Automatic Semantic Role Labeling for European Portuguese

Rui Pedro Talhadas dos Santos

Dissertação
Mestrado em Ciências da Linguagem

Trabalho efetuado sob a orientação de:
Prof. Doutor Jorge Manuel Evangelista Baptista (U.Algarve / FCHS)
Prof. Doutor Nuno João Neves Mamede (U. Lisboa / IST)

Faro, 2014

Automatic Semantic Role Labeling for European Portuguese

Declaração de autoria do trabalho

Declaro ser o(a) autor(a) deste trabalho, que é original e inédito. Autores e trabalhos consultados estão devidamente citados no texto e constam da listagem de referências incluída.

Acknowledgements

First of all, I can not thank my supervisor, Prof. Jorge Baptista, enough for his tireless support and help throughout all my years in the university and specially during this dissertation. Without his expertise and support this dissertation would not have been possible.

Secondly, I would like to thank my co-supervisor, Prof. Nuno Mamede, for his precious support in the course of my dissertation.

I have to thank my mother for all the unbelievable efforts she has done so that I could have the opportunity to achieve all I have achieved in my life, most times putting me in front of her needs.

Nonetheless, I have to thank Marie Bordone for always being there and for cheering me up in the hardest times.

Last but not the least, I have to thank my closest friends for all the the support they have given me.

Faro, April 23^{th} 2014
Rui Talhadas dos Santos

Resumo

Esta tese aborda a tarefa de Anotação de Papéis Semânticos (APS) em Português Europeu. A APS pode ser usada em diversas aplicações de Processamento de Linguagem Natural (PLN) tais como, Resolução de Anáforas, Recuperação/Extração de Informação, Sumarização Automática, etc. Um conjunto de 37 papéis semânticos, consensual e de uso geral, foi definido com base nos trabalhos relacionados relevantes e recorrendo a propriedades suficientemente reprodutíveis.

Foi também elaborado um conjunto de diretrizes de anotação, a fim de esclarecer como deveriam ser atribuídos aos argumentos verbais, em contexto, os seus respetivos papéis semânticos.

Com base nestes elementos, foi construído um módulo de APS, que se encontra integrado na cadeia de Processamento de Linguagem Natural STRING, desenvolvida no INESC-ID Lisboa.

Para este módulo, foram utilizadas as informações de um banco de dados léxico-sintáticos, ViPEr, que contém a informação linguística relevante para mais de 6.000 verbos plenos (ou lexicais, ou distribucionais) do Português Europeu, e a base de dados foi enriquecida manualmente com as informações referentes ao papéis semânticos de todos os argumentos verbais (sujeito e complementos essenciais).

O módulo de APS é composto por 183 regras de correspondência de padrões para a marcação de sujeito (N0), primeiro (N1) e segundo (N2) complementos essenciais das construções verbais, e também permite a atribuição de papéis semânticos para outros constituintes sintáticos, adjuntos adverbiais, tais como os complementos de tempo, de modo, os complementos locativos, instrumentais, comitativos, entre outros (tanto essenciais como circunstanciais).

Este módulo foi testado num corpus de textos reais, de natureza tipológica variada e abordando diversos tópicos, o qual foi manualmente anotado por dois linguistas especificamente para este propósito. Após esse processo de anotação manual, o corpus, que contém 655 papéis semânticos, foi usado como um corpus de referência (*golden standard*) para a comparação automática com a saída do sistema.

Considerando-se que o módulo de APS opera nos últimos passos da cadeia de processamento, foi alcançada uma precisão relativamente alta (69,9 % em uma avaliação estrita e 77,7 %, quando a avaliação inclui resultados parciais), embora a abrangência (ou *recall*) tenha sido baixa (17,9 %), o que deverá constituir um dos objetivos do trabalho futuro.

Abstract

This thesis addresses the task of Semantic Role Labeling (SRL) in European Portuguese. SRL can be used in a number of NLP application, namely Anaphora Resolution, Question Answering, Summarization, etc.

A general-purpose, consensual set of 37 semantic roles was defined, based on a survey of the relevant related work, and using highly reproducible properties. A set of annotation guidelines was also built, in order to clarify how semantic roles should be assigned to verbal arguments in context.

A SRL module was built and integrated in a fully-fledged Natural Language Processing (NLP) chain, named STRING, developed at INESC-ID Lisboa.

For this module, the information from a lexicon-syntactic database, ViPEr, which contains the relevant linguistic information for more than 6,000 European Portuguese full (or lexical, or distributional) verbs, was used and the database manually enriched with the information pertaining to the semantic roles of all verbal arguments.

The SRL module is composed of 183 pattern-matching rules for labeling of subject (N0), first (N1) and second (N2) essential complements of verbal constructions and also allows the attribution of SR to other syntactic slots in the case of time, locative, manner, instrumental, comitative and other complements (both essential and circumstantial).

This module was tested in a small corpus that was specifically annotated for this purpose. After this manual annotation, the corpus containing 655 semantic roles was used as a golden standard for automatic comparison with the system's output.

Considering that the SRL module operates at the last stages of the processing chain, a relatively high precision was achieved (69.9% in a strict evaluation and 77.7%, when evaluation included partial matches), though the recall was low (17.9%), which calls for future improvements.

Resumo Alargado

Esta dissertação aborda a tarefa de Anotação de Papéis Semânticos (APS) em Português Europeu. Os papéis semânticos correspondem à relação semântica estabelecida entre um verbo (ou outro elemento predicativo) e os seus respetivos argumentos e complementos, correspondendo, *grosso modo*, à noção de *lead*: "Quem faz o Quê, a Quem, Como, Quando e Onde". A APS pode ser usada em diversas aplicações de Processamento de Linguagem Natural (PLN), tais como Resolução de Anáforas, Recuperação/Extração de Informação, Sumarização Automática, etc.

Os objectivos principais deste trabalho consistem em: (i) a definição clara e precisa de um conjunto de papéis semânticos (PS) com um certo grau de generalidade e elevada reprodutibilidade a fim de poderem ser utilizados em diferentes tarefas e aplicações de PLN; e (ii) a construção de um módulo de anotação automática de papéis semânticos para o Português Europeu, a ser integrado numa cadeia de Processamento de Linguagem Natural, a STRING [1].

Quanto ao primeiro objetivo, foram analisados os principais trabalhos relacionados com os papéis semânticos, em particular mais utilizados na literatura, e foi constituída uma lista de todos os SR encontrados e suas respetivas definições. Estes foram analisados de forma crítica, a fim de construir uma proposta aplicável aos dados das construções verbais o Português Europeu. No total, a proposta apresenta 37 papéis semânticos, incluindo alguns subtipos (e.g. vários subtipos de `AGENTE`, em função do tipo de predicado), os quais constituem o conjunto utilizado neste estudo. Quando agrupados nos principais papéis, abstraindo dos diferentes subtipos, totalizam 19 papéis semânticos distintos.

A par deste processo, foi elaborado um conjunto de diretrizes de anotação a fim de esclarecer os anotadores de como deveriam ser atribuídos aos argumentos verbais, em contexto, os seus respetivos papéis semânticos. Estas diretrizes serviram igualmente para orientar o processo de anotação do corpus de avaliação do módulo de anotação automática.

