DEVERBAL NOUN COMPLEMENTATION RULES APPLIED TO SEMANTIC ROLE LABELING

Liliana MAMANI SÁNCHEZ
Instituto Militar de Engenharia, Rio de Janeiro, Brazil

Claudia OLIVEIRA
Macquarie University, Sydney, Australia

RESUMO

Este trabalho apresenta um método de etiquetagem de papéis semânticos em predicações nominalizadas. O método evita o problema da não disponibilidade de recursos computacionais robustos para o português, explorando a correspondência sistemática entre as estruturas argumentais de verbos e das nominalizações correspondentes, e utilizando a Web como corpus *para inferir probabilisticamente a delimitação dos argumentos e suas relações gramaticais dentro da estrutura sintática do verbo base.*

ABSTRACT

This work presents a procedure of semantic role labeling in deverbal noun predications. The procedure avoids the problems related to the unavailability of robust computational resources for Portuguese by exploring the systematic correspondence between the argument structures of verbs and of their corresponding nominalizations, and by using the Web as a corpus in order to infer probabilistically the boundaries of the arguments and their grammatical relations in the verbal syntactic structure.

PALAVRAS-CHAVE

Nominalização deverbal. Estrutura argumental nominal. Etiquetagem de papéis semânticos. Semântica computacional. Lingüística computacional.

KEY WORDS

Deverbal nominalization. Noun argument structure. Semantic role labeling. Computational semantics. Computational linguistics.

Introduction

The objective of this work is to describe the automatic identification and classification of the argument structure of a noun phrase (NP) whose head is a deverbal noun, using the argument structure of the base verb and a set of rules, described in Meyer (1991), that map the verbal complementation pattern to the corresponding noun complementation pattern. These rules are based on the regularities of verbal complementation within certain semantic verb classes. The research has produced information for a frame-like description of nominalizations to be stored in a lexicon that is able to aid the automatic labeling of semantic roles in Portuguese texts.

Semantic Role Labeling (SRL) is a high level language processing task that has taken center stage in the Computational Linguistics research community in the last few years, mainly because many sophisticated applications require language understanding, which is intrinsically linked to the semantics of predication. In addition, SRL often requires a wealth of language resources, such as treebanks, lexicons, parsers, and chunkers, which are only now becoming freely available.

In general, natural language predication occurs in the domain of a sentence, the linguistic unit built around a verb. The semantic role of a sentence constituent is determined by its relationships to the predicate's head, the verb. The concept of "semantic role", also referred to as thematic role or theta-role in linguistic theories, provides the link between the syntactic and the semantic levels of linguistic analysis.

The sets of semantic roles that can be assigned to sentence constituents vary enormously. As with most semantic categories, many different role sets have been proposed by numerous theoretical frameworks, ranging from a small set of very abstract and general roles (van Valin, 2004; Dowty, 1991), to large sets of specific roles which are particular to certain verbs or verb groups, as in the FrameNet project (Johnson and Fillmore, 2000). Nevertheless, some characterization of a small group of semantic roles, such as agent, experiencer, theme (patient), instrument and location, never fail to be included.

In this work we propose a procedure that assigns semantic roles to phrases that participate in the argument structure of a deverbal noun predication. The main procedure includes two phases: first, the identification of arguments, and second, the classification of arguments. At the identification phase, the prepositional phrases attached to the nominalization are parsed. At the argument classification phase, the predicate information repository in its nominal version and a selection of linguistically based heuristics are used to assign semantic roles to the prepositional phrases. The procedure has been tested and evaluated using a list of deverbal nouns.

The remainder of the paper is organized as follows. In section 1, we present the linguistic aspects involved in the formation, the behavior, and the semantic characterization of deverbal noun predicators; in section 2 the computational task of semantic role labeling is described, first, in general terms and, then, specifically in the case of noun predicators; in section 3 we describe the SRL procedure proposed and in section 4 we report the experimental results obtained in the evaluation of the implementation; section 5 brings our conclusions and directions for future work.

1 Deverbal noun predicates: linguistic aspects

From a logical perspective, a predicate is an expression that can be true of something, expressing either a property (a unary relation) or a relation between objects, and in this sense, only verbal expressions are considered predicates. In traditional grammar, a sentence is formed by a *subject* and a *predicate* containing a verb, and the verb's required or optional arguments: objects (direct, indirect, prepositional), predicatives and adverbials (either obligatory or adjuncts). On the other hand, deverbal nominalizations enable the reference to a verbal process regardless of the particular circumstances of the verbal predication, such as person, tense, mood, etc., and can, therefore, be used to construct the expression of a predication.

1.1 Deverbal nominalization

Deverbal nominalization is a process that affects morphological, syntactic and semantic properties of lexical items, involving, in many cases, derivation by suffixation. In Portuguese it seems that the verb and its nominalization have a paradigmatic relationship (Basilio, 1980), that is, they are part of "a morphological structure represented as a set of paths between a base and appropriate morphological operations" (Pounder, 2000). Basilio's hypothesis derives from the fact that a large proportion of verbs in Portuguese have regular nominalizations that maintain a clear syntactic-semantic correspondence with their verbal base in terms of argument structure and semantic roles. Verbs that do not have a nominal counterpart can be characterized as colloquial verbs, copula verbs, or verbs that have been diachronically blocked. In contrast to nominalization, word formation processes that generate denominal verbs are syntactically and semantically unpredictable.

From an empirical perspective, the nominalization process is dominant, amongst Portuguese word formation processes, in productivity. In Sandman (1989), the author shows that nominalizations are produced twice as much as derived verbs and adjectives together. Table 1 shows some of the most productive deverbal nominalization processes in Portuguese.

TABLE 1 - Deverbal nominalization by suffixation.

Morphology	***Semantics***	***Examples***
X_V -> [X_V *ção*]$_N$	to designate action or result	produzir -> produção, falar -> falação *produce -> production, speak -> speech*
X_V -> [X_V *mento*]$_N$		treinar -> treinamento, comprometer -> comprometimento *to train -> training, to commit -> commitment*
X_V -> [X_V *agem*]$_N$		lavar-> lavagem, montar -> montagem *to wash -> washing, to assemble -> assemblage*
X_V -> [X_V *dor*]$_N$	to designate agent or instrument	atirar -> atirador, secar -> secador *to shoot -> shooter, to dry -> drier*
X_V -> [X_V *nte*]$_N$		vigiar -> vigilante, desinfetar -> desinfetante *to watch -> watchman, to desinfect -> desinfectant*

It is possible that a verb provides the base for more than one nominalization, in particular when a nominalized form acquires a particular meaning and there is the need to express another meaning: *casar* (to marry) -> *casamento* (marriage), *casório* (marriage, informal); *ondular* (to ondulate) -> *ondulação* (ondulation), *ondulamento* (hair waving). As a reference to the verbal process (Ex. 1.b.), verb nominalizations are abstract nouns fulfilling a function of textual construction and cohesion, in which case the nominal construction may preserve the arguments of the verbal construction. On the other hand, nominalizations can also fulfill a designating role by, in many cases, incorporating the verbal object in order to denote an entitative concept, undergoing a concretization of their meaning through a metonymic association (Ex. 1.c.). In this case, the arguments of the verb are frequently omitted.

