

Automatic Detection of Non-deverbal Event Nouns in Spanish

Gabriela Resnik
Universidad Nacional
de General Sarmiento
gresnik@ungs.edu.ar

Núria Bel
IULA
Universitat Pompeu Fabra
nuria.bel@upf.edu

Abstract

In this work we present the results of an experiment for the automatic identification of non-deverbal event nouns in Spanish such as *fiesta* or *terremoto*, which constitute a separate lexical class that can only be distinguished by its syntactic behaviour, because, and contrary to de-verbal event nouns, non-deverbal nouns cannot be identified by its morphological features. Experimental results confirm the existence of such lexical class and the possibility of automatically identifying its members by means of syntactic cue-based automatic lexical acquisition methods.

1 Introduction

In this paper we report our experiments to confirm the existence of a particular lexical class formed by non-deverbal event nouns in Spanish, such as *fiesta* ('party') or *terremoto* ('earthquake'), by means of automatic lexical acquisition methods. In contrast with nouns created through derivation from verbal bases –nominals like *elaboración* ('elaboration') or *entrenamiento* ('training')– non-deverbal event nouns show a special syntactic behaviour that seems to indicate the existence of a different lexical class. Our proposal analyses the lexico-syntactic properties of this proposed class of nouns and enables their detection through a set of syntactic contexts that reflect its lexical singularities. The consistency of the proposed class is further confirmed by the evidence that an automatic lexical acquisition method can identify the members of such a class correctly using the differentiating syntactic contexts that are taken as cues by the learner.

In what follows, firstly, we state some of the syntactic properties of non-deverbal event nouns which set them apart from both non-eventive nouns and deverbal event nouns (i.e. nominaliza-

tions or nominals) and establish them theoretically as a unique lexical class. Secondly, we describe how we map the differentiating syntactic contexts into cues that we considered relevant for the automatic detection of these nouns. Finally, we describe the experiment of automatic lexical acquisition based on those syntactic cues and show the results obtained in the classification of non-deverbal event nouns from the cues observed on a corpus of 21 million tokens from Spanish newspapers.

2 Some Properties of Non-Deverbal Event Nouns in Spanish

Our hypothesis is that Spanish event nouns like *accidente* ('accident') or *guerra* ('war') are members of a separate lexical class that can be distinguished by its syntactic behaviour both from other event nouns such as nominalizations as from non eventive nouns.

As regards nominalizations, Grimshaw (1990) states a relevant difference between process nominals and result nominals; this distinction can be applied to certain ambiguous nouns, such as *examination*, which has a process reading in a sentence like *The examination of the candidates took three hours* (= 'process of examining') and a non-eventive, result reading in a sentence like *The written examination consists of 100 multiple choice items* (= 'set of exercises'). These two kinds of nominals contrast in several lexico-syntactic properties: the obligatory/optional internal argument realization, manner of external argument realization, determiner selection and their possibility of controlling infinitival clauses, among others. Grimshaw's analysis of nouns such as *trip*, which she refers to as simple event nouns (as opposed to complex event nouns, i.e. process nominals), practically assimilates them to result nominals, because of a supposed set of shared syntactic properties.

This similarity between simple event nouns and result nominals is accepted, without much debate, in Picallo's (1991, 1999) analysis of Catalan and Spanish nominalizations, and in Alexiadou's (2001) work on nominalizations in Greek, English and other languages.

Pustejovsky (1995) proposes a richer representation of the various types of event nouns based in the differences found in some of the levels in which the GL model structures lexical information: argument structure, event structure and qualia structure. In this proposal, event structure is part of any lexical class, not just verbs and nominalizations, and therefore the eventive status of non-deverbal nouns such as Spanish *guerra* ('war') or *fiesta* ('party') must be encoded in that level of representation. The lexical entry for a noun like *guerra* would look like this (adapted from Pustejovsky 1995: 162):

$$\left[\begin{array}{l} \text{guerra} \\ \text{ARGSTR} = \left[\begin{array}{l} \text{D - ARG}_1 = x : \text{ind} \\ \text{D - ARG}_2 = y : \text{ind} \end{array} \right] \\ \text{EVENTSTR} = \left[\begin{array}{l} \text{E}_1 = e_1 : \text{process} \\ \text{HEAD} = e_1 \end{array} \right] \\ \text{QUALIA} = [\text{AGENTIVE} = \text{GUERRA_ACT}(e_1, x \oplus y)] \end{array} \right]$$

The eventive meaning of *guerra*, defined in the event structure as being of type 'process' and in qualia structure with an agentive value, accounts for the occurrence of this lexical item in contexts like those in 1, which show how certain 'aspectual' verbs select an event as its argument:

1. a. *La guerra empezó en 1903.*
The war started in 1903
'The war started in 1903.'
- b. *Terminó la guerra.*
Ended the war
'The war is over.'

