingüísticos para la distinción evento y resultado en los sustantivos deverbales'. In *Proceedings of the 1st International Conference on Corpus Linguistics*. Murcia (Spain).

Peris et al. (2010). 'ADN-Classifier: Automatically assigning denotation types to nominalizations'. *In Proceedings of the 7th International Conference on Language Resources and Evaluation* (LREC-2010). Valleta. (Malta).

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

Picallo, C. (1999). 'La estructura del sintagma nominal: las nominalizaciones y otros sustantivos con complementos argumentales'. In Bosque & Demonte (Ed.) *Gramática descriptiva de la lengua española*. Madrid: Real Academia Española / Espasa Calpe.

Santiago, R. and E. Bustos (1999). 'La Derivación Nominal'. In Bosque & Demonte (Ed.) *Gramática descriptiva de la lengua española*. Madrid: Real Academia Española / Espasa Calpe.

Scott, W. A. (1955). 'Reliability of content analysis: The case of nominal scale coding'. *Public Opinion* Quarterly, 19(3): 321–325.

Siegel, S. and N. J. Castellan. (1988). *Nonparametric Statistics for the Behavioral Sciences*. McGraw-Hill, New York.

Taulé, M. et al. (2008). ' Ancora: Multilevel Annotated Corpora for Catalan and Spanish'. In *Proceedings of 6th International Conference on Language Resources and Evaluation*. Marrakesh (Morocco).

Vendler, Z. (1967). *Linguistics in Philosophy*. Cornell University Press, Ithaca, N.Y.