Example 1:

a. A editora publicou rapidamente este livro.
   (The publisher published quickly this book.)

b. A rápida publicação deste livro pela editora foi comemorada.
   (The quick publication of the book by the publisher was celebrated.)

c. Esta publicação vai esgotar-se rapidamente.
   (This publication will sell out quickly.)

1.2 Deverbal noun predication

Predication is the assignment of properties to things by means of language (Foltran, 2003), and the predicator is its linguistic expression. Traditional grammar divides declarative sentences into *subject* and *predicate*, the first part being the recipient or subject of the properties, and the second part containing the relational or nominal properties being assigned: the subject is the entity about which a statement is made; the predicate is all that is said of the subject (Said Ali, 1966, p. 105). A syntactic-semantic relationship is made explicit in Mateus et al.'s definition: the subject is the syntactic function of the constituent that occurs as external argument of the predicator (Mateus et al., 1989, p. 161), i.e. outside the domain of the verb phrase (Crystal, 2003).

The process of deverbal nominalization in Portuguese often produces, on the syntactic-semantic level, an argument structure that bears a regular relationship to the base verb's argument structure, preserving both the valence of the head noun and the semantic functions of the arguments. Thus, the nominalized noun is a predicator that is in an embedded construction, acting as a term in a predication of a higher level (Camacho and Santana, 2004). A great deal of work, in particular from a functionalist perspective, has explored the argument structure of nominalizations. The work described in the present paper is mostly based on Meyer (1991), which proposes a set of formal rules that map verbal and nominal argument structures according to a typology introduced by Peres (1984).

Deverbal noun predicators are subject to the same unclear distinction between arguments (obligatory) and adjuncts (optional) as their verbal counterparts. We adopt Meyer's distinction between core (primary) and optional (secondary) arguments, and follow her choices with respect to the argument structures of the predicators in Peres's typology. It is also worth pointing out that deverbal noun predicators may be less strict with respect to obligatory arguments: even when the transitive base verb requires an obligatory direct object, the nominalized construction often does not realize it syntactically. This may be due either, for example, to the anaphoric omission of the verbal argument, or to the concretization of the nominalization.

1.3 Semantic Roles

A semantic role is the conceptual relationship between the argument and the predicator in a clause. Even though the theoretical status of semantic roles is still an unresolved issue in Linguistics, the most common understanding is that semantic roles are semantic/conceptual elements (Jackendoff, 1972), somewhere in the semantics-syntax interface.

In many linguistic theories semantic roles are assumed to be the source of, or to constrain, grammatical relations, assuming that grammatical relations are projected from predicate argument structures represented in the lexicon. The idea dates back to Fillmore (1968), who first proposed that subject selection is sensitive to a hierarchy of "cases", i.e. semantic relations. According to Fillmore, there is a mapping between a list of semantic relations and a list of grammatical relations which is controlled by a correspondence strategy. For instance, Fillmore puts forwards the following hierarchy for subject selection:

Agent < Instrumental < Objective.

If the case frame (argument structure) of a verb contains an Agent, then it is realized as the grammatical subject, otherwise Instrumental is promoted to subject, and so forth. To illustrate, ex. 2 shows different senses for the verb *quebrar* (to break), an ergative verb, i.e. a verb whose subject when intransitive corresponds to its direct object when transitive (Crystal, 2003). In 2.a. the subject is the Agent, in 2.b. it is the Instrumental and in 2.c. it is the Objective.

Example 2:

a. Monael *quebrou* o computador com o martelo.
   (Monael broke the computer with the hammer.)

b. O martelo *quebrou* o computador.
   (The hammer broke the computer.)

c. O computador *quebrou*.
   (The computer broke.)

Fillmore's work has had an enormous influence in Computational Linguistics, especially with the development of Frame Semantics (Fillmore and Atkins, 1992), which has become one of the standards of lexical-semantic knowledge representation.

Common characteristics of the different approaches to semantic roles and their linguistic functions are (Dowty, 1991):

- completeness - every argument of every predicator is assigned a semantic role;
- uniqueness - every argument of every predicator is assigned only one semantic role;
- distinctness - every argument of every predicator is distinguished from the other arguments by the role it is assigned;
- independence - each role is given a consistent semantic definition that applies to all predicators and all situations. Thus, role definitions do not depend on the meaning of the particular predicator or on the other semantic roles it assigns.

There is no agreement on how many roles are needed and what they are, which may be an indication that the semantic role list approach is still rather immature. Proposals range from just a few to hundreds of them. A related problem is what Dowty (1991) calls "role fragmentation", illustrated by the role Agent, "one of the most frequently cited roles, and it is in some sense a very intuitive role, but it is one of the hardest to pin down" (Dowty, 1991:553). Jackendoff (1983:176) distinguishes Agent from Actor by the willfulness or intentionality of the participant, a similar criteria behind van Valin's (1990) distinction between Agent and Effector; Cruse (1973:18-21) divides agency into four types: volitive, effective, initiative, and agentive; Lakoff (1977) proposes up to fourteen different roles as types of agent.

In this work we use the list of roles proposed by Peres – CAUSER, OBJECT, EXPERIENCER, OWNER, GIVER, RECEPTOR, PLACE, SOURCE, and GOAL – which presents a good compromise between the distinction of the semantic functions in the clause and the effectiveness of the Semantic Role Labeling procedure we propose.

2 Semantic role labeling

Grammatical descriptions frequently resort to the categorization of words and linguistic expressions according to features that locate them in a given linguistic system (Oliveira and Freitas, 2006). In Computational Linguistics, labeling (also known as tagging or annotation) is the task of assigning such categories, and involves the determination of the boundaries of the expressions and the choice of which label to assign them. Text labeling can be manual or automatic. At this point in time, the first approach results more accurate and reliable but considerably more costly.

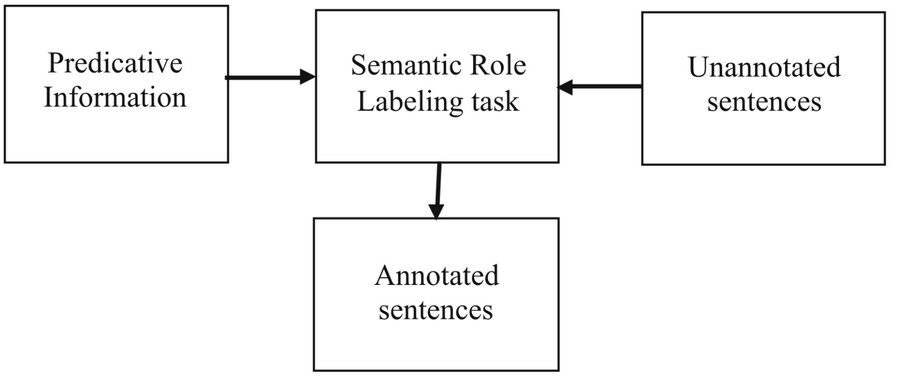

FIGURE 1 - The semantic role labeling process.

Semantic Role Labeling (SRL) is the computational task of assigning semantic roles to portions of text. As such, it is a kind of semantic analysis that can be of great relevance in Natural Language Processing (NLP) Systems such as question answering systems, machine translators, language generators and other semantic-centered resources and tools. In recent years a number of SRL systems have been developed using different approaches and technologies: from semi-automatic to fully-automatic systems; from hand-coded probabilistic techniques, such as Gildea and Jurafsky's (2002) for verb arguments and Lapata's (2002) for deverbal noun arguments, to Machine Learning techniques (Carreras and Márquez, 2005).

As a computational process (Fig. 1), SRL takes as input the predicative information stored in a lexical resource, and a set of unannotated sentences; the output is a set of sentences with semantic information attached to the arguments of the predications. In the case in focus, the predicative information is relative to nominalizations, and instead of sentences the input and the output are noun phrases.

2.1 Semantic role labeling of deverbal noun predicates

The SRL task in general can be methodologically divided in two subtasks, the argument identification phase and the argument classification phase, both particularized for deverbal noun predications as follows.

1. The argument identification phase aim is to parse the predication into a nominalized head and its arguments in the form of prepositional phrases. In this phase the main problem is the prepositional phrase attachment task, which consists in deciding how to group correctly the sequence of prepositional phrases in an expression.
2. The argument classification phase attempts to assign a semantic role to the arguments that were recognized in the identification phase.

Example 3 below illustrates the input/output of the identification and classification phases.

Example 3:

Input string: a transformação do tricampeão mundial em uma lenda no país mais moderno do planeta.
(the transformation of the three times world champion in a legend in the most modern country in the world)
identification phase output: 1. do tricampeão mundial (of the three times world champion); 2. em uma lenda (in a legend); 3. no país mais moderno do planeta (in the most modern country in the world).
classification phase output: 1. object; 2. goal; 3. place.

There is a variety of linguistic features and information that may be used in SRL procedures. Apart from other annotation tools for part-of-speech, named entities (proper nouns) and chunkers, the main resource involved in SRL is some kind of lexical database containing the predicate-argument structures of the predicators. It is worth mentioning two such resources: the FrameNet and the PropBank.

The Berkeley FrameNet project is an ongoing effort of building an on-line lexical resource for English, based on frame semantics and supported by corpus evidence (Baker et al., 1998). Its aim is to document the range of semantic and syntactic valences of each word in each of its senses, through computer-assisted annotation of example sentences and automatic tabulation and display of the annotation results. PropBank, on the other hand, is a corpus annotated with verbal predications and their arguments (Palmer et al., 2005) based on Levin's verbal classification (Levin, 1993). As by-products, both FrameNet and PropBank can be used to aid SRL systems.

Such resources are unfortunately not available for Portuguese. In order to overcome this difficulty we have built a small set of verbal predicate-argument structures and implemented a procedure that derives the corresponding deverbal nouns predicate-argument structures.

2.2 Inferring deverbal noun complementation from verbal complementation

Meyer (1991) proposes the formalization of nominal complementation, based on the analysis of the structure and the semantic-pragmatic-syntactic behavior of nominalizations. The author proceeds by mapping argument structure slots of both predication forms, nominal and verbal, with respect to order of arguments and selection of preposition, which results in an inventory of predicators organized with the following information: i. predicator lexical form; ii. part of speech; iii. required number of arguments; iv. semantic functions of arguments; v. selectional constraints over the arguments.

Meyer categorizes predicators in four macro-classes – actional, processual, posicional and stative predicators – each further divided into four classes – experiencial, possessives, locatives and basic predicators. This last class is intended to cover the unmarked relations, those that do not involve the participation of entities that could be interpreted as "experiencers", "owners", "givers", etc. Despite the division into classes, predicators share several properties, especially syntactic properties, which makes it possible to determine common syntactic-semantic structures. Meyer explores this characteristic in order to build a set of seven rules summarizing the ordering of arguments and the selection of preposition in deverbal nominalization complementation. The main elements of the rules are:

1. the categories of the predicators to which the rule apply (AC, BA, EX, PS, etc);
2. slots defining the syntactic structure of the constructions containing the nominalization as predicator and the verb as predicator (E1, E2, …);
3. the syntactic function of the element filling the slots in the constructions: subject (SU), direct object (OD), indirect object (OI), etc.;
4. the semantic function for each slot: causer (C), object (O), giver or dative (D), etc., and X representing an unspecified semantic function;
5. prepositions, lexicalized or with a general label (prep), preceding an argument slot;
6. correspondence between slots, indicated by a superscripted index (1, 2, …);
7. a relationship symbol "->", meaning that the structure that precedes this symbol yields the structure following the symbol;
8. arguments (A1, A2, …).

```
@          AC    BA     (A1)C       (A2)O       (A3)X
           EX
           PS
-> (E1)1 SU  V  (E2)2 OD prep (E3)3 OI
           :  (E1) = (A1)
              (E2) = (A2)
              (E3) = (A3)
           -> FNDS de (E1)2 ? prep (E2)3? por  (E3)1?
                   :  (E1) = (A2)
                      (E2) = (A3)
                      (E3) = (A1)
```

FIGURE 2 - Rule for action, basic, experiential, or possessive predicators.

An example of rule (Meyer's rule 3) is shown in Fig. 2, where the argument structure is composed by the arguments A1, A2 and A3 with the semantic function of causer (C), object (O) and X, respectively. Slots E1, E2 and E3 determine the syntactic structure of the predication. E1 takes the function of subject (SU), E2 takes the function of direct object (OD) and the slot E3 takes the function of indirect object (OI). These slots have to be filled by the arguments that perform some semantic function related with the predicator V: E1, E2 and E3 filled with A1, A2 and A3 respectively.