Two entities of semantic type 'ind' jointly constitute the agent of this event, which is encoded by means of the symbol \oplus in the list of arguments of predicate GUERRA_ACT, the value of the agentive quale. These arguments are represented as 'default arguments' (D-ARG) in the level of argument structure, hence their syntactic realization is optional, as already noted in Grimshaw (1990), as well as varied, as can be seen in 2:

2. a. *Se conocieron durante la guerra.*
CL-REFL met during the war
'They met during the war'.
- b. *Se conocieron durante la guerra ruso-japonesa.*
'They met during the russo-japanese war'.
- c. *Se conocieron durante la guerra entre Rusia y Japón.*
CL-REFL met during the war between Russia and Japan
'They met during the war between Russia and Japan'.

In this proposal, then, the obligatory vs. optional argument realization is what distinguishes process nominals from result and non-deverbal nominals and it must be encoded in the lexical entry for the noun in terms of ARG vs. D-ARG.

Some evidences (Resnik 2004 and forthcoming) have been found that non-deverbal event nouns in Spanish cannot be assimilated to either result nominals or simple non-eventive nouns, in spite of their shared properties. In the next sections, we briefly present some syntactic contexts that support the view that non-deverbal event nouns are a separate lexical class and that the consistency of this separate class can be confirmed also by means of an experiment for identifying them automatically.

2.1. Argument Structure Realization

As regards argument structure, non-deverbal event nouns behave exactly as result nominals. Picallo (1999) shows that external argument realization distinguishes process and result nominals: process nouns project an agentive PP headed by *por* (*parte de*) (see 3a) while result nominals, in contrast, realize their external argument with a genitive DP (case-marked with the preposition *de*), as can be seen in example 3b. Non-deverbal nouns indeed behave like result nominals, as shown in 3c and d.

3. a. *la composición continua de obras para piano por (parte de) Satie*
the composition continual of works for piano by Satie
'the continual composing of piano works by Satie'
- b. *las sutiles composiciones para piano de Satie*
the subtle compositions for piano of Satie
'Satie's subtle piano compositions'
- c. **la fiesta por parte del equipo*
the party by the team
'the party by the team'
- d. *la fiesta del equipo*
the party of the team
'the team's party'

Besides, Spanish data confirm Grimshaw's (1990) claim that non ambiguous process nominals require internal argument realization (as in 4a and b), while such realization is optional with

result nominals and event non-deverbal nouns (4 c and d).

4. a. *la traducción de artículos por parte del estudiante*
the translation of articles by the student
'the translation of articles by the student'
- b. **la traducción por parte del estudiante*
the translation by the student
'the translation by the student'
- c. *La traducción (de Burton) (de Las mil y una noches)*
the translation of Burton of The Arabian Nights
'Burton's translation of *The Arabian Nights*'
- d. *La conferencia (de Juan) (sobre la globalización) duró demasiado.*
The lecture of John on globalization lasted too much
'John's lecture on globalization was too long.'

In fact, the argument structure of event non-deverbal nouns (and of result nouns) has the same properties as that of any argument taking non-deverbal noun, such as *foto* ('photo') or *mapa* ('map'), which do not denote events.