A fim de validar o conjunto de papéis semânticos assim constituído, foi pedido a um conjunto de 4 anotadores (três linguistas e um informático com experiência em PLN) que marcasse os papéis semânticos num corpus de frases reais. Foi fornecido aos anotadores o conjunto de diretrizes acima referido. O corpus a anotar era constituído por 68 frases selecionadas aleatoriamente do corpus de avaliação (ver adiante) e apresentava um conjunto de 113 papéis semânticos, que os anotadores deveriam assinalar. O processo de anotação baseou-se no preenchimento de um formulário *on-line*. Neste formulário, as frases eram apresentadas uma a uma, destacando-se o verbo e o respetivo argumento/complemento, para os quais era pedido que se assinalasse a respetiva relação semântica. O anotador podia escolher qualquer um dos 37 papéis semânticos do conjunto previamente definido ou, em alternativa, indicar "nenhum dos anteriores". No total, foi alcançado um acordo entre anotadores (Cohen k) de 52 %, considerando todos os subtipos distintos. Quando os subtipos são agregados nos principais papéis semânticos correspondentes, este valor sobe nove pontos percentuais, atingindo-se 61 % de concordância entre anotadores. Estes resultados indicam que a tarefa de anotação de papéis semânticos é bastante exigente, mesmo para anotadores

[1] https://string.l2f.inesc-id.pt/

humanos.

Foram ainda revistos vários trabalhos sobre a Anotação de Papéis Semânticos já realizados para o Português. A maioria desses trabalhos adota uma abordagem da tarefa de anotação baseada em métodos aprendizagem automática. Apenas um dos trabalhos encontrados adoptava uma metodologia baseada em regras. Os resultados de avaliação reportados não podem ser diretamente comparados com os deste estudo, na medida em que os *corpora* utilizados são de dimensões e natureza textual muito diferente. Contudo, importa referir que alguns destes trabalho recorrem a texto simplificado, o que facilita a tarefa de anotação, ou se limitam a texto de tipo jornalístico.

Com base no conjunto de papéis semânticos por nós definido, foi enriquecida uma base de dados léxico-sintáticos, constituída pelos verbos plenos do Português Europeu, o ViPEr (Baptista 2012), acrescentando-lhe a informação relativa aos papéis semânticos de todos argumentos (sujeito e complementos essenciais) das diferentes construções verbais. Este recurso linguístico contém informação sintática (distribucional, estrutural e transformacional) relevante para mais de 6.300 construções de verbais do Português Europeu. Um total de 13.201 papéis semânticos foram manualmente anotados no ViPEr.

Com base neste recurso, foi então construído o módulo de Anotação de Papéis Semânticos, que se encontra atualmente integrado na cadeia de Processamento de Linguagem Natural STRING, desenvolvida no INESC-ID Lisboa. A estratégia adotada para a APS enquadra-se na cadeia de processamento STRING, prevendo que, antes desse módulo entrar em ação, tenham já tido lugar as tarefas de anotação e desambiguação morfossintática (*POS tagging*); a análise sintática (*shallow parsing*) dos constituintes elementares (*chunking*): identificação de grupos nominais (`NP`), preposicionais (`PP`), etc.; a extração das dependências sintáticas fundamentais (*deep parsing*): sujeito, complemento direto, modificador, etc.; e, finalmente, a desambiguação automática das construções verbais (Travanca 2013). O módulo de APS associa, então, a informação do ViPEr à construção verbal previamente identificada, nomeadamente o respetivo papel semântico dos vários argumentos (sujeito e complementos essenciais) desse emprego do verbo. Assim, por exemplo, se o sistema atribuiu a uma dada instância do verbo *contar*, de entre as várias construções representadas no ViPEr, a que corresponde ao sentido de "*O Pedro conta com o João*", é então atribuída a essa instância os traços correspondentes aos papéis semânticos do sujeito e do complemento tal como representados na base de dados. Esses traços são depois convertidos nos papéis semânticos associados aos argumentos do predicado, durante a fase de extração de eventos (os predicados da frase) e respetivos participantes (os argumentos desses predicados).

O módulo de APS é composto por 183 regras de emparelhamento de padrões para a marcação de sujeito (`N0`), primeiro (`N1`), segundo (`N2`) e terceiro (`N3`) complementos essenciais das construções verbais. Este conjunto de regras inclui ainda um conjunto de heurísticas que também permite a atribuição de papéis semânticos para outros constituintes sintáticos (tanto essenciais como circunstanciais), em especial para os adjuntos adverbiais, tais como os complementos de tempo, de modo, os complementos locativos, instrumentais, comitativos, entre outros. O módulo permite ainda, com base na sua identificação prévia, a atribuição de papéis semânticos às construções passivas, tanto com o verbo auxiliar *ser* como com o verbo *estar*.

O módulo de APS foi testado num corpus de textos reais, de natureza tipológica variada e abordando diversos tópicos. No total, o corpus é composto por 373 frases, com 7,834 palavras. Trata-se, tanto quanto sabemos, de um dos maiores corpus manualmente anotado com papéis semânticos existente para o Português.

Para fazer a avaliação do módulo de APS, o corpus foi conjuntamente anotado por dois linguistas. Após esse processo de anotação manual, o corpus, que apresenta 655 papéis semânticos, foi usado como um corpus de referência (*golden standard*) para comparação automática com a saída do sistema.

Considerando-se que o módulo opera nos últimos passos da cadeia de processamento, foi alcançada uma precisão relativamente alta (68.4 %, numa avaliação estrita, e 76.4 %, quando a avaliação inclui

resultados parcialmente corretos), embora a abrangência (ou *recall*) tenha sido baixa (19.8 %), atingindo-se uma medida-F de 30.1 % (avaliação estrita) a 31.5 % (na avaliação relaxada).

Estes resultados podem estar associados a diversos fatores, em particular ao facto de o módulo ter sido integrado nas fases finais da cadeia de processamento de linguagem natural, STRING, após as tarefas de *POS tagging*, *parsing*, e de desambiguação verbal, o que implica que os erros cometidos nas fases anteriores da cadeia se reflitam nos resultados, em particular numa menor abrangência. Uma outra razão para os resultados obtidos vem do facto de o corpus ser extremamente variado do ponto de vista textual e de o mesmo não ter sido simplificado, ao contrário do que foi observado na revisão de outros trabalhos relacionados.

Como trabalho futuro parece ser importante garantir uma maior abrangência (*recall*), o que poderá implicar o desenvolvimento de heurísticas mais flexíveis do que as utilizadas atualmente. Por outro lado, é desejável a extensão da APS a outras categorias morfossintáticas, em particular os adjetivos e os nomes predicativos.