From the elements identified in the verbal predication and the rule in figure 2, it is possible to make inferences about the nominal complementation structure: the suffixal deverbal nominalized form (FNDS) is the predicator, the sentence's E1 is mapped onto FNDS's E2, the sentence's E2 is mapped onto FNDS's E3, and the sentence's E3 is mapped onto FNDS's E1. The mapping makes it possible to create correspondences between the arguments filling the slots in the verbal predication and those filling slots in the nominal predication. For example, in the second structure, the argument A2, which fills E1, has the semantic function of object; A3, which fills E2, has the semantic function X; and A1, which fills E3, is a causer preceded by "*por*" (*by*).

2.2.1 Predicative information repository

In order to build the necessary argument structure information input for the method, we characterize the predicative information for a given predicator as: i. a set *P* of semantic roles for the predicator; ii. a set *Prep* of prepositions; iii. the essential arguments; iv. the set S of syntactic schemes for different senses of the predicator. Two syntactic schemes s_i and s_j of *S* can have the same syntactic structure but different senses, which is characterized, in the present procedure, in terms of semantic role differences.

Thus, the verbal predicative information repository, which stores the input information for the rules of nominalization, contains the description of specific verbs, their different senses and their respective syntactic structures. From verbal predicate information and using the inductive rules, we formulate an algorithm for deriving nominalization predicative information. The algorithm is shown in Fig. 3.

```
Let v be the verbal predicator,
   Choose an appropriate rule R for the predicator class of v
   Generate from v its abstract nominalization n
   Retrieve the syntactic scheme set S from the predicative information
   repository for v
   For each syntactic scheme s_i from S
        Apply R for getting the syntactic scheme s'_i
        Append s'_i to the syntactic scheme set S'
   Write down the predicative information for n with S'
```

FIGURE 3 - Predicative information derivation algorithm.

2.2.2 Repository of phrases with nominalized predicators

The selected phrases for the repository were obtained from two resources: Corpus NILC/São Carlos (Santos and Sarmento, 2003) and corpus MAC-MORPHO (Marchi, 2003), both annotated with part of speech information, useful in cases where the nominalization and the verbal forms are identical. The selection of phrases was made amongst singular forms only, given the concrete nature of plural forms.

The construction of the predicative information and of the phrase repository followed Meyer's recommendations strictly, except with respect to the concrete/abstract senses of deverbal nouns and the distinction between core and optional arguments.

3 The proposed labeling procedure

The SRL task can be understood in an abstract way as follows. Let *R* be a set of semantic roles for the verb *v* and its nominalization *Nom*; let *A* be the set of arguments (core and optional) in a predication *Nom* $comp_1\ comp_2 \ldots comp_n$; then the semantic role labeling of the predication is a relation between *R* and *A* with the properties: i. each role in *R* can be related to at most one argument in *A*; and ii. each argument in *A* can be related to at most one role in *R*.

The sub-tasks in the procedure can be viewed in Fig. 4 below. Previously to the argument identification and classification phases, there is a need for a **pre-processing phase**, when the input phrases are parsed and the necessary linguistic information is gathered from the parsed trees. We used the parser PALAVRAS (Bick, 2000), a well-known Portuguese NLP tool, available on the Internet.

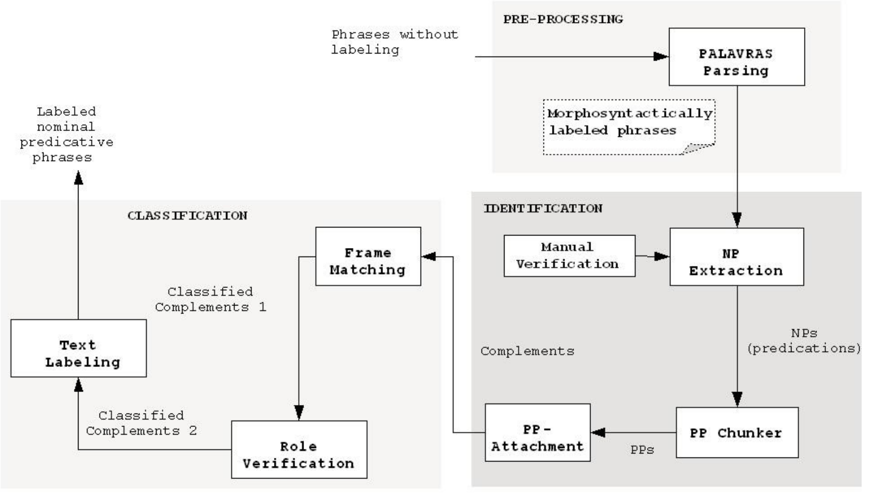

FIGURE 4 - Block diagram of the proposed procedure.

Even though the reported precision in PALAVRAS is 99% for part of speech and 96% for syntactic tagging, we suspect that the parser's precision rate drops when considering only phrases that require prepositional phrase attachment, which is one of the main challenges in the argument identification phase.

The goal in the argument identification phase is to extract the prepositional phrases sp_1, sp_2, ..., sp_n containing the arguments of the nominalized predicator. It consists of three sub-tasks: i. the delimitation of the noun phrase whose head is the deverbal noun; ii. chunking the noun phrase into head and simple prepositional phrases, i.e. containing a single preposition; iii. building the list of arguments by deciding the prepositional phrase attachment problems.

It is worth noting that the identification phase starts with the manual verification of the pre-processed phrases: the research reported here intended to evaluate the SRL procedure itself, and failing to review the pre-processed input phrases would mean that the error produced by the parser would be amplified in the final results.

Chunking the noun phrase into simple prepositional phrases is straightforward in most cases. The exceptions have been analyzed in detail in (Mamani Sánchez, 2007). From a sequence *Nom* $sp_1, sp_2, \ldots, sp_n$, where each sp_i is a simple prepositional phrase $sp_i = prep_i\, n_i$ there is a number of parsing possibilities, as illustrated by Ex. 4. The phrase in 4.a. is ambiguous between two bracketing possibilities 4.b. and 4.c., in other words, the prepositional phrase *dos democratas* can be attached to either *o envio* or to *a campanha*.

Example 4:

a. o envio para a campanha dos democratas
(the postage for the campaign of the democrats)

b. o envio [para a campanha] [dos democratas]

c. o envio [para a campanha dos democratas]

This problem can be formally understood as follows. Given the 5-tuple ($Nom, prep_1, n_1, prep_2, n_2$), the prepositional phrase attachment problem (PPAP) consists in deciding whether $sp_2 = prep_2, n_2$ is attached to *Nom* or to n_1 (attachment sites).

It is notoriously hard to obtain good results for the PPAP. Robust linguistic resources such as treebanks and ontologies can be helpful, but for languages in which these resources are unavailable the Web has been increasingly used in cunning ways. Volk (2001) has proposed a set of simple statistics of term co-occurrence that has been used in disambiguation tasks. We adapted Volk's measures for the PPAP, based on the assumption that, in the Web, the most frequent combination, either *Nom* sp_2 or $n_1\ sp_2$, should be the choice. The adapted measures are:

Measure 1. $cooc(N, prep_2, n_2) = freq\ (N, prep_2, n_2)\ /freq(N)$

Measure 2. $cooc(N, prep_2) = freq\ (N, prep_2)\ / freq(N)$

The measures are calculated for $N = Nom$ and $N = n_1$. Measure 1 is more strict since n_2 might be an infrequent word and even the Web might be statistically insufficient. When this is the case, measure 2 is employed, taking into account only the affinity of the preposition with either *Nom or* $N = n_1$.

The PPAP decision algorithm is shown in Fig. 5 below. As an illustration, considering Ex. 4, the algorithm would calculate the frequencies of the phrases *envio dos democratas* (*postage of the democrats*) and *campanha dos democratas* (*campaign of the democrats*).

```
Function attachment (n0, n1, prep2, n2 )

If cooc(n0, prep2, n2) ≠ 0 and cooc(n1, prep2, n2) ≠ 0
        If cooc(n0, prep2, n2) > cooc(n1, prep2, n2)
                assign attachment to n0
        else  if cooc(n0; prep2; n2) < cooc(n1; prep2; n2)
                      assign attachment to n1
else if cooc(n0; prep2; n2) > arbitrary threshold
        assign attachment to n0
else if cooc(n1; prep2; n2) > arbitrary threshold
        assign attachment to n1
else
        undefined attachment.
```

FIGURE 5 - Attachment decision algorithm.

What remains to be done in this phase is to construct a list of arguments according to the attachment decisions made by algorithm in Fig. 5. Considering Ex. 4, if the results are in favor of *campanha dos democratas,* the list of arguments would consist of {*para a campanha dos democratas*}; otherwise, if *envio dos democratas* is found to be more frequent, the list of arguments would turn out to be {*para a campanha, dos democratas*}. It is worth pointing out that, if the attachment of $prep_2$ n_2 to the competing noun heads remains undefined, neither n_0 $prep_2$ n_2 nor n_1 $prep_2$ n_2 will be included in the list of arguments produced by the identification phase.

The next phase in the procedure is the argument classification phase, which uses the predicate-argument information and the list of identified arguments to perform a frame matching procedure, similar to Swier and Stevenson's (2004). The assignment algorithm is shown in Fig. 6, and is applied to core arguments only.

Let *Nom* be the predicator and *listaSP* be the list of prepositional phrases $p_i\ np_i$ constructed in the identification phase.

Let C = {$<pv_1, SR_1>, <pv_2, SR_2>, ..., <pv_r, SR_r>$} a subcategorization frame belonging to the predicative information repository for *Nom*, where $<pv_j, SR_j>$ is a pair (preposition, semantic role).

For each $p_i\ np_i$ of *listaSP*
    For each $<pv_j, SR_j>$ in C
        If $p_i = pv_j$, $p_i \neq pv_k$ for $k \neq j$, and SR_j has not yet been assigned to an argument
            Assign the semantic role SR_j to np_i and mark SR_j as assigned.
    If np_i was not assigned some semantic role
        Mark np_i with a generic role of adjunct.
Finish.

FIGURE 6 - Semantic role assignment (frame matching) algorithm.

There are cases of arguments that cannot be assigned a role by the frame matching procedure because: i. two roles require the same preposition (ambiguity); or, ii. the information available is not sufficient to decide the assignment. In these cases it is necessary to perform a role verification procedure, using the possible grammatical relations fulfilled by the arguments in the verbal construction. For instance, prepositions *de* (of) and *por* (by) frequently introduce, in the nominal construction, the verb's subject; preposition *de* (of) is also frequent before objects. These considerations are the basis for a procedure, defined in figure 7 below, designed to increase the role labeling rate, using once again the Web as a corpus.

The procedure operates on the basis of three statistical measures of co-occurrence, which are applied to each argument that has not had a role assignment, where *N* is the argument's head noun and *V* is the base verb of the deverbal noun.

1. Subject co-occurrence (cooSuj): verifies the plausibility of the construction *N V;*
2. Object co-occurrence (cooObj): verifies the plausibility of the construction *V N;*
3. Object co-occurrence (cooObjInd): verifies the plausibility of the construction *V prep N.*

Considering example 5 below, both [civis] (*civilians*) and [capital] (*capital*) can be labeled with either OBJECT or ORIGIN. The ambiguity might be resolved by counting and comparing the frequencies of the phrases a., b. and c. and their inflexional variations.

Example 5:

A ONU cancelou a retirada de [civis] d[a capital] por causa de combates de artilharia.
(The UN cancelled the removal of civilians from the capital because of artillery combats.)

a. [civis retiram] (*civilians remove*) and [a capital retira] (*the capital removes*);
b. [retira civis] (*remove civilians*) and [retira a capital] (*remove the capital*);
c. [retira de civis] (*remove from civilians*) and [retira da capital] (*remove from the capital*).

Let $P = \{SR_1, SR_2, \ldots, SR_r\}$ be the semantic role list for the predicator *Nom*, let $C = \{comp_1, comp_2, \ldots, comp_s\}$ be a complement list for *Nom* (typically, the complements that remained unlabeled by the semantic role assignment procedure).

To each semantic role SR_i there is an associated subset of C, henceforth $lComp_i$, containing the complements that may be labeled with SR_i (identical prepositions).

To each complement $comp_j$ there is an associated subset of P, henceforth $lPap_j$, containing the semantic roles that may be assigned to $comp_j$ (identical prepositions).

1. For each semantic role SR_i in P:
   a) If the list $lComp_i$ of candidate complements for SR_i has a single element $comp_j$
      1. If the list $lPap_j$ of candidate roles for $comp_j$ contains only SR_i
         (i) Assign the semantic role SR_i to $comp_j$ (i.e. label $comp_j$ with SR_i).
   b) Otherwise ($lComp_i$ has more than one element or $lPap_j$ has more than one element)
      1. Determine the syntactic function *fSin* of SR_i in the verbal predicative information.
      2. For each element *comp* of $lComp_i$
         (i) Calculate the co-occurrence measure *cooc* for *fSin* (cooSuj , cooObj ou cooObjInd);
         (ii) Construct the triplet *t* = <*SR, comp, cooc*> and append it to the tuple list *lCoocs.*
2. Order *lCoocs* decreasingly by *cooc.*
3. Work down the list of triplets *t* = <SR_i, $comp_i$, $cooc_i$>, assigning the role SR_i to $comp_i$ (i.e. labeling $comp_i$ with SR_i), adding $comp_i$ to the labeled complement list, marking the role SR_i as assigned, and removing triplets with either SR_i or $comp_i$ from the remaining list.
4. The unlabeled complements that remained in the list of triplets are added to the unlabeled complement list.

Finish.

FIGURE 7 - Algorithm for role verification by preferential syntactic function.

A detailed description of the proposed algorithms can be found in (Mamani Sánchez, 2007) with thorough analysis of particularized cases and exceptions.

4 Experimental Results

With the implementation of the proposed procedure we were able to set up evaluation experiments and empirically observe the correctness of the method. Two sets of experiments were designed; the first set did not use measure 2, in contrast to the second one.

We chose a set of predicators so as to cover the seven mapping rules and to obtain a reasonable amount of corpus occurrences. The repository of argument structures was build manually, using XML as markup language.

As a simple measure of correctness we used

$$Mc = NCorr / NCases,$$

where *NCorr* is the number is the number of predications correctly labeled and *NCases* is the total number of predications processed.

The final results can be classified into three categories, summarized in table 2 below, according to the possible source of the errors: results of type 1 are those correctly labeled using adequate predicate information; type 2 are those incorrectly labeled; type 3 are those correctly labeled but the predicate information was considered inadequate. An instance of result type 3, in Ex. 6, demonstrates the need for improving and, above all, expanding the predicate information repository. Although the labeling procedure agrees with the available frame for nominalized predicate "remoção" (removal), the label CAUSER (AGENT) is not appropriate for "abrasão" (abrasion), which should be identified as an adjunct.

Example 6:

a. remoção das bordas por abrasão
   (removal of the edges by abrasion)
b. Predicative Information: *Noun* de [OBJECT] de [ORIGIN] para [GOAL] por [CAUSER]
c. Labelling: [OBJECT as bordas] [CAUSER abrasão]

TABLE 2 - Evaluation criteria for results.

	Result	Predicative Information Repository
Type 1	Correct	Correct
Type 2	Incorrect	Not considered
Type 3	Correct	Incorrect

Tables 3 to 6 below summarize the results with the information displayed in their columns in the following order.

Column 1: Nominal predicator;
Column 2: Classifiable cases: number of predications with one or more arguments (the sum of columns 3 and 4);
Column 3: Number of errors in the identification phase, where the arguments were badly parsed either by the chunker, or by the prepositional phrase attachment procedure, including cases of undefined attachments;
Column 4: Well formed cases: number of predications with a single argument plus the number of predications correctly formed in the identification phase (the sum of columns 5 and 6, or the sum of columns 7 and 8);
Columns 5 and 6 report the correctness of the method if we consider correct the cases in which the final labels are consistent with available predicative information and the information is correct (type 1). Columns 7 and 8 contain the instances in which the final labels are consistent with available predicative information but the information is incorrect or insufficient (type 3).
Column 5: Number of correct role assignments (type 1);
Column 6: Number of errors in the classification phase (type 1);
Column 7: Number of correct role assignments (type 3);
Column 8: Number of errors in the classification phase (type 3).

The experiments were organized as follows:

- Experiment 1 - Group 1 - results in table 3.
    - Predicators: *perda (loss), herança (inheritance), ganho (gain), saída (exit), descida (descension), deslocamento (dislocation), retirada (retreat), remoção (removal), extirpação (extirpation)* and *extração (extraction)*.
    - Measure 1 for PPAP.
- Experiment 1 - Group 2 - results in table 4.
    - Predicators: *corrida (run), caminhada (walk), padecimento (endurance), sofrimento (suffering), aquisição (acquisition), criação (creation), destruição (destruction), treino (training), sustentação (support), adaptação (adaptation), comparação (comparison), transformação (transformation), envio (postage), doação (donation), solicitação (request), crescimento (growth), degeneração (degeneration), aprendizado (learning), esquecimento (forgetting), conhecimento (knowledge), adoração (adoration), desconhecimento (ignorance), aproximação (aproximation), subida (ascendence)* and *situação (situation)*.
    - Measure 1 for PPAP.
- Experiment 2 - Group 1 - results in table 5.
    - Predicators: *perda (loss), herança (inheritance), ganho (gain), saída (exit), descida (descension), deslocamento (dislocation), retirada (retreat), remoção (removal), extirpação (extirpation)* and *extração (extraction)*.
    - Measure 2 for PPAP.
- Experiment 2 - Group 2 - results in table 6.
    - Predicators: *corrida (run), caminhada (walk), padecimento (endurance), sofrimento (suffering), aquisição (acquisition), criação (creation), destruição (destruction), treino (training), sustentação (support), adaptação (adaptation), comparação (comparison), transformação (transformation), envio (postage), doação (donation), solicitação (request), crescimento (growth), degeneração (degeneration),*

aprendizado (learning), esquecimento (forgetting), conhecimento (knowledge), adoração (adoration), desconhecimento (ignorance), aproximação (aproximation), subida (ascendence) and *situação (situation)*.

- Measure 2 for PPAP.

TABLE 3 - Semantic role labeling results for Experiment 1 – Group 1.

Predicator	Classifiable cases	Errors in Ident.	Well-formed cases	Type 1 correct cases	Type 1 errors Classif.	Type 3 correct cases	Type 3 errors Classif.
Perda	34	6	28	26	2	27	1
Herança	22	2	20	18	2	18	2
Ganho	30	3	27	18	9	22	5
Saída44	3	41	29	12	32	9	
Descida	29	6	23	18	5	21	2
Deslocamento	34	5	29	24	5	24	5
Retirada	41	10	31	24	7	27	4
Remoção	39	2	37	34	3	34	3
Extirpação	5	0	5	5	0	5	0
Extração	39	4	35	32	3	33	2
Totals	317	41	276	228	48	243	33

TABLE 4 - Semantic role labeling results for Experiment 1 – Group 2.

Predicator	Classifiable cases	Errors in Ident.	Well-formed cases	Type 1 correct cases	Type 1 errors Classif.	Type 3 correct cases	Type 3 errors Classif.
Corrida	24	4	20	15	5	20	0
Caminhada	26	4	22	14	8	16	6
Padecimento	2	0	2	2	0	2	0
Sofrimento	21	2	19	15	4	18	1
Aquisição	36	5	31	27	4	29	2
Criação	42	10	32	32	0	32	0
Destruição	34	5	29	29	0	29	0
Treino	29	4	25	17	8	22	3
Sustentação	35	7	28	16	12	27	1
Adaptação	37	6	31	30	1	30	1
Comparação	33	3	30	26	4	28	2

Transformação	25	9	16	10	6	12	4
Envio	47	19	28	24	4	24	4
Doação	30	9	21	13	8	16	5
Solicitação	33	8	25	12	13	25	0
Crescimento	34	4	30	21	9	30	0
Degeneração	28	3	25	22	3	25	0
Aprendizado	21	1	20	14	6	19	1
Esquecimento	18	2	16	12	4	16	0
Conhecimento	15	2	13	12	1	13	0
Adoração	20	3	17	6	11	17	0
Desconhecimento	39	6	33	21	12	33	0
Aproximação	38	6	32	14	18	31	1
Subida40	3	37	27	10	37	0	
Situação	17	3	14	13	1	14	0
Totals	724	128	596	444	152	565	31

TABLE 5 - Semantic role labeling results for Experiment 2 – Group 1.

Predicator	Classifiable cases	Errors in Ident.	Well-formed cases	Type 1 correct cases	Type 1 errors Classif.	Type 3 correct cases	Type 3 errors Classif
Perda	34	8	26	24	2	25	1
Herança	22	2	20	17	3	17	3
Ganho	30	4	26	17	9	20	6
Saída	44	3	41	28	13	31	10
Descida	29	6	23	18	5	21	2
Deslocamento	34	7	27	23	4	23	4
Retirada	41	11	30	25	5	27	3
Remoção	39	3	36	33	3	34	2
Extirpação	5	0	5	5	0	5	0
Extração	39	5	34	31	3	32	2
Totals	317	49	268	221	47	235	33

TABLE 6 - Semantic role labeling results for Experiment 2 – Group 2.

Predicator	Classifiable cases	Errors in Ident.	Well-formed cases	Type 1 correct cases	Type 1 errors Classif.	Type 3 correct cases	Type 3 errors Classif
Corrida	24	3	21	16	5	21	0
Caminhada	26	4	22	15	7	18	4
Padecimento	2	0	2	2	0	2	0
Sofrimento	21	1	20	15	5	18	2
Aquisição	36	5	31	28	3	30	1
Criação	42	8	34	34	0	34	0
Destruição	34	3	31	31	0	31	0
Treino	29	4	25	17	8	23	2
Sustentação	35	7	28	15	13	27	1
Adaptação	37	3	34	34	0	34	0
Comparação	33	1	32	30	2	32	0
Transformação	25	8	17	12	5	14	3
Envio	47	14	33	32	1	33	0
Doação	30	6	24	19	5	23	1
Solicitação	33	7	26	15	11	26	0
Crescimento	34	5	29	21	8	29	0
Degeneração	28	3	25	22	3	25	0
Aprendizado	21	1	20	14	6	19	1
Esquecimento	18	1	17	13	4	17	0
Conhecimento	15	2	13	12	1	13	0
Adoração	20	1	19	8	11	19	0
Desconhecimento	39	3	36	23	13	34	2
Aproximação	38	5	33	13	20	31	2
Subida	40	3	37	26	11	36	1
Situação	17	1	16	14	2	16	0
Totals	724	99	625	481	144	605	20

Table 7 below shows the measure of correctness (*Mc*), in percentage form, according to groups of predicators and correctness types 1 and 3.

TABLE 7 - Correction percentages.

	Group of predicators	Correctness Type	Classifiable cases (%)	Well-formed cases (%)
Experiment 1	Group 1	Type 1	71.92	82.61
		Type 3	76,66	88.04
	Group 2	Type 1	61.33	74.50
		Type 3	78.04	94.80
Experiment 2	Group 1	Type 1	69.72	82.46
		Type 3	74.13	87.69
	Group 2	Type 1	66.44	76.96
		Type 3	83.56	96.80

Correctness type 1 is more restrictive than correctness type 3, hence the lower percentages associated with type 1 in all experiments. The percentage of correctness in the set of classifiable cases reflects the accumulated errors generated in the identification phase and in the classification phase; the percentage in the set of well-formed cases reflects the errors generated in the classification phase only.

Experiments 1 and 2 yielded very similar results in both groups of predicators, which leads us to conclude that the choice between measure 1 or measure 2 for the PPAP decision does not have a significant impact in the overall procedure. It is interesting to note that, even though the identification phase of the procedure is slightly improved with measure 2, as shown by the comparison between the number of well-formed cases for each experiment, the increase in errors in the classification phase outweighs the gain.

5 Concluding Remarks

This work involves a great deal of high level linguistic and computational knowledge, coordinating different technologies ranging from the exploration of automatic lexical resources and corpus, to the construction of efficient algorithms. It starts from a well-founded set of linguistic rules that correspond verbal to nominal predication syntactic-semantic patterns, translates these rules into a computer-based lexicon of nominalizations, and proposes the use of the Web as a corpus to derive the information that is necessary to perform the semantic role labeling of a set of sentences.

The results are considered very positive for such a high level linguistic task as SRL. A comparative analysis with other tools was not possible given that we could not find similar proposals for Portuguese. English-oriented projects rely heavily on powerful resources such as the English Nombank (Meyers et al., 2004) and are, therefore, able to achieve better performance results. Xue (2006) describes the effort to implement SRL of Chinese nominalizations using the Chinese Nombank, and discusses the negative impact of using manually annotated verb data to improve the SRL accuracy of nouns. The differences between Portuguese and Chinese nominalized predications would have to be better understood before we could effectively analyse he author´s findings.

In the analysis of the weaknesses of our procedure, we identified a few points that could be crucial to improve the quality of the results. With respect to the predicative information, it is noticeable that most deverbal noun predications involve only one complement. Also, in the experiments, few predications contained every argument specified in the predicative information derived from the verb. For instance, the role CAUSER can rarely be assigned in deverbal noun predications, in contrast to verbal predications. On average, we observed that deverbal noun predications have fewer arguments than verbal predications. These findings suggest a reformulation of the derivation of predicative information, in the direction of simplifying the nominal frame, thus reducing the chances of ambiguities in the SRL task.

The reduction of ambiguities can also be achieved through the incorporation of a named entity repository, a kind of resource that is already available for Portuguese (Santos and Cardoso, 2007). This type of lexicon associates proper names to types of entities (places, people, institutions, etc), and frequently indicates locative or agentive roles of the complements, thus facilitating role assignment. The annotation of collocations and other forms of composite expressions, such as complex prepositions and fixed expressions would also avoid a great deal of ambiguity in many levels of semantic analysis.

A frequent cause of errors is the lack of morphosyntactic tagging in Web texts. On the other hand, part of speech annotated corpora such as Cetenfolha or MAC-Morpho, are insufficiently large for this kind of approach, as we had the opportunity to attest.

References

BAKER, C. F.; FILLMORE, C. J. and LOWE, J. B. 1998. The Berkeley FrameNet Project. In: PROCEEDINGS OF THE COLING-ACL, volume 1, 86-90.

BASILIO, M. 1980. *Estruturas Lexicais do Português*: uma abordagem gerativa. Petrópolis: Vozes, Coleção Perspectivas Lingüísticas, 1980.

BICK, E. 2000. *The Parsing System "Palavras"*: Automatic Grammatical Analysis of Portuguese in a Constraint Grammar Framework. Ph.D. thesis, Aarhus University.

CAMACHO, R., and SANTANA, L. 2004. A expressão argumental dos nomes deverbais. *Estudos Lingüísticos* XXXIII(3): 214-219.

CARRERAS, X., and MÀRQUEZ, L. 2005. Introduction to the CoNLL-2005 Shared Task: Semantic Role Labeling. In: PROCEEDINGS OF THE NINTH CONFERENCE ON COMPUTATIONAL NATURAL LANGUAGE LEARNING (CoNLL-2005), 152-164.

CRUSE, D. A. 1973. Some Thoughts on Agentivity. *Journal of Linguistics* 9: 1-204.

CRYSTAL, D. 2003. *A Dictionary of Linguistics and Phonetics*. Oxford: Blackwell Publishing.

DOWTY, D. R. 1991. Thematic Proto-Roles and Argument Selection. *Language* 67: 547-619.

FILLMORE, C. 1968. The Case for Case, in *Universals in Linguistic Theory*, E. Bach and R.T. Hams (eds.). New York: Holt, Rinehart and Winston.

FILLMORE C. and B. T. S. ATKINS 1992. Towards a Frame-Based Lexicon: the Case of RISK. In: A. LEHRER and E. KITTAY (eds.). *Frames and Fields.* NJ: Erlbaum Publ.

FOLTRAN M. J. 2003. Relações de Predicação. In: Maria José Foltran, Esmeralda Vailati Negrão, Ana Lúcia Müller (orgs.), *Semântica Formal.* Sao Paulo: Contexto.

GILDEA, D., and JURAFSKY, D. 2002. Automatic Labelling of Semantic Roles. *Computational Linguistics* 28(3): 245-288.

JACKENDOFF, R. 1972. *Semantic Interpretation in Generative Grammar.* Cambridge, MA: The MIT Press.

JACKENDOFF, R. 1983. *Semantics and Cognition*. Cambridge, MA: The MIT Press.

JONHSON, C. and FILLMORE, C. 2000. The FrameNet tagset for frame-semantic and syntactic coding of predicate-argument structure. In: PROCEEDINGS OF THE 1st MEETING OF THE NORTH AMERICAN CHAPTER OF THE ASSOCIATION OF COMPUTATIONAL LINGUISTICS, April 29-May 4, 2000, Seattle WA, pp. 56-62.

LAKOFF, G. 1977. *Linguistic Gestalts*. Chicago Linguistics Society 13: 236-87.

LAPATA, M.. 2002. The Disambiguation of Nominalizations. *Computational Linguistics*. 28(3): 357-388.

LEVIN, B. 1993. (ed.). *English Verb Classes and Alternations*: A Preliminary Investigation. Chicago: The University of Chicago Press.

MAMANI SÁNCHEZ, L. 2007. *Atribuição de Papéis Semânticos a Argumentos de Nominalizações*: um Método Semi-automático. Master's Dissertation. Instituto Militar de Engenharia, Brazil

MARCHI, A. R. 2003. Projeto Lácio-web: Desafíos na construção de um córpus de 1, 1 milhão de palavras de textos jornalísticos em português do Brasil. In: 51º SEMINÁRIO DO GRUPO DE ESTUDOS LINGÜÍSTICOS DO ESTADO DE SÃO PAULO (GEL 2003).

MATEUS, M. H. M.; BRITO, A. M.; DUARTE, I.; and FARIA, I. H. 1989. *Gramática da Língua Portuguesa*. Lisboa: Caminho.

MEYER, R. M. 1991. A *complementação da forma nominalizada de verbal sufixal e a conceituação do complemento nominal.* Doctoral Thesis, Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Letras, Brazil.

MEYERS, A.; REEVES, R.; MACLEOD, C.; SZEKELY, R.; ZIELINSKA, V.; YOUNG, B. and GRISHMAN, R. 2004. The Nom-Bank Project: An Interim Report. In: PROCEEDINGS OF THE NAACL/HLT WORKSHOP ON FRONTIERS IN CORPUS ANNOTATION, Boston.

OLIVEIRA, C., and FREITAS, M. C. 2006. Classes de palavras e etiquetagem na Lingüística Computacional. *Calidoscópio*. 4(3).

PALMER M., KINGSBURY P. and GILDEA D. *2005.* The Proposition Bank: An Annotated Corpus of Semantic Roles. *Computational Linguistics 31 (1): 71-106.*

PERES, J. 1984. *Elementos para uma gramática nova*. Almeidina: Coimbra.

POUNDER, A. 2000. *Processes and Paradigms in Word-formation Morphology.* Berlin: Walter de Gruyter.

SAID ALI, M. 1966. *Gramática Elementar da Língua Portuguesa*. 9a. ed. São Paulo: Melhoramentos.

SANDMANN, A. J. 1989. *Formação de Palavras no Português Brasileiro Contemporâneo*. 2a. ed. Curitiba: Ed. UFPR.

SANTOS, D. and SARMENTO, L. 2003. O projecto AC/DC: acesso a corpora /disponibilização de corpora. In: ACTAS DO XVIII ENCONTRO DA ASSOCIAÇÃO PORTUGUESA DE LINGUÍSTICA.

SANTOS, D. and CARDOSO, N. 2007. *Reconhecimento de entidades mencionadas em português*: Documentação e actas do HAREM, a primeira avaliação conjunta na área. Lisboa: Linguateca.

SWIER, R. S., and STEVENSON, S. 2004. Unsupervised Semantic Role Labelling. In: LIN, D., E WU, D., (eds.) In: PROCEEDINGS OF EMNLP 2004, 95-102. Barcelona, Spain: Association for Computational Linguistics.

VAN VALIN, R. D., Jr. 1990. Semantic Parameters of Split Transitivity. *Language* 66: 221-60.

VAN VALIN R. D., Jr. 2004. Semantic macroroles in Role and Reference Grammar. Rolf Kailuweit and Martin Hummel (eds.) *Semantische Rollen*, 62-82. Tübingen: Narr,

VOLK, M. 2001. Exploiting the WWW as a corpus to resolve PP attachment ambiguities. In: PROCEEDINGS OF CORPUS LINGUISTICS 2001, Lancaster, UK.

Xue, N. 2006. Semantic role labeling of nominalized predicates in Chinese. In: PROCEEDINGS OF THE MAIN CONFERENCE ON HUMAN LANGUAGE TECHNOLOGY CONFERENCE OF THE NORTH AMERICAN CHAPTER OF THE ASSOCIATION OF COMPUTATIONAL LINGUISTICS, 431-438.