2.2. Event Structure Realization

Aspectual and Occurrence Predicates

One of the differences between result nouns and non-deverbal event nouns concerns its co-occurrence with aspectual and occurrence predicates. Aspectual verbs (*empezar*, *terminar* or *durar*, 'to begin', 'to finish' or 'to last') and occurrence verbs (such as *suced*, *ocurrir* or *tener lugar*, 'to happen', 'to occur' or 'to take place') select process nominals or non-deverbal event nouns as subjects (5 and 6 a and b), but do not accept result nominals, as can be seen in examples 5c and 6c (in these examples, the result reading is forced by the presence of the modifier *de piedra* or the demonstrative article, respectively). Thus, result nouns and non-deverbal event nouns cannot be under the same lexical class:

5. a. *La construcción/demolición del edificio empezó a las 10.*
The building/demolition of the building started at 10
'The construction/demolition of the building started at 10'.
- b. *La conferencia/película empezó a las 10.*
The lecture/film started at 10
'The lecture/film started at 10'.
- c. **La construcción de piedra empezó a las 10.*
The building of stone started at 10
'The stone building started at 10'.
- d. **La carretera empezó a las 10.*
The road started at 10
'The road started at 10'.
6. a. *La construcción de la Gran Muralla tuvo lugar en el siglo III a.C.*
The building of the Great Wall took place in the century III B.C.

- 'The building of the Great Wall took place in III B.C.'
- b. *La fiesta/clase/ceremonia tuvo lugar el 15 de abril.*
The party/class/ceremony took place the 15 of April
'The party/class/ceremony took place on April 15th'.
- c. **Esta famosa construcción tuvo lugar en el siglo III a.C.*
This famous building took place in the century III B.C.
'This famous building took place in III B.C.'
- d. **Esta ciudad tuvo lugar en el siglo III a.C.*
This city took place in the century III B.C.
'This city took place in III B.C.'

As can be observed from the examples 5d and 6d, result nouns behave exactly as any non-eventive simple nouns like *carretera* ('road') or *ciudad* ('ciudad').

Aspectual Quantifiers

Another difference between result nouns and non-deverbal event nouns is a different selection of aspectual quantifiers. Both process nominals and non-deverbal event nouns denoting an atelic event are compatible with an aspectual quantifier which bounds the duration of the event, as shown in 7a and b; in contrast, this type of quantifier is rejected by result nominals (7c) and any non-eventive simple noun (7d):

7. a. *dos años de construcción*
a year of building
'two years of building'
- b. *una hora de clase*
a hour of class
'an hour of class (time)'
- c. **dos años de construcción de piedra*
a year of building of stone
'two years of stone building'
- d. **dos años de zona/provincia.*
two years of area/province
'Two years of area/province'.

In 7c, the adjunct *de piedra* forces the result reading of *construcción*, which accounts for the ungrammatical effect. This means that in 5a, the seemingly ambiguous nominal must be interpreted as denoting a process¹.

Aspectual Prepositional Phrases

Besides, process nominals and non-deverbal event nouns may appear as complements of aspectual PPs, headed by complex prepositions such as *en (el) medio/(la) mitad de* ('in the middle/midst of'), *al final/principio de* ('at/in the end

¹ Nominals and non-deverbal event nouns denoting telic events, e.g. *llegada* ('arrival') or *accidente* ('accident'), reject these quantifiers, so this test applies only to a subset of the nouns under analysis.

of’) and *durante* (‘during’), which is shown in 8a and b. Result nominals, as well as simple non-eventive nouns, are not possible in these contexts (8c and 8d):

8. a. *Renunció en mitad de la filmación (de la película)/de la construcción (del puente)*
Resigned in middle of the shooting (of the film)/of the building (of the bridge)
‘S/he resigned in the middle of the shooting (of the film)/ building (of the bridge)’.
- b. *Renunció en mitad de la película/obra.*
Resigned in middle of the film/play
‘S/he resigned in the middle of the film/play’.
- c. **Renunció en mitad de la construcción de piedra.*
Resigned in middle of stone building
‘S/he resigned in the middle of the stone building’.
- d. **Renunció en mitad del autobús/de la prensa.*
Resigned in middle of-DET bus/of the press
‘S/he resigned in the middle of the bus/press’.

2.3. Count/mass Distinction

Determinant Selection

Process nominals are mass nouns and result nominals are count nouns, a fact that has well known effects in determiner selection which is restricted to the definite article in the case of process nominals (9a), whereas result nominals admit indefinites and demonstratives as well (9b):

9. a. *La/una/esta construcción azteca los dejó admirados.*
The/a/this building aztec CL-ACC left amazed
‘They were amazed at the/an/this aztec building’.
- b. *La/*una/*esta construcción del puente los dejó admirados.*
The/a/this building of the bridge CL-ACC left amazed
‘They were amazed at the/a/this building of the bridge’.

Non-deverbal event nouns, in contrast, are not a homogeneous class: some are countable (e.g. *conferencia*, *accidente*) and others are mass nouns (*silencio*, *pánico*), hence their behaviour towards determiners is not uniform². 10a shows that countable nouns admit the same variety of determiners as result nominals, whereas mass nouns behave like process nominals (10b):

10. a. *La/una/esta conferencia del profesor los dejó admirados.*
The/a/this lecture of the professor CL-ACC left amazed
‘They were amazed at the professor’s lecture’
- b. *La foto muestra el/*un pánico del público/??este pánico del público.* (OK if demonstrative is read as non-deictic)

² Cf. Resnik (2008) for a detailed analysis of the compatibility of eventive Spanish nouns with determiners.

The picture shows the/a panic of the audience/this panic of the audience

‘The picture shows the/a/this panic in the audience’

This behavior towards determiner does not support identification of non-deverbal event nouns with result nominals, as proposed in Grimshaw (1990: 59). In fact, in this respect, non-deverbal event nouns pattern along simple non-eventive nouns, which can also be primarily mass (*humo*, ‘smoke’) or count (*paquete*, ‘package’).

Pluralization

Being count nouns, result nominals can be pluralized, as opposed to process nominals, (cf. 11a vs. 11b). Non-deverbal event nouns are not a uniform class regarding countability, hence count nouns admit the plural morpheme (11c) and mass nouns do not (11d), just like it happens with any simple non-eventive noun:

11. a. *Las publicaciones del profesor son recientes.*
The publications of the professor are recent
‘The professor’s publications are recent’.
- b. **Las publicaciones de artículos son recientes.*
The publications of articles are recent
‘The publications of articles are recent’
- c. *Las conferencias del profesor son recientes.*
The lectures of the professor are recent
‘The professor’s lectures are recent’.
- d. **Los pánicos de los invitados duraron un instante.*
The panics of the guests lasted an instant
‘The guests’ panics lasted an instant’.

2.4. Properties of Non-deverbal Event Nouns in Spanish: a Summary

The analysis of the contexts provided in sections 2.1, 2.2 and 2.3 shows that non-deverbal event nouns occur in unique syntactic patterns: they are similar to result nouns (and simple non-eventive nouns) regarding the realization of argument structure, yet they pattern along process nominals regarding event structure, in view that they accept the same range of adjuncts and quantifiers as these nouns. As to other nominal properties, such as the mass/count distinction, the contexts show that non-deverbal event nouns are not quite like any of the two kinds of nominals, and they behave like simple non-eventive nouns. The table below summarizes the lexico-syntactic properties of the different kinds of nouns described in the previous sections:

	NDV E N (fiesta)	PR-N (construcción = event)	RES-N (construcción = resulting object)	NDN (mapa)
Obligatory	no	yes	no	no

internal argument				
External argument realization	genitive DP (<i>de</i>)	PP headed by <i>por</i> (<i>parte de</i>)	genitive DP (<i>de</i>)	genitive DP (<i>de</i>)
Subject of aspectual verbs (<i>empezar, durar, etc.</i>)	yes	yes	no	no
Aspectual quantifier (<i>dos años de, etc.</i>)	yes	yes	no	no
Complement of <i>durante, en la mitad de, etc.</i>	yes	yes	no	no
Count/mass distinction (determiners, plural forms)	mass/count	mass	count	mass/count

Table 1. Lexico-syntactic properties of Non-Deverbal Event Nouns (NDV E N), Process Nouns (PR-N) and Result Nouns (RES-N).

3. Automatic Detection of Non-deverbal Event Nouns

We have shown some of the singularities of non-deverbal event nouns as a lexical class in contrast with other eventive and non-eventive nouns. Besides, we have also obtained the confirmation of our hypothesis about the lexical class of non-deverbal event nouns from experiments with machine learning methods. In these experiments, we use the characterization of the contexts where we hypothesize that members of this class occurs to confirm that an automatic learner can usefully rely on them to automatically classify which ones of a set of non-deverbal nouns are indeed eventive and which ones are not.

Some work related to our experiments can be found in the literature dealing with the identification of new events for broadcast news and semantic annotation of texts which are possible applications of automatic event detection (Allan et al. 1998 and Saurí et al. 2005, respectively, for example). For these systems, however, it would be difficult to find non-deverbal event nouns because of the absence of morphological suffixes, and therefore they could benefit from our learner.

3.1. Cue-based Lexical Information Acquisition

According to the linguistic tradition, words that can be inserted in the same contexts can be said to belong to the same class. Thus, lexical classes are linguistic generalizations drawn from the

characteristics of the contexts where a number of words tend to appear. One of the approaches to lexical acquisition proposes to classify words taking as input characteristics of the contexts where they occur. The idea behind is that differences in the distribution of the contexts will separate words in different classes, e.g. the class of transitive verbs will show up in passive constructions, while the intransitive verbs will not. Thus, the whole set of occurrences (tokens) of a word are taken as cues for defining its class (the class of the type), either because the word is observed in a number of particular contexts or because it is not. Selected references for this approach are: Brent, 1993; Merlo and Stevenson, 2001; Baldwin and Bond, 2003; Baldwin, 2005; Joanis and Stevenson, 2003; Joanis et al. 2007.

Different supervised methods of Machine Learning (ML) techniques have been applied to cue-based lexical acquisition. A learner is supplied with classified examples of words represented by numerical information about matched and not matched cues. The final exercise is to confirm that the data characterized by the linguistically motivated cues support indeed the division into the proposed classes. This was the approach taken by Merlo and Stevenson (2001), who worked with a Decision Tree and selected linguistic cues to classify English verbs into three classes: *unaccusative*, *unergative* and *object-drop*. Animacy of the subject, for instance, is a significant cue for the class of object dropping verbs, in contrast with verbs in *unergative* and *unaccusative* classes. Baldwin and Bond (2003) used a number of linguistic cues (i.e. co-occurrence with particular determiners, number, etc.) to learn the countability of English nouns. Bel et al. (2007) proposed a number of cues for classifying nouns into different types according to a lexical typology. The need for more general cues has also been proposed, such as the part of speech tags of neighboring words (Baldwin, 2005), or general linguistic information as Joanis et al. (2007), who used the frequency of filled syntactic positions or slots, tense and voice features, etc., to describe the whole system of English verbal classes.

3.2. Cues for the Detection of Non-deverbal Event Nouns

As we have seen in section 2, non-deverbal event nouns can be identified by its occurrence in particular syntactic contexts and its co-occurrence

with particular lexical items as indicative cues of a class of non-deverbal event nouns in Spanish. We have used 11 cues for separating non-deverbal event nouns of non event nouns. These cues are the following:

Cues from 1-3. As we have seen in section 2.2, nouns occurring in PP's headed by prepositions such as *durante* ('during'), *hasta el final de* ('until the end of'), *desde el principio de* ('from the beginning of'), and similar expressions are considered to be eventive. Thus, the occurrence after one of such expressions will be indicative of event nouns.

Cues from 4-8. Nouns occurring as external or internal arguments of verbs such as *ocurrir* ('occur'), *producir* ('produce' or 'occur', in the case of ergative variant *producirse*), *celebrar* ('celebrate'), and others with similar meanings, are also events as we have justified in section 2.2. Note that we identify as 'external arguments' the nouns occurring immediately after the verb in particular constructions, as our *pos-* tagged text does not contain information about subjects (see below). In many cases it is the internal argument occurring in these contexts. These verbs tend to appear in 'presentative' constructions such as *Se produjo un accidente* ('An accident occurred'), with the pronoun *se* signaling the lack of external argument. Verbs like *ocurrir* appear in participial absolute constructions or with participial adjectives, which means they are unaccusatives.

Cues 9 and 10. The presence of quantifying expressions such as *dos semanas de* ('two weeks of') or similar would indicate the eventive character of a noun occurring with it, as shown in section 2.2.

Cue 11. As seen in 2.2., non-deverbal event nouns do have an external argument that can also be realized as an adjective. The alternation of DP arguments with adjectives was then a good cue for detecting non-deverbal events, even when some other nouns may appear in this context as well. For instance: *fiesta nacional* ('national party') vs. *mapa nacional* ('national map').

In the selection of these cues we have concentrated in those that separate the class of non-deverbal event nouns from the class formed by simple non-eventive nouns and result nouns. If it is the case that these are really cues for detecting non-deverbal event nouns, the learner should

confirm it by classifying non-deverbal event nouns correctly, separating them from other types of nouns.

4. Experiment and results

For our experiments we have used Regular Expressions to implement the patterns just mentioned, which look for the intended cues in a part-of-speech tagged corpus. We have used a corpus of 21M tokens from two Spanish newspapers (*El País* and *La Vanguardia*). The positive or negative results of the n-pattern checking in all the occurrences of a word are stored as binary values of an n-dimension vector. Thus, a single vector summarizes all the occurrences of a word (the type) by encoding how many times each cue has been observed or whether it has not been observed (zero value). One of the main characteristics of cue-based lexical classification is the impact of the sparse data problem: most of the possible contexts searched for as cues do not show up in the text. As a consequence, the vectors representing the words contain mostly zero values that can cause an overfitting, as the learner groups all zero values as one class. In order to avoid this effect we produced a vector that has been smoothed by replacing zero values with a mean value obtained from the test data following a standard smoothing technique (Mingers, 1989).

We used a Decision Tree (DT) classifier (Quinlan, 1993) in the Weka (Witten and Frank, 2005) implementation. DT's perform a general to specific search in a feature space, selecting the most informative attributes for a tree structure as the search proceeds. The goal is to select the minimal set of attributes that efficiently partitions the feature space into classes of observations and assemble them into a tree. In our experiments, we have taken as input all the occurrences (tokens) of a word that together cause the type to be classified as an eventive noun (class 1) or not (class 0). For the experiment we used a 10-fold cross-validation testing. The evaluation was done comparing the results of the DT classification with manually annotated gold-standard files made of 99 non-eventive and 100 eventive nouns, all of them non-deverbal (see Annex I). As we have seen in section 2, result and simple non-deverbal nouns have the same syntactic properties (with the exception of determiner selection, which does not seem to be a useful cue at this stage). In this first experiment, then, we decided to use only non-deverbal non-eventive

nouns such as *mapa*, as nouns like *construcción* would have given extremely ambiguous results, given their systematic process/result polysemy. Besides, the use of the nominalization morphemes as a cue will clearly separate them from non-deverbal nouns, making the experiment trivial.

The results of the experiment are the following:

Correctly Classified Instances 83.45 %
 Incorrectly Classified Instances 16.58 %
 Kappa³ statistic 0.6684

For each class we got the following precision (proportion of correct assigned classes) and recall (the proportion of gold-standard nouns for which our learner has assigned the same class) figures:

Class	Precision	Recall
0	0.845	0.82
1	0.824	0.848

Table 2: Results of the experiment in terms of Precision and Recall

There were 18 non-event nouns out of 99 that were wrongly considered to be event nouns, while 15 non-deverbal event nouns out of 100 were not identified as such. An analysis of the results confirms that low frequency (that is those that occur few times in the whole corpus) event nouns tend to be classified as non-eventive. Low frequency nouns cannot show the cues that confirm its event nature, just because they do not show up enough. As for the significance of the different cues, a Chisquared analysis of features indicates that the preposition *durante* (‘during’) is the best feature for identifying these non-deverbal event nouns, followed by other temporal prepositions such as *después de* (‘after of’) and similar.

5. Conclusions

In this paper we have suggested that the existence of a lexical class of exclusively non-deverbal event nouns in Spanish can be justified by differences in distributional behavior between these and the members of the class of result nouns, with which they are normally assimilated in the literature. In order to get further support for this view, we have carried out an experiment,

which consisted in training a DT with the information that is supposed to justify the existence of the class. The results of the classifier, beyond an 83%, show the validity of the approach.

References

- Alexiadou, A. (2001). *Functional Structure in Nominals: Nominalization and Ergativity*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Allan, J.; Papka, R.; Lavrenko, V. (1998). *On-line New Event Detection and Tracking*. SIGIR98, Melbourne, Australia.
- Bel, N.; Espeja, S.; Marimon, M. 2007. Automatic Acquisition of Grammatical Types for Nouns. In *HLT 2007: The Conference of the NAACL*. Companion Volume, Short Papers.
- Grimshaw, J. (1990). *Argument Structure*. Cambridge: The MIT Press.
- Joanis, E; Stevenson, S; and James, D. 2007. A General Feature Space for Automatic Verb Classification. *Natural Language Engineering*.
- Merlo P. and Stevenson S. 2001. Automatic Verb Classification based on Statistical Distribution of Argument Structure, *Computational Linguistics*, 27:3.
- Mingers, J. 1989. An empirical comparison of pruning methods for decision tree induction. *Machine Learning*, 4 (2), 227–243
- Saurí, R.; R. Knippen, M. Verhagen and J. Pustejovsky. 2005. Evita: A Robust Event Recognizer for QA Systems. *Proceedings of HLT/EMNLP 2005*: 700-70
- Picallo, M. C. (1991). “Nominals and Nominalizations in Catalan”. *Probus*, 3. 279-316.
- Picallo, M. C. (1999). “La estructura del sintagma nominal: Las nominalizaciones y otros sustantivos con complementos argumentales”, en Bosque, I. & V. Demonte (eds.) *Gramática descriptiva de la lengua española*. Madrid: Real Academia Española / Espasa Calpe. Vol. 1, Cap. 6, 363-394.
- Pustejovsky, J. (1995). *The Generative Lexicon*. Cambridge: MIT Press.
- Quinlan, R.J. 1993. C4.5: Programs for Machine Learning. Series in Machine Learning. Morgan Kaufman, San Mateo, CA.
- Resnik, G. (2004). *Los nombres eventivos no debverbales en español*. Proyecto de Tesis. Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra.
- Resnik, G. (2008) “La expresión del aspecto en el dominio nominal: determinantes y cuantificadores

³ Kappa is an index which compares the agreement against that which might be expected by chance. 0.6 is within what is considered to be significantly good.

de nombres eventivos”., *XI Congreso Nacional de Lingüística*. Sociedad Argentina de Lingüística/Universidad del Litoral. Santa Fe, 9-12 de abril de 2008.

Resnik, G. (forthcoming) “La determinación de la eventividad nominal en español”. En G. Ciapuscio (ed.) *De la palabra al texto: estudios lingüísticos del español*.

Witten, I. H. and Frank E. 2005. *Data Mining: Practical machine learning tools and techniques*. 2nd Edition, Morgan Kaufmann, San Francisco.

Annex I: eventive and non-eventive nouns

Eventive	Non-eventive
fiesta	mapa
feria	antología
festival	característica
boda	droga
funeral	plasma
velorio/velatorio	teléfono
ceremonia	montaña
evento	tubo
picnic	estética
cóctel	cliente
té	escena
banquete	colectividad
festín	canal
ágape	arquitectura
tertulia	cara
campaña	levedad
cónclave	estadio
cumbre	batuta
asamblea	súbdito
sesión	ciudad
misa	madera
vacaciones	cifra
receso	habitación
excursión	fotocopia
trayecto	vivienda
travesía	gas
clase	literatura
conferencia	especie
curso	paisaje
taller	diferencia
workshop	carretera
congreso	seguridad
simposio	red
jornadas	contraseña
tumulto	rodilla
coloquio	virus
entrevista	cantidad
audiencia	provincia
concierto	detalle
ópera	público
serenata	garganta
espectáculo	maqueta
show	dato

programa	volcán
película	cárcel
ciclo	familia
discurso	dinero
sermón	estereotipo
torneo	tarifa
campeonato	compañía
carrera	justicia
rally	humo
tormenta	balneario
tempestad	paquete
temporal	prensa
borrasca	vehículo
terremoto	dueño
sismo	prejuicio
huracán	banda
maremoto	consorcio
sequía	economía
catástrofe	figura
cataclismo	mar
desastre	pancarta
tragedia	grupo
holocausto	arma
drama	informe
incendio	diario
accidente	trama
impacto	zona
siniestro	misterio
caos	facultad
crisis	cadáver
guerra	nivel
batalla	pista
conflicto	columna
paz	combustible
silencio	estructura
ruido	ruta
escándalo	alimento
lío	herramienta
follón	factura
problema	miembro
motín	forma
huelga	tema
incidente	fuelle
boicot	temperatura
pánico	euro
miedo	ilusión
pasión	punto
furor	batería
rabia	silueta
siesta	unidad
frío	organismo
calor	norma
hambre	via
pereza	planta
dolor	autobús
fiebre	perspectiva
gripe	antena