Palavras-Chave
Keywords

Palavras-Chave

Anotação de Papéis Semânticos
Processamento de Língua Natural
Léxico-gramática
Regras de emparelhamento de padrões
Português Europeu

Keywords

Semantic Role Labeling
Natural Language Processing
Lexicon-grammar
Pattern-matching rules
European Portuguese

Contents

List of Figures

List of Tables

List of Acronyms

Acronym	Designation in English	Designation in Portuguese
CDIR	**D**irect **C**complement	Complemento direto
CINDIR	**I**ndirect **C**complement	Complemento indireto
HMM(s)	**H**idden **M**arkov **M**odel(s)	Modelo(s) oculto(s) de Markov
L^{2}F	Spoken Language Systems Laboratory	**L**aboratório de Sistemas de **L**íngua **F**alada
MARv	**M**orphosyntactic **A**mbiguity **R**esol**v**er	Desambiguador Morfossintático
MOD	**M**odifier	Modificador
NE(s)	**N**amed **E**ntity(ies)	Entidade(s) Mencionada(s)
NER	**N**amed **E**ntities **R**ecognition	Reconhecimento de Entidades Mencionadas
NLP	**N**atural **L**anguage **P**rocessing	Processamento de Língua Natural
NP	**N**oun **P**hrase	Sintagma nominal
POS	**P**art **o**f **S**peech	Categoria gramatical ou (morfos)sintática
PP	**P**repositional **P**hrase	Sintagma preposicional
QA	**Q**uestion **A**nswering	Sistemas de Pergunta e Resposta
RuDriCo	**Ru**le-**Dri**ven **Co**nverter	Conversor baseado em regras de transformação declarativas
SR	**S**emantic **R**ole	Papel Semântico
XIP	**X**erox **I**ncremental **P**arser	

Chapter 1

Introduction

Semantic roles (SR) (or *theta* or *thematic roles*) correspond, grossly, to the well-known notion of lead: "Who did What to Whom, How, When and Where" [56]. The concept of SR, though not exactly the term, is already present in linguistics, associated with the discussion on the semantic or syntactic values of *Case* and the development of Case Theory [20], particular with the work of Fillmore (1968) [29] and the notion of *Deep Case*, and later derived in Linking Theory [44]. A general understanding of SR underlies Zellig Harris' (1991) [42] definition of the fourth basic restriction, called *linearization*. This constraint is responsible, not only for the sequential arrangement of words (or morphemes) in the linear order of discourse (or in the sentence), but also for the semantic relations those words (or morphemes) hold between them and the operator they depend on. The specific linguistic devices used by languages to express this constraint can vary (case, word-order, and pre- or postpositions). This diversity of linguistic devices and the correspondence between pre- and postpositions, word order and cases, as mechanisms to express the same notions across languages (or across language stages) is reported from the early 19th century ([16], apud [20] (p.59)), though the different systems (adpositions and cases) may coexist in the same language (Latin being an obvious case).

In languages such as Portuguese, whose order of constituents in the sentence is relatively stable, there is a great regularity between the function and the position of syntactic constituents, on the one hand, and the semantic role they play in relation to the operator on which they depend, on the other hand. Thus, for example, the subject of a verb is often the `AGENT` of the process, while the direct complement is, in most cases, its `OBJECT`[1]:

(1) *O Pedro/subject-`AGENT-GENERIC` moldou o barro/direct complement-`OBJECT-GENERIC`*
'Peter shaped the clay'

The semantic role is often directly related to the syntactic function that the constituent plays in the sentence. However, it is not always possible to predict, from its syntactic function, the semantic role a constituent plays. For example, in the next sentence, the direct complement has a locative interpretation:

(2) *O Pedro/subject-`AGENT-GENERIC` atravessou a sala/dir. compl.-`LOCATIVE-PLACE`*
'Peter crossed the room'

Moreover, certain transformations modify sentences, changing the arrangement of its constituents

[1] A more complete definition of these roles will be provided in 3. Here, the reader can take them in its informal sense.

relatively to the core of the predication, without, however, changing their respective thematic roles. This is the case of the active-passive pair in verbal constructions:

(3) *O Pedro/subj-AGENT-GENERIC já leu o texto da Ana/dir. compl.-OBJECT-GENERIC*
'Peter read Ana's text'

(4) *O texto da Ana/subj-OBJECT-GENERIC já foi lido pelo Pedro/comp.-AGENT-GENERIC*
'Ana's text was read by Peter'

or in the standard-converse pair, in nominal predicate constructions with support verb [34]

(5) *O João/subj.-AGENT-GENERIC deu uma leitura rápida ao texto da Ana/ind. compl.-OBJECT-GENERIC*
'John gave a quick reading to Ana's text'

(6) *O texto da Ana/subj.-OBJECT-GENERIC levou uma leitura rápida por parte do Pedro/compl.-AGENT-GENERIC*
'Ana's text has had a quick reading by Peter'

The semantic roles, also known as thematic or theta-roles, can, thus, be defined as the semantic relations between the predicate and its arguments. This semantic level overlaps the syntactic level, captured by the parsing NLP task, that is, the analysis of the sentence and the identification of its immediate constituents (shallow parsing), as well as the extraction of the syntactic dependencies they hold with the main verb and between them (deep parsing). In other words, for numerous NLP applications an adequate representation of the meaning of the sentence may just require the mere identification of the sentence's constituents and their syntactic dependencies. However, for more complex processes, a deeper semantic analysis may be useful or even necessary. In the context of natural language processing, the task of syntactic parsing and semantic analysis are normally performed sequentially.

1.1 Motivation

The annotation of semantic roles can be used in several NLP applications of natural language processing, for example in Question Answering (QA) [62]. In QA systems, it is intended that the relevant and correct information be found in texts as the answer to questions formulated in natural language. Imagine, for example, the question:

(7) *Onde/LOCATIVE-PLACE está sepultada D. Inês de Castro/PATIENT*
'Where is D. Inês de Castro buried?'

With this information, the system can search the available corpus, trying to find phrases where *D. Inês de Castro* plays the role of `PATIENT` of verbs such as *sepultar* (to bury), and that some element is associated to the semantic role of `LOCATIVE-PLACE`. Thus, for example, by analyzing a sentence like:

(8) *D. Inês de Castro/PATIENT foi sepultada no Mosteiro de Alcobaça/LOCATIVE-PLACE*
'D. Inês de Castro was buried on the Monastery of Alcobaça'

the QA system would associate the interrogative pronoun constituent *onde* (where) in the question to the named entity *Mosteiro de Alcobaça*, as they both play the same role for the same, or a similar, predicate.

1.2 Objectives

The goal of this research project is to build a rule based module that performs automatic Semantic Role Labeling. This module is to be integrated on an existing NLP system, named STRING [2] (Mamede *et al.*, 2012 [45]). The module will consist of a set of rules for SRL for the European Portuguese full (or *lexical* or *distributional*) verbs. The SRs will be extracted using the set of rules implemented on the Xerox Incremental Parser, XIP[3], the parsing module of STRING.

STRING is an NLP chain with a modular structure that executes all the basic processing tasks, such as: tokenization, part-of-speech tagging, morphosyntactic disambiguation, chunking (shallow parsing) of elementary syntactic constituents and syntactic analysis. The SRL task is a preliminary semantic processing of texts, as it tries to capture the semantic relations between predicates and their arguments. STRING is already able to extract some semantic properties: It performs Named Entity Recognition (person, organization, place, time expressions) [52], event and relation extraction and it uses a lexicon that has been partially encoded with semantic properties (based on Bick, 2006 [18]).

The objectives on this work are manyfold: First, we intend to present an overview of extant semantic roles' definitions and compare them in order to define in a precise and clear way a set of SR that can be used in a reproducible way for corpus annotation and NLP. Secondly, a corpus of significant size will be manually annotated, by two linguists, and will be used to evaluate the SRL module. In order to validate the set of SR here defined, a smaller portion of that corpus will be annotated by several people, and the inter-annotator agreement measured. Finally, a SRL rule-based module will be developed and integrated in a fully-functional NLP chain, STRING, and the results are to be evaluated and discussed.

1.3 Structure

In the next chapter, the state-of-the-art of the research in semantic role labeling is briefly reviewed, with special focus on Portuguese related work. In Chapter 3, an overview of extant definitions of semantic roles is provided and discussed, in view of defining the relevant set of SR used in this project. In Chapter 4, STRING's architecture is explained and the semantic role module is presented. Chapter 5 presents the evaluation methods and the final results, along with the validation of the SR set, and Chapter 6 presents the conclusions and future work.

[2] https://string.l2f.inesc-id.pt/
[3] http://open.xerox.com/Services/XIPParser

Chapter 2

Related Work

In this chapter, we briefly present the main sources used for a broad review of definitions of semantic roles (Subsection 2.1) and then focus on related work done in Portuguese SRL 2.2.

Though the general concept of semantic role is relatively clear, it is much more difficult to define precisely the requirements for a given constituent to be tagged as performing a certain role in relation to the operator (verb) it depends on in the sentence. In fact, there is abundant discussion, in the vast literature already available, about the number of SR and the specifics on the definition of each SR (Allen, 1994 [4]; Gildea and Jurafsky, 2002 [33]; Palmer *et al.*, 2010 [56]; Ruppenhofer, 2010 [59]; Oliveira, 2010 [53]; Wikipedia[1], s/v). Before any SR labeling is undertaken, it is necessary to clearly define which semantic roles are to be used by a system and to define them in an unambiguous, clearly reproducible way.

2.1 Main sources for a broad review of semantic roles definitions

As concepts and their wording are intrinsically connected, and in order to provide a operative definition, an extended set of SR and their definitions were collected and compared, basically from the following sources:

- Fillmore (1968) [29], robably being the first person to propose the notion of semantic roles as an extension of the grammatical concept of *case*;
- Dowty (1991) [27], who posit the concept of 'proto-roles' who give rise to specific semantic roles by modulating different semantic primitives (dynamicity, volition, etc);
- Allen (1994) [4], as one of the first authors that motioned NLP towards the automatic semantic analysis;
- Gildea and Jurafsky (2002) [33], who present a list of semantic roles, though they have not provided any definitions and the examples there given are relatively poor;
- Palmer *et al.* (2005) [55], introducing an annotated corpus with semantic roles, the PropBank corpus;

[1] http://en.wikipedia.org/wiki/Thematic_relation

- Bick (2007) [19], where a set of semantic prototypes was presented as the basis for a rule-based, semantic role labeling system for Portuguese in a constraint grammar framework;
- Palmer *et al.* (2010) [56], a recent state-of-the-art review of the field of SRL.
- These definitions were also compared with those presented in the FrameNet[2] perspective (Ruppenhofer *et al.*, 2010)[59];
- The definitions of PAPEL (Oliveira *et al.*, 2010 [53]) were also consulted, being a Portuguese linguistic resource that presents several semantic relations.
- Finally, the extensive and well-sourced article of Wikipedia was checked regarding the same terminological issues of naming and defining SRs.

In Chapter 3 the definitions of SR (or other relevant semantic relations among sentence constituents) will be reviewed and compared, in order to achieve an operative set of semantic roles. In the following section we focus on the related work for Portuguese SRL.

2.2 Semantic Role Labeling in Portuguese

Semantic role labeling in Portuguese is a relatively new topic of research. The following works were considered relevant for this dissertation:

Bick (2007) [19], presented an automatic system for semantic role labeling, based on his previous work on semantic prototypes (Bick, 2006 [18]). On this work, he used 35 different semantic roles, such as `§AG` (for agent), `§PAT` (for patient) or `§LOC` (for location), and developed a constraint grammar of about 500 mapping rules and a small number of disambiguation rules. The semantic roles are marked on the head of the argument. For example, the following rule selects the agent role (`§AG`) to a subject of a speech verb, if this is not in the passive construction..

```
SELECT (§AG) (0 @SUBJ) (p V-SPEAK LINK NOT 0 PAS) ;
```

These rules are only applied after a preprocessing stage consisting of dependency-based rules, not presented by the author. This preprocessing handles, among others, the semantic prototype labels, namely the inheritance from nouns to pronouns and relative clauses.

Bick applied the mapping rules on input from the European Portuguese part of the Floresta Sintá(c)tica treebank[3] (Afonso *et al.* (2002))[1]. A section of 2,500 words was used for testing. Before the semantic role labeling, different PALAVRAS modules were used, such as dependency conversion, semantic prototype tagging and named entity classification. The automatic annotator assigned 884 semantic role labels in total, of which 84 were wrong, having obtained a recall of 86.8%, a precision of 90.5%, and an F-score of 88.6%.

Bick presents a very interesting and useful list of semantic roles, which were considered to the building of the set of semantic roles used in this dissertation. Bick uses manually developed rules to annotate the verbal arguments on the corpus. In this dissertation, the labels are also applied to the head of the NP.

[2] https://framenet.icsi.berkeley.edu/fndrupal/
[3] http://www.linguateca.pt/floresta/

Amancio *et al.* (2011) [6], presents a technique which ultimate goal is to build an automatic question categorization system that assigns *wh*-question labels to verbal arguments in a sentence. *wh*-questions labels are tags that represent the semantic relation between the verb and its arguments.

The corpus was composed of 104 general news articles, manually simplified in the PorSimples project[4] (Caselli *et al.* (2009) [22]). The corpus was then syntactically annotated by the parser PALAVRAS [17], with no revision. The 9,820 verbal arguments of the corpus were then labelled with the question labels, using the SALTO annotation tool (Burchardt *et al.* (2006) [21]). This tool was originally developed for the annotation of semantic roles and semantic classes in the FrameNet paradigm, and can be used for other tasks such as annotation of discourse structure or anaphoric relations.

The verbal arguments were marked with 68 different tags like the following: "Quem-direita" (Who-Right) and "Quem-esquerda" (Who-Left), "O quê" (What), "Qual" (Which) and "Quais" (Which [plural]). Some tags include the flags *left* or *right*, and that is explained by the syntactical position of the argument in relation to its predicate, corresponding, respectively, to the subject and to the indirect object, preventing ambiguity on question answering. For example, "Quem-direita" (Who-Right) and "Quem-esquerda" (Who-Left) tags prevent the ambiguity between the subject and an direct complement: *[O Pedro]/Quem-esquerda adora [a Ana]/Quem-direita.*

This technique has no labels to allow the system to distinguish between animate/ inanimate nouns. Because of this, they had to unify some labels, such as "O quê" (What) and "Quem" (Who), into "quê_quem" (what_who). After this unification, they have ended up with 57 semantic tags. Notice that these set includes tags involving the prepositions introducing the arguments: "There is a lot of labels because in Portuguese the preposition that introduces indirect object is moved to the left of the *wh*-question. Questions answered by adverbials are: "Onde?" "Quando?" "Com que frequência?" "Por quanto tempo?" "Quanto?" "Por quê?" "Como?" "Para quê?" "Em que direção?" and combinations of prepositions with the *wh*- questions "onde", "quando" and "quanto" (Por onde?, De onde?, De quando?, A quanto?, etc.)." (idem [6] p.46)

The evaluation was conducted on Weka (Witten *et al.* (2009) [41]) using six machine-learning algorithms. From the six used, SMO, SimpleLogistic (Maximum Entropy) and J48 had the best results with F-measure scores of 0.79, 0.78 and 0.74, respectively.

Though Amancio's work presents an interesting approach on semantic role labeling, it will not be considered in this work because our work focuses on a rule-based approach, rather than a strictly automatic annotation of the corpus that uses only the syntactic information to label the verbal arguments. The labels chosen by the author are easy for human annotation because they are very intuitive. This means the margin of error is small, but it also means that not much information can be retrieved from the label, in respect to the SR of the argument. An example of this is the verb *influenciar* (to influence), where both N0 and N1 positions can be filled by "O quê" (What) or "Quem" (Who), , depending on distributional features of the corresponding head nouns. However, it does not allow for the attribution of the CAUSE role to the subject (or the potential non-agentive reading of a human subject); secondly, depending on the definitions of SR used, the human/non-human distributional feature of the N1 head noun can yield the same or different SRs, though it is apparent that the semantic interpretation of that verb argument position is the same, irrespective of the lexical fulfillment of that syntactic slot. Still, this is a good approach to the SRL task: On the one hand, the corpus used is quite large and the labels are very reproducible. On the other hand, the corpus used is composed of simplified sentences, which clearly eases the SRL task in a significant way. Some labels are not generic enough, such as the case of locatives, and there are even cases where some labels are ambiguous, such as "Como?", where one can have a manner SR ("efficiently"), an instrument SR ("with a knife"), or even a time SR ("frequently").

Manchego *et al.* (2012) [5], present a methodology for a machine learning semi-supervised approach

[4] http://nilc.icmc.usp.br/porsimples/index.php/Tools

for Semantic Role Labeling of Brazilian Portuguese. This paper is only a methodological proposal, and no actual work was done. The proposed architecture included three phases: pruning, argument identification and argument classification. Manchego *et al.* planed on exploring the use of three sets of semantic roles: first, the one introduced by Gildea and Jurafsky [33], used in most SRL systems; the second by Amancio *et al.*[6], because it proved to be useful in the automatic labeling of questions for Brazilian Portuguese; and the third one by Morante *et al.* (2007) [50], who created a SRL system for Spanish; this SR set was considered, according to the authors, because it should have a stronger relation with Portuguese than with English. The corpus they plan on using is the PropBank.Br, which has 1,068 target verbs.

The system is to be evaluated in the tasks of argument identification, argument classification and combination of identification and classification (the system must classify the nodes as being a specific argument or as not being an argument of a verb). This methodology uses only machine learning approach, and so, it will not be considered on this dissertation.

Sequeira *et al.* (2012) [61], presents a preliminary machine learning approach for obtaining a semantic role labeler for Portuguese, using Arg0 and Arg1 labels, which correspond, grossly, to the verb's subject and complement. This paper presents not only the first version of a functional semantic role labeler, but also an important element for SRL for Portuguese, a classified corpus, BosqueUE. It is a subset of Bosque 8.0[5], composed of CETEMPublico's[6] sentences that ended with a punctuation mark. In this way, 4,416 sentences were selected to integrate the corpus.

Besides marking the semantic role of each sentence's chunk, the corpus was also marked with morphologic, syntactic and other kind of features. BosqueUE was built using the same format as the corpus used in CONLL'2004[7], and contained 7 features for each word: word, lemma, part-of-speech tag, chunks with IOB, semantic roles with IOB, named entities with IOB [8] and clauses. The only semantic roles here used were Arg0 (=Subject) and Arg1(=Object), which widens the chance to correctly tag the SR, compared to this dissertation's larger set of semantic roles. The predicate, mostly verbs, was also targeted.

The annotated corpus was tested using several MinorThird's algorithms, and the best results were obtained with SVMCMM (Conditional Markov Models trained with Support Vector Machines) and CRF (Conditional Random Fields).

Apparently, the authors do not clearly distinguish the concepts of syntactic function from the semantic role. Therefore, the problem was, in our opinion, over-simplified as not all subjects nor all direct complements play the same semantic role regarding the verb they depend on. Furthermore, other (essential) complements may be involved in a semantic predicate, besides the direct object, the obvious case being the addressee (in a communication predicate) or the recipient (in an object-transference predicate), among many others. Alternatively, the authors may rely on a previous verb-per-verb description of the semantic roles associated to Arg0 and Arg1 of each verb, but no mention of such resource is made in the paper.

Both corpora used, the BosqueUE corpus and the one from CONLL'2004, were evaluated, and the results were better in the latter one. On BosqueUE, the best results were 0.549, 0.306 and 0.171 for the predicate, Subject and Object, respectively. On CONLL'2004, the best results were 0.836, 0.523 and 0.234 for the predicate, Subject and Object, respectively, using the CRF and SVMCMM algorithms. This difference can be explained by the variance on the size of the datasets used.

Considering this work is only a preliminary machine learning approach to a semantic role labeler, it presented promising results.

[5]http://www.linguateca.pt/floresta/corpus.html
[6]http://www.linguateca.pt/cetempublico/
[7]http://www.lsi.upc.edu/ srlconll/st04/st04.html
[8]IOB tags mark each token as B= begin, I= inside and O= other, so that the boundaries of the overlapping layers, such as the semantic roles, can be identified.

Fonseca *et al.* (2012) [30], presented an adaptation of the system SENNA (Collobert *et al.* (2011) [23]). SENNA is an NLP tool that was built for the English language, and can perform part-of-speech tagging, chunking, named entity recognition and semantic role labeling.

The semantic roles tags used by SENNA are the ones used in CONLL, and all its results are compared with the CONLL tasks. Fonseca *et al.* propose separating words into their lemmas and flexional attributes, believing the system will benefit from the use of unlabeled data, considering the small amounts of labeled resources in Portuguese. Regarding the semantic role labeling task, it is unclear how the tags are assigned, in the authors' article.

In the original architecture of SENNA, considering it was designed for the English language, the words are not lemmatized. Fonseca *et al.* state that in Portuguese, the results would not be good, due to the great number of possible inflections. To solve this problem, and adapt the tool for Portuguese language, the authors proposed that the words were lemmatized, and stored in the tool's dictionary.

It is an interesting paper, and its results are awaited. The original SENNA tool presents state-of-the-art results for the English language (75.49% for F1 measure), and this adaptation can be a good way for semantic role labeling for Portuguese, though the semantic roles used, as well as in all the works that follow CONLL semantic roles tags (*Arg-n*), do not capture all information that we think it would be possible to obtain, when assigning a semantic role to a verb's argument.

Zílio *et al.* (2013) [67] aim to present a methodology to develop a lexical resource with semantic information, to compare semantic roles in specialized and non-specialized language and to observe manual semantic role labeling by human annotators. The author describes this lexical resource as being a mix of PropBank and VerbNet, because it contains sentences extracted from corpora as basis for annotation, as PropBank, and uses descriptive semantic roles, like VerbNet.

The corpora are composed of a specialized medical corpus and a non-specialized corpus taken from a Brazilian newspaper. The former containing approximately 1.6 million words, and the latter containing approximately one million words. Both corpora would be analyzed by the PALAVRAS parser [17]. A rule-based frame extractor (Zanette *et al.* (2012) [66]) will then extract the verbs, and for each verb in the sentence it searches for its dependencies. After this, the data is then stored in a database. To facilitate the annotation, an interface was built, which allows the annotator to choose the semantic relation for ARG_1 and ARG_2.

In total, 38 semantic roles were used in this paper, and they were taken from VerbNet [60]. Small changes were made to the list of semantic roles, in which the most relevant was the creation of the hypernym TARGET, which includes BENEFICIARY and PATIENT, in verbs were both semantic roles can be used.

One of the authors annotated a sample of the corpora, namely sentences including each of the 25 more frequent verbs in each corpus (occurring at least 10 times in each corpus). In total, 3,400 sentences were annotated, 1,790 of the specialized cardiology corpus and 1.610 of the newspaper corpus.

In the Cardiology corpus, the most frequent syntactic-semantic relation is subject as THEME[9], occurring 181 times, with a frequency of 10.11%. The second most frequent is subject as THEME, followed by an adverbial adjunct <in location>, occurring 121 times, with a frequency of 6.76%.

In the non-specialized newspaper corpus, the most frequent syntactic-semantic relations are the combinations of subject as AGENT, with a direct object, as THEME, occurring 171 times, with a frequency of 10.62%, and intransitive predicates with subject as THEME, occurring 114 times, with a frequency of 7.08%. The frequency of the sentences structures' on each corpora were compared, and the top structures are similar, and in some sentences, equal. The main difference between them is that, in the cardiology

[9]VerbNet considers THEME is "used for participants in a location or undergoing a change of location.".

corpus, the subjects are most of the times omitted. Another interesting difference is that in both newspaper and cardiology corpora, the most frequent semantic roles in subject are respectively, `AGENT` and `INSTRUMENT`.

Besides the methodology above, Zílio *et al.* intended to evaluate an inter-annotator agreement on semantic role labeling, where 10 annotators labeled the SRs on 25 simple sentences from the corpora above mentioned. The semantic role labeling proved to be a extremely difficult task, for only an inter-annotator agreement of approximately 0.25 was achieved.

Finally, as this paper only presents an methodology, no actual SRL system has been produced and evaluated.

Oliveira *et al.* (2010) [53], created a linguistic resource for Portuguese, PAPEL 2.0. This resource is made of relations between words, extracted automatically from a general language dictionary of Portuguese. The relations to be extracted are, for example, synonymy/antonymy, those of hypernym (type, gender, class, way of), meronymy (part of, member of), cause, finality (used to). PAPEL is a free resource that can be fully downloaded.

Grammars were built to extract each of the relations as trigrams consisting of the word-pair and the semantic relation holding between them; *e.g.* `casa SYNONYM lar`. After the grammars were applied, PAPEL 2.0 contains around 100,000 lexical items and close to 200,000 semantic relations, of which 79,161 are synonymity relations, and 62,591 are hypernym relations. n this way, synonyms and hypernyms are, by far, the most frequent relations in this resource, with the third most frequent semantic relation being Property, with 24,061 instances.

Other linguistics tools that are possible to use with the informations retrieved from PAPEL are FOLHEADOR [10], and VARRA (Freitas,2013 [31]). FOLHEADOR is an interface that allows the user to type in a word and the system shows a list of all the relations where that word is present. It is even possible to filter the results by gramatical category. VARRA is a manual validation system that allows users to validate PAPEL's semantic relations annotations, classifying the annotation from 1 to 5.

The information in PAPEL could be used to help annotate SR in unlabeled data, extending through the synonymy and other semantic relations the SR previously marked in a manually produced corpus. Naturally this is out of the scope of the current project.

[10] http://www.linguateca.pt/Folheador/

Chapter 3

Definitions of Semantic Roles

In this chapter, we review some of the terms and definitions of semantic roles, while trying, at the same time, to define an operative set of semantic roles, with a sufficiently clear-cut and reproducible definition, to be used in corpus annotation, building lexical resources and in view of natural language processing.

In PropBank (Palmer *et al.* 2005) [55], the notions of Arg_0, Arg_1 and Arg_2 are very close to the notions of the syntactic roles of subject, direct object and indirect object, respectively. In the PropBank framework, very specific SRs were created, and so, there are hundreds of them, specifically designed to match the verbs (often just one) they were devised for, which sometimes can be seen as tautologic. In fact, this methodology is justified by the belief that the semantic roles a predicative element (mostly verbs) assigns to its arguments is directly derived from its meaning, and it presupposes that at a later stage of processing, once the main Arg_n are determined, some mapping procedure will be carried out to link the (generic) argument status of those constituents with the (verb-specific) semantic roles and their tags, as determined by the predicative element they depend on.

One of the goals for the definition of a relevant set of semantic roles is to annotate ViPEr (Baptista 2012) [10], a database of European Portuguese full (or lexical or distributional) verbs. The semantic roles' choice was made carefully, in order to have just the necessary SRs to label correctly all the verbal constructions represented in ViPEr. Some of the SRs found on state-of-the-art review were found to be inadequate for this goal and where either adapted or dropped. Some other semantic roles were added, because none of the previous SR found in the literature was adequate to fill in the syntactic slots present in ViPEr.

Though Palmer *et al.* (2005) and our perspective here are both verb-oriented, in this work, a clear-cut distinction is established between the syntactic function a constituent may have and the semantic role it performs in relation to the verb it depends on. Besides, the semantic roles will be chosen in a less-specific way than in PropBank, in order to achieve some degree of generality.

In order to ensure reproducibility, the definition of a semantic role (or rather, the definition of the semantic prototype of that role) is methodologically/procedurally defined in function of the normal distribution for a given syntactic slot of a predicate type; the awarding of a semantic role to any constituent of a sentence in a particular instance of a predicate/operator (be it a verb, an adjective, a noun, etc.) results from a judgment of (dis)similarity between that particular instance of a predicate-argument relation and the set of distributions provided in the mentioned role's definition. This methodology is illustrated in the sections below.

We set out by assuming a major distinction between the semantic roles pertaining to human and non-human arguments. Evidence from many languages suggest that this distinction is pervasive across languages, often having a formal (morphological and/or syntactical) counterpart (e.g. the s/he-it distinction in English, or the *quê-quem* opposition in Portuguese). Table 3.1 presents a synopsis of the set of semantic roles associated with the human feature, while Table 3.2 present the roles associated to the non-human arguments. These are complemented by other, circumstantial roles, such as locative and time-related roles. In the end, a set 37 semantic roles has been defined. The definitions and examples of these are shown in the next subsections.

In the following paragraphs, we briefly comment on the terminology for designating SRs and the definitions provided for each one, in order to better justify the set of SR that we have adopted for this research project.

3.1 AGENT

The first and most consensual semantic role is that of `AGENT`. It is usually defined as the entity being represented by a human noun, i.e. a noun referring to a person who willingly performs an action. The concept of **human** may be disputed in some situations, but its core meaning can be linked to the referentiality of such nouns. Although **human** is a semantic concept, we adopt M. Gross (1975) [35] more formal definition, considering this feature to result from placing any noun as the subject of verbs like *dizer* (to say), *pensar* (to think), or as the direct complement of verbs such as *alegrar* (to please) or *irritar* (to irritate).

Fillmore (1968)[29] (apud [56]), refers the fact that animal-denoting nouns and human nouns can often be used as the semantic role `AGENT`. Besides, we include the concept of **human-collective** as an extension by metonymy of the concept of **human**, to cover organizations and similar named entities [52].

The related concept of **action** inserts some degree of circularity in the definition, but it is still a useful component of the definition as this noun can function as a valid classifier of many predicates: compare *ler* (to read), *saltar* (to jump), *comer* (to eat), *etc.*, which are undoubtedly classified as **actions**; with the more dubious *morrer* (to die), *pensar* (to think), *apreciar* (to appreciate), *ficar localizado* (be located), *etc.* for which "action" does not seem to be an adequate classifier.

In the reviewed literature, it is not clear whether the distributional restriction to human nouns is agreed upon by all the authors, but it seems implied in the Wikipedia definition (use of the adverb *deliberadamente* (deliberately)), in the attribute "capable of volition" of Palmer *et al.* (2010) [56], or in Allen's (1994) [4] use of the adjective "intentional". FrameNet definition is somewhat vague, as to be "engaged" does not necessarily involves **volition**.

The notion of **motion** is often associated with the concept of `AGENT`, as in the subject of verbs like *correr* (to run), *nadar* (to swim), as well as that of **manipulation** of an `OBJECT-GENERIC`, *i.e.*, the subject of *fritar* (to fry), *sublinhar* (underline), *cortar* (cut).

Since `AGENT` is a very broad concept, more specific subtypes of this SR can be defined. These are presented below.

A more specialized meaning is that of `AGENT-SPEAKER`, with **communication** or **declarative** verbs, like the subject of *contar* (tell) or *declarar* (declare); this SR is to be related with their counterparts in a communicative process, namely `MESSAGE`, `TOPIC` and `ADDRESSEE` (see section 3.14):

(9) *O Pedro/`AGENT-SPEAKER` contou uma história/`MESSAGE` ao João/`ADDRESSEE`*
'Peter told John a story'

(10) *O Pedro/AGENT-SPEAKER declarou que isso era importante/MESSAGE*
'Peter declared that that was important'

The semantic role sub-type of AGENT-GIVER can be defined for the subject of most **transfer** predicates as in *dar* (to give), *oferecer* (to offer), *distribuir* (to distribute);

(11) *A avó/AGENT-GIVER deu umas meias/OBJECT-GENERIC ao neto/RECIPIENT*
'The grandmother gave her grandson some socks'

The AGENT-GIVER role also applies to verbs like *atirar* (to throw), as in:

(12) *O manifestante/AGENT-GIVER atirou uma pedra/OBJECT-GENERIC ao carro da polícia/LOCATIVE-DESTINATION*
'The manifestant threw a rock at the police car'

In this case, there is a **motion transfer** involved, so a locative complement is assigned a DESTINATION role (see 3.15)

The AGENT-GIVER is to be used in actions of **transfer** of an OBJECT, which are performed intentionally or willingly. It involves **motion** (of the subjects) and it also involves **transference** of an OBJECT-GENERIC from a person, the AGENT-GIVER, to a place, LOCATIVE-DESTINATION.

Conversely, the AGENT-TAKER semantic role was defined for the subject of verbs like *roubar* (to steal) that is also said to be agentive, but where the **transference** of the OBJECT-GENERIC has the opposite orientation.

(13) *O Pedro/AGENT-TAKER roubou o livro/OBJECT-GENERIC ao João/VICTIM*
'Pedro stole John's book'

There is dispute as whether AGENT should be associated with the notion of CAUSE. For example, *causador* (causer) is explicitly used in the definition of PAPEL and in Allen's. In almost all definitions (but not in Allen's), the semantic role is referred to as an entity, designated by a nominal or a nominalized verb: "causador da acção" (causer of the action)[53], "initiator of action" [56], or else, implied by the subjectless use of third-person singular "performs the action"(Wikipedia). Allen's (1994) definition[4] is, by far the most divergent, as it is centered on the abstract and subjectless noun "causation", though modified by "intentional".

The CAUSE semantic role can often compete with AGENT for the same syntactic slots, and in some cases both interpretations can be associated to a **human** noun, as in the subject of psychologic verbs (Mendes, 2001 [49]) like *irritar* (to irritate) or *agradar* (to please); in these semantic predicates, a **causal**, non-agentive interpretation can be found in sentences like the following example, where one clearly can not say that the subject is doing any action at all:

(14) *A Ana/CAUSE irrita-me/PATIENT com aquela carinha sonsa.*
'Ana irritates me with that underhand face'

In this case, *Ana* has the semantic role of CAUSE, while in:

(15) *A Ana/CAUSE irritou o Pedro/PATIENT com insultos.*
'Ana irritate Pedro with insults'

an agentive reading seems preferable.

On the other hand, certain actions are ambiguous between an agentive and non-agentive reading:

(16) *O Pedro/AGENT-GENERIC partiu a janela/OBJECT-GENERIC*
'Peter broke the window'

In this case, it may be preferable to leave the distinction between AGENT and CAUSE underspecified, as an AGENT-CAUSE semantic role (see also 3.10).

In view of the above, we adopt a definition where AGENT is always expressed by a **human** (eventually a human-collective) noun. It is associated with the concepts of **volition** and **motion**. When associated with another role, OBJECT-GENERIC (see below), is often associated with the general prototype of **transfer** of an OBJECT-GENERIC. For these predicates, the specific subtypes of AGENT-GIVER and AGENT-TAKER were devised. Finally, this SR can also be specified as an AGENT-SPEAKER, in the case of **communication** predicates, for the entity with an active role in the speech act.

Finally, for verbs implying the **apparition** of a new OBJECT-GENERIC that comes into being at the conclusion of a process, like *confecionar* (to prepare), *construir* (to build),The SR AGENT-CREATOR was adopted from FrameNet [59], in order to distinguish this predicates from those where the OBJECT-GENERIC pre-exists the process and only undergoes some type of **manipulation** or **transformation**.

(17) *O pedreiro/AGENT-CREATOR construiu a casa/OBJECT-GENERIC*
'The worker has built the house'

(18) *O cozinheiro/AGENT-CREATOR confecionou uma refeição/OBJECT-GENERIC* (apparition)
'The cook has prepared a meal'

3.2 PATIENT

The term "patient", derived from the Greek "pathos" 'suffering', though its negative polarity is not necessarily implied in the use of the word as a semantic role. Like AGENT-GEN, it is not clearly associated to **human** entities. In fact, both Palmer *et al.* (2010) [56] and Wikipedia define the term by the subjectless verbs "affected" and "undergo". Again, Allen (1994) [4] clearly uses the term "thing", which, while not explicitly denying the possibility of a human entity, tends to be interpreted in a non-human reading.

One of Palmer *et al.* [56] examples is:

(19) 'John/AGENT-GENERIC broke the window/PATIENT'

In this sentence it is possible to observe that the role PATIENT is not strictly associated with **human** entities.

FrameNet defines PATIENT as the entity being acted on, and that may, but need not, undergo a change. Though PATIENT is defined in such broad terms, many different subtypes of PATIENT are also defined:

- SUSPECT in the action "arrest" (sic)
  The police/ AUTHORITIES *arrested Harry*/ SUSPECT

- VICTIM in the action “attack" (sic);
  The USA/ ASSAILANT *invaded Iraq*/ VICTIM

- EMPLOYEE in the action of “employing" (sic);
  The boss/ EMPLOYER *hired John*/ EMPLOYEE

and many others subtypes.

As one can see by these terms, they can be considered to be either too specific to the predicates' arguments they are applied to or even circular in their definition, since they seem to result from the very predicate they try to describe. In our view, though the concept of semantic role may occasionally require some adaptation to specific semantic predicate types, such *minutiae* leads to a loss in generality and, thus, less utility for a general linguistic description such as the one here envisaged.

Finally, we have chosen not to use the term THEME here, often used for the same semantic roles as `PATIENT`, because we find it to be not sufficiently clearly defined in the references. The term used in this project will be `PATIENT`, to refer to the **human** entity, **affected** (positively or negatively) by the process.

3.3 BENEFICIARY

In several dictionaries, such as the Oxford Dictionary, the `BENEFICIARY` represents a person who derives advantage from something. Unlike the previous roles, this one is defined as being strictly applied to entities with the **human** feature. Palmer *et al.* (2010) [56], Allen (1994) [4] and also Wikipedia consider the `BENEFICIARY` as a human entity, referring to the entity, to whose benefit the action is performed” and, the person for whom an action is done”. This SR always refers to an entity outside the predicate's argument domain (essential complements), that is, it is not one of its essential arguments.

Examples of this semantic role are:

(20) *O João/`AGENT-GENERIC` descascou uma laranja/`OBJECT-GENERIC` para a Joana/`BENEFICIARY`*
‘John peeled Joana an orange’

(21) *Os pais/`AGENT-GENERIC` arrendaram um apartamento/`OBJECT-GENERIC` para o filho/`BENEFICIARY`*
‘The parents rented a house for John’

As we can observe, in both actions the `AGENT` performs an action on behalf of the `BENEFICIARY`.

3.4 VICTIM

While `BENEFICIARY` had a positive connotation, the semantic role of `VICTIM` has been proposed [59] for the person affected by a process, though it is here used only for those constituents that are not an argument but mere adjuncts of the predicate.

(22) *O João/`AGENT-GENERIC` estragou o relvado/`OBJECT-GENERIC` à Ana/`VICTIM`*
‘Peter ruined the lawn’

3.5 EXPERIENCER

This semantic role is common to all the sources reviewed, and their definitions of `EXPERIENCER` are very similar. In Ruppenhofer *et al.* (2010) [59], it is defined as being "the person or sentient entity that experiences or feels the emotions". This SR is used in examples such as the following:

(23) *O Pedro/`EXPERIENCER-GENERIC` sentiu medo*
'Peter felt fear'

(24) *A Ana/`EXPERIENCER-GENERIC` tem frio*
'Ana is cold'

The semantic role `EXPERIENCER` will be used in this project as being the entity with the human feature that experiences a physical or psychological state. It is non-agentive and usually **non-volitional**.

This SR is also given to the subject of predicates of **mental perception**, such as *pensar* (to think) and to **volitive** verbs, such as *querer* (to want) of ViPEr's class 6. Volition and imperative verbs require a non-agentive human subject and an `OBJECT-F` complement, usually expressed by a subjunctive sub-clause and receive the sub-type feature of `-VOL`:

(25) *O Pedro/`EXPERIENCER-GENERIC` pensou na Ana/`OBJECT-GENERIC`*
'Peter thought of Ana'

(26) *O João/`EXPERIENCER-VOL` quer ir ao cinema/`LOCATIVE-DESTINATION`*
'Peter wants to go to the movies'

Certain verbs of **mental perception** (v.g. *pensar em, pensar de, perceber de, pescar de, saber de, tratar de, palrar de*) require a `TOPIC` complement (see below in 3.14).

3.6 POSSESSOR

There is no consensus on the term for this semantic role. Ruppenhofer *et al.* (2010) [59] named it OWNER and Allen (1994) [4] named it AT-POSS. Nevertheless, all authors agree on the definition. It is agreed that it is the entity that "owns a possession" [59] (which is a somewhat tautologic definition). Allen (1994) [4] defines as the "current possessor", though this aspectual value conveyed by the adjective *current* may not be absolutely necessary. The most basic verb that expresses **possession** is *to have* (*ter*, in Portuguese):

(27) *O João/`POSSESSOR` tem uma linda bicicleta/`OBJECT-GENERIC`*
'John has a beautiful bicycle'

or *belong to* (*pertencer*, in Portuguese), where the relative position of `OBJECT-GENERIC` and `POSSESSOR` are reversed.

(28) *Esta linda bicicleta/`OBJECT-GENERIC` é do/pertence ao João/`POSSESSOR`*
'This beautiful bicycle is John's/belongs to John'

The entity to which this SR is assigned is always attached to a non-human concrete noun, often an OBJECT, as in the next example:

(29) *A bicicleta/OBJECT-GENERIC do João/POSSESSOR foi roubada*
'John's bicycle was stolen'

This SR will be used, from now on, in this project, as referring to the entity that possesses an OBJECT, which involves a static, durative or permanisive aspect, and which is distinct from the AGENT-GIVER, which has a dynamic feature.

3.7 RECIPIENT

The RECIPIENT is many times related with the AGENT-GIVER. The term is suggested by Ruppenhofer *et al.* (2010) [59], where this role is defined as indicating "the entity that ends up in possession of the theme". Allen (1994) [4] uses the term TO-POSS, though recognizing the existence of the term RECIPIENT and defines it as the "final possessor". An example of the SR can be seen in the example of the semantic role AGENT-GIVER:

(30) *A avó/AGENT-GIVER deu umas meias/OBJECT-GENERIC ao neto/RECIPIENT*
'The grandmother gave her grandson some socks'

(31) *O patrão/AGENT-GIVER pagou (o salário) ao Pedro/RECIPIENT*
'The boss payed Pedro(his salary)'

The RECIPIENT always involