Recognizing Deverbal Events in Context

Irene Russo, Tommaso Caselli, Francesco Rubino

Istituto di Linguistica Computazionale, CNR, Pisa
firstName.secondName@ilc.cnr.it

Abstract. Event detection is a key task in order to access information through content. This paper focuses on events realized by deverbal nouns in Italian. Deverbal nouns obtained through transpositional suffixes (such as -*zione*; -*mento*, -*tura* and -*aggio*) are commonly known as nouns of action, i.e. nouns which denote the process/action described by the corresponding verbs. However, this class of nouns is also known for a specific polysemous alternation: they may denote the result of the process/action of the corresponding verb. This paper describes a statistically based analysis that helps to develop a classifier for automatic identification of deverbal nouns denoting events in context by exploiting rules obtained from syntagmatic and collocational cues identified by linguists.

1 Introduction

In Italian, deverbal nouns obtained through transpositional suffixes (such as -*zione*; -*mento*, -*tura*; and -*aggio*) are commonly known as nouns of action (*nomina actionis*) or nominalizations, i.e. nouns which denote the process/action described by the corresponding verbs. This class of nouns is also known for a specific lexical ambiguity phenomenon: they may denote the result of the process/action of the corresponding verbs. Commonly, these two different denotations of deverbal nouns are named *event* (example 1) and *result* (example 2) reading:

(1) La **costruzione**$_{EVENT}$ del ponte é durata tre anni.
The building of the bridge lasted three years.
(2) Questa *costruzione*$_{RESULT}$ é imponente. *This building is huge.*

This paper focuses on a statistically based analysis for the disambiguation of Italian deverbal nouns in context, using syntagmatic and collocational cues that are specific for the identification of the eventive reading. The classifier has been built using J48, the rule version of the decision tree classifier C4.5, and distributed through the Weka platform [1]. Two different training sets have been used: the It-TimeBank corpus[1], a corpus of Italian newpaper articles annotated with the Italian version of the TimeML specifications [2] and the La Repubblica Corpus [3]. From each set of data we have extracted co-occurence frequencies with a list of relevant syntagmatic cues (namely verbs and adjectives) identified through a detailed review of linguistically oriented works such as [4], [5], [6]. Next to this set of linguistically informed cues, we have also experimented the use of parts-of-speech (POS) sequences, which from previous works in word sense disambiguation tasks have proved useful ([7] among others).

In addition to the development of a classifier for disambiguating the eventive reading of deverbal nouns in Italian, we also want to verify the usefulness of the linguistically informed features, i.e. how powerful they are in discriminating the correct reading, by exploiting different types of training data, namely manually annotated tokens (single sentence level) *vs.* distributional frequencies of pre-classified types (global corpus level). We test if a combination of relevant lexical cues useful for broad semantic classification out of context and syntactic patterns essential for the discrimination in context can help for the disambiguation of deverbal nouns.

The remaining of the paper is structured as follows: section 2 describes the set of linguistically motivated features which emerges from the review of previous works. Section 3 is devoted to the description of the classifier by means of the the experiments conducted and their evaluation. In section 4 the methodology we have adopted is compared with previsous works in NLP on this subject. A tentative comparison of the results is outlined though the data sets used for the evaluation are different. Finally, section 5 reports on the conclusions and future developments.

[1] The corpus is still under development and not officially distributed.

2 Syntagmatic and lexical cues for deverbal nouns disambiguation

To automatically detect nouns that denote an event, morphological suffixation provides an important cue. However, deverbal nouns exhibit a peculiar and complex kind of logical polysemy [8]. The deverbal nouns can denote the act but also the result of an action, both as an abstract result (as in "*L'**espressione** di Maria fu inopportuna*" [Maria's statement was out of context]) or as a concrete result (as in "*L'**espresione** scritta alla lavagna era scorretta*" [The formula on the blackboard was wrong]). In these cases, the new meaning is the object of the verb though in other cases the non-eventive meaning can denote a result state ("*la **coagulazione** del sangue*" [blood clotting]), an instrument ("*L'**illuminazione** della sala fu rimessa in funzione*" [The illumination of the room was brought back into operation]), a material ("*la **segatura***" [the sawdust]), a person or object responsible for the action ("*la **difesa** accusò i giudici*" [the defense accused the judges]), the place where the predicate is realized ("*la sua **sistemazione** era un lussuoso appartamento*" [his accommodation was a luxury apartment]), the modality ("*la **classificazione** dei libri è pessima*" [the book classification is wrong]).

The theorical literature on this subject such as [5], [9], [6], point out the selection of specific cues for the identification of the two possible readings. For clarity's sake we report in Table 1 this set of cues.

[10] verified through a corpus-based quantitative analysis on a set of 842 Spanish deverbal nouns (over a total of 3,075 occurrences and 1,121 senses) the relevance of these cues. We claim that their results can be applied to Italian provided the high similarity of the two languages.

Among the scholars there is not a complete agreement on these features. Moreover, a small set of cues is suggested but no effort is made to establish the nature of their discriminative role (i.e are they dichotomous?), neither to rank the cues on the basis of their discriminative strength. As a consequence, linguistic theories lack of classification rules that instead are strictly necessary for computational systems. The identification of the most relevant cues and corresponding values must be carefully conducted since we aim at

Table 1. Cues for the identification of the eventive *vs.* non-eventive reading of deverbal nouns.

Features/cues	**Event reading**	**Non-eventive reading**
Obligatory realization of verb argument structure by means of a PPs	+	-
Pluralization	+	-
Telicity of the verb	+	-
Verb grammatical class	+	+
Type of determiner	+	-
Aspectual modifiers	+	-
Agent-oriented modifiers	+	-
Co-occurrence with eventive predicate	+	-
Complement clause at the infinitive	+	-
by-phrases, relational adjectives and possessive determiners as realizations of the subject of the deverbal noun	+	-

automatically detect them in text.

In the remaining of this section, we will go through some of the features listed in 1. Descriptive statistics on the features are reported to briefly asses their import. The figures have been obtained through a test set of 581 deverbal nouns extracted from the It-TimeBank (see section 3 for details on the resource) Corpus. The test set contains 440 occurrences of eventive nouns and 141 non-eventive nouns.

Obligatory realization of the argument structure with a PP . One of the most controversial point is related to the role of argument structure. [5] claims that only complex event nouns[2] have argument structure and its realization is compulsory. On the other hand, other scholars ([9], [6] among others) consider the presence of argument structure as an ancillary element for the disambiguation of deverbal nouns. These authors go even further in claiming that all event nouns, both complex and simple, can have arguments and its overt (i.e. superficial) realization is not necessary in order to instantiate the event reading. For instance, the noun "*fucilazione*" [shooting] has event readings both in example 3 and in 4, though in 4 there is no overt (superficial) argument realization:

[2] In her account, Grimshaw distinguishes among three types of nominalizations, namely (i.) complex event nominals, which requires the obligatory realization of the verb argument structure, (ii.) simple event nominals, which have event reading but do not realize argument structure and (iii.) result nouns.

(3) La **fucilazione**$_{EVENT}$ della prigioniera$_{Arg1}$ da parte dei soldati$_{Arg0}$.
The shooting of the prisoner by the soldiers.

(4) La **fucilazione**$_{EVENT}$ ha avuto luogo nella piazza.
The shooting took place in the square.

The results from [10]'s analysis have provided a partial support to this latter hypothesis. They have observed that almost every eventive reading of deverbal nouns (98%) presents a realization of the argument structure. However, they have also observed that there are cases in which the argument structure is not realized and argument structure can be realized by constituents other than PPs, such as possessive determiner. As for Italian [8] argues that predicate arguments can be omitted but they are frequently expressed through the preposition "*di*" and that possessive adjectives can express arguments as well.

On the basis of these results, the presence of the argument structure could be a discriminating cue but the automatic detection of internal arguments of deverbal nouns is not an easy task due to the fact that their identification is subordinated to the identification of the status of the deverbal noun (eventive *vs.* non-eventive). In Table 2 we report the percentages of nouns co-occurring with the realization of the argument structure in the dataset. If the argument structure is preferentially realized through the PPs "*di/del*", it is apparent that eventive nouns are more often followed by this kind of phrases with respect to non eventive nouns. From the data, it seems that possessive modifiers tend to co-occur with non-eventive readings against linguists' intuitions.

Table 2. Co-occurrence percentages of the cues for argument structure realization.

Noun type	**Possessive modifiers**	**PPs Di**	**/ Del**
eventive deverbal nouns	0.8%	47%	40%
non eventive deverbal nouns	2.5%	28%	22%

Pluralization The occurrence in plural forms of a deverbal noun is considered as a discriminating cue for detecting its non-eventive reading. As a matter of fact, [10] reports that 98% of the plural instances of deverbal nouns have a non eventive reading. On our dataset (see Table 3): deverbal eventive nouns are less frequently pluralized with respect to non-eventive nouns, even if the difference is not striking.

Table 3. Percentages of singular and plural occurrences.

Noun type	Singular	Plural
eventive deverbal nouns	87%	13%
non eventive deverbal nouns	59%	41%

Type of determiner According to the theoretical literature, if the determiner of a deverbal noun is a definite article, the noun will have an eventive reading. As reported in [10], this hypothesis is not verified by a corpus analysis. Demonstratives tend to prefer resultative (i.e. non-eventive) readings. Even if these features are reported in literature, it is hard to define how they can be used to disambiguate the correct reading of the deverbal nouns because the differences in percentages between eventive and non-eventive readings are not significant. However, they are retained in our analysis because linguists' intuitions report on their role.

Table 4. Co-occurrence percentages of determiners.

Noun type	il/la	un/una	demonstrative
eventive deverbal nouns	39%	13%	1%
non eventive deverbal nouns	33%	10	3.8%

Aspectual modifiers [10] did not report any figures on collocational cues in their study. However, it is possible to identify a rich list of relevant lexical items which could help in the classification of eventive nouns out of context and their identification in context. We manually selected a set of 53 high frequency adjectives and 41 verbs that can be reputed good collocational cues for the identification of eventive readings. In particular, we focus our attention on a selection of aspectual concurrent adjectives (e.g. "*annuo*" [yearly], "*contemporaneo*" [contemporary], "*immediato*" [immediate]) that modify more frequently eventive nouns. Other potentially interesting lexical cues are agent-oriented adjectives (e.g. "*abile*" [able], "*moderato*" [moderate], "*volontario*" [voluntarily])) that tend to co-occur with eventive nouns. Finally, we consider the co-occurrences of the nouns either as the object or as the subject of eventive predicates, such as "*continuare*" [to continue], "*finire*" [to finish], "*rimandare*" [to postpone] and so on and so forth.

3 Towards the classifier: experiments and results

In the development of the classifier we want to compare, on one hand, syntagmatic and collocational information from manually annotated corpora with co-occurrence frequencies from large corpora extracted after a coarse grained annotation derived from a lexical resource. On the other hand, we want to test the relevance of the cues suggested by linguists with similar cues extracted without previous assumptions.

The data set we have used to train the Weka version of the C4.5 algorithm is composed by three different sets of data: two training datasets, the It-TimeBank corpus and the La Repubblica Corpus [3], and one test set, composed by a TimeML-compliant manually annotated data from the La Repubblica Corpus.

The It-TimeBank is an Italian corpus composed by 149 newspaper articles, for a total of more than 63 thousand tokens, with 18,312 of them being labelled as nouns. Six annotators have manually applied the TimeML specifications [2] by distinguishing between temporal expressions, events and signals. As far as the event annotation is concerned the corpus contains 8,138 tokens annotated as events (including verbs, nouns, adjectives and prepositional phrases), 3,695 of whom are realized by nominal tokens. As already stated, we have a grand total of 581 deverbal tokens realized by means of transpositional suffixes, which count 440 event tokens and 141 non-eventive ones. Inter-annotator agreement on event annotation is K = 0.87 and average precision and recall 0.89, which guarantee a reliable supervised data set. A subset of 31,000 tokens of this corpus has been released for the SemEval 2010 TempEval-2 task [11].

The La Repubblica set is a training dataset composed by 1054 high frequency nouns and subdivided in two sub-sets: 566 deverbal nouns exclusively eventive such as "*pulitura*" [cleaning], "*proliferazione*" [proliferation] selected according to the transpositional suffixes in analysis, and 488 non eventive nouns such as "*aula*" [classroom], "*testo*" [text]. These nouns have been extracted automatically by associating to each noun in the corpus its highest hyperonym in MultiWordNet [12].

As test set we have 444 sentences randomly extracted from La Repubblica corpus containing a deverbal noun. They have been manually annotated by the authors: 281 sentences contain an eventive occurrence of a deverbal noun while 163 contain a non-eventive one.

The features' extraction has been automatically performed on a dependency parsed version of the datasets [13].

3.1 Experiment 1: type occurrences and token occurrences of eventive and non-eventive nouns

We apply the J48 classifier provided by Weka, with La Repubblica data as training set. Distributional patterns have been largely used to find semantically related nouns at type level in large corpora [14] and have proved their utility for semantic classification tasks. For instance, [15] obtain an accuracy of 75% for the classification of eventive nouns. But the reverse is true: from previously classified semantic items discriminative distributional patterns for token occurrences can be induced.

We have considered as baseline the most frequent class as the correct one, i.e the eventive reading, which corresponds to the 63.2%. The accuracy obtained against the test set is 71.5%, which ouperforms the baseline of 8 points, with an overall F-measure of 0.69. If we split the results on the basis of the readings, or classes, of the deverbal nouns, the results show that the classifier performs better on eventive readings (F-measure = 0.80) than on non-eventive ones (F-measure = 0.51). Detailed results are reported in Table 5 under the heading "*La Repubblica*".

3.2 Experiment 2: token occurrences as training

The second experiment uses the It-TimeBank corpus as training set. The results are lower than those obtained when using the La Repub-

Table 5.

Noun type	La Repubblica				It-TimeBank			
	Accuracy	Precision	Recall	f-meas.	Accuracy	Precision	Recall	f-meas.
event reading		0.72	0.88	0.80		0.63	1	0.77
not-event reading		0.68	0.41	0.51		0	0	0
event + not-event	71.5%	0.70	0.71	0.69	63.5%	0.40	0.63	0.49

blica Corpus. We obtain an overall accuracy of 63.5% (F-measure = 0.49), which is very close to the baseline. It is striking to observe how with this highly supervised training set the classifier performance is worse. In particular, no non-eventive reading of the nouns in the test test is correclty classified. The details are reported again in Table 5 under the heading "*It-TimeBank*".

3.3 Experiment 3: POS sequences as disambiguating cues

Event noun detection for event extraction systems is partially akin to word sense disambiguation because the aim is to test algorithms for automatic detection/identification of nouns denoting events in context. Methodologies that proved their utility for WSD tasks can be tested on event nouns detection in context. For this reason, we evaluate the relevance of single POS preceding or following our key words, performing classifications even on the basis of sequences of POS, a methodology that [7] reputed partially good for WSD of nouns. More generally, our aim is to test the role of POS sequences as not theoretically predetermined features that are similar, in terms of structural information, to more specific patterns listed by linguists. The results are discouraging (see Table 6), showing that even wider POS sequences as 5-grams are not able to help in this classification task.

Table 6. POS n-grams as disambiguating cue.

POS sequence	Accuracy - La Repubblica	Accuracy - It-TimeBank
P-1, P0, P+1	41%	70%
P0, P+1, P+2	63.2%	63.2%
P-2, P-1, P0	63.2%	63.2%
P-2, P-1, P0, P+1, P2	37.3%	63.2%

4 Related works

In recent years there has been an increasing interest in the NLP community for automatic event identification as the development of different systems for the identification of event nouns shows [16], [17] [18] [19] among others).

To the best of our knowledge, our methodology (and the results obtained) can be directly compared with [20], even if they did not focus specifically on deverbal nouns. They propose a weakly-supervised method for detecting nominal events mentions that classify noun phrases on the basis of a combination of word sense disambiguation and lexical acquisition techniques. Our training and test sets are smaller but we show how, with a list of linguistically informed cues, our methodology slightly outperforms their results for eventive reading of deverbal nouns (88% *vs.* 87.7%) while is lower for non-eventive ones (41% *vs.* 60%).

Finally, comparing our results with [19] is not possible because precision and recall are reported for the component of the classifier that integrates information from a lexical resource with information extracted from a corpus. Using just corpus data, as we did in our experiments, they report an accuracy of 80%, which yields an accuracy which is lower than their baseline (82.1%). In the overall, our classifier seems to be better for the classification of eventive readings of deverbal nouns, while it seems less promising for the classification of non eventive nouns. This may be due also to the fact that the features' set we have identified is mainly focused on eventive readings.

5 Conclusion and future works

The availability of methodologies able to identify the correct denotation of deverbal nouns is essential because it can help to build better event extraction system but it can also improve the performance of more complex NLP systems such as anaphora resolution, subcategorization frames, paraphrase detection and temporal processing. Our classifier can be integrated in a broader event extraction system for Italian but it can be used also for automatically annotate or add semantic information to large corpora reducing the manual effort and costs for their realization.

In this paper we show how to classify deverbal nouns in context as eventive or non eventive using syntagmatic and collocational information relative to past encounters of nouns tagged with the help of a lexical resource such as MultiWordNet.

We have showed that linguistically informed syntagmatic and lexical patterns perform better than POS sequences, at least for this task.

Future work will focus on automatic identification of nouns denoting events, going beyond the present case study on deverbal nouns. Of course, some integrations in the fatures' set to improve the identification of non-eventive readings are necessary, together with a more detailed classification of these occurrences (e.g. result/state *vs.* concrete object). The role of manually annotated data as training set such as the It-TimeBank is not clear due to its dimension with respect to the class of deverbal nouns. With a richer training set manually annotated we will gain clearer evidence on the utility of annotated corpora.

Acknoledgments

This work has been supported by the EU founded project PANACEA (Grant agreement no.: 7FP-ITC-248064).

References

1. Witten, I.H., Frank, E.: Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann (2005)
2. Pustejovsky, J., Castao, J., Ingria, R., Saurì, R., Gaizauskas, R., Setzer, A., Katz, G.: TimeML: Robust specification of event and temporal expressions in text. In: Fifth International Workshop on Computational Semantics (IWCS-5). (2003)
3. Baroni, M., Bernardini, S., Comastri, F., Piccioni, L., Volpi, A., Aston, G., Mazzoleni, M.: Introducing the "`la Repubblica`" corpus: A large, annotated, `TEI(XML)`-compliant corpus of newspaper italian. In: Proceedings of the Fourth International conference on Language Resources and Evaluation (LREC-04). (2004)
4. Vendler, Z.: 4. In: Linguistics in Philosophy. Cornell University Press, Ithaca, NY (1967)
5. Grimshaw, J.: Argument Structure. MIT Press, Cambridge, Massachusetts (1990)
6. Alexadiou, A.: The Functional Structure in Nominals. Nominalization and Ergativity. John Benjamins, Amsterdam/Philadelphia (2001)
7. Mohammad, S., Pedersen, T.: Combining lexical and syntactic features for supervised word sense disambiguation. In: Proceedings of the Conference on Computational Natural Language Learning. (2004) 25–32

8. Gaeta, L.: Nomi d'azione. In Grossmann, M., F., R., eds.: La formazione delle parole in italiano. Niemeyer, Tubingen (2004) 314–351
9. Pustejovsky, J.: The Generative Lexicon. MIT Press (1995)
10. Peris, A., Taulé, M.: Evaluaciòn de los criterios linguìsticos para la distinciòn evento y resultado en los sustantivos deverbales. In: Proceedings of the 1st International Conference on Corpus Linguistics (CILC-09). (2009)
11. Pustejovsky, J., Verhagen, M.: Semeval-2010 task 13: Evaluating events, time expressions, and temporal relations (tempeval-2). In: Proceedings of the Workshop on Semantic Evaluations: Recent Achievements and Future Directions (SEW-2009), Boulder, Colorado, Association for Computational Linguistics (June 2009) 112–116
12. Pianta, E., Bentivogli, L., Girardi, C.: Multiwordnet: Developing and aligned multilingual database. In: Proceedings of the First International Conference on Global WordNet, Mysore, India (January 2002) 293–302
13. Attardi, G., Dell'Orletta, F.: Reverse revision and linear tree combination for dependency parsing. In: NAACL-Short '09: Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics, Companion Volume: Short Papers, Morristown, NJ, USA, Association for Computational Linguistics (2009) 261–264
14. Hindle, D.: Noun classification from predicate-argument structures. In: Proceedings of the 28th Annual Meeting of the Association for Computational Linguistics, Pittsburgh, Pennsylvania, USA, Association for Computational Linguistics (June 1990) 268–275
15. Qian, T., Durme, B.V., Schubert, L.: Building a semantic lexicon of english nouns via bootstrapping. In: Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics, Companion Volume: Student Research Workshop and Doctoral Consortium, Boulder, Colorado, USA (june 2009)
16. Saurí, R., Knippen, R., Verhagen, M., Pustejovsky, J.: Evita: A robust event recognizer for qa systems. In: Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing (HLT/EMNLP-05). (2005) 700–707
17. Resnik, G., Nùria, B.: Automatic detection of non-deverbal event nouns in spanish. In: Proceedings of the 5th International Conference on Generative Approaches to the Lexicon, Pisa: Istituto di Linguistica Computazionale (2009)
18. Eberle, K., Faass, G., Heid, U.: Corpus-based identification and disambiguation of reading indicators for german nominalizations. In: Corpus Linguistics 2009, Liverpool, UK (2009)
19. Peris, A., Taulé, M., Boleda, G., Rodrìguez, H.: Adn-classifier:automatically assigning denotation types to nominalizations. In Calzolari, N., Choukri, K., Maegaard, B., Mariani, J., Odijk, J., Piperidis, S., Rosner, M., Tapias, D., eds.: Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10), Valletta, Malta, European Language Resources Association (ELRA) (May 2010)
20. Creswell, C., Beal, M.J., Chen, J., Cornell, T.L., Nilsson, L., Srihari, R.K.: Automatically extracting nominal mentions of events with a bootstrapped probabilistic classifier. In: COLING-ACL '06: Proceedings of the COLING/ACL on Main conference poster sessions, Morristown, NJ, USA, Association for Computational Linguistics (2006) 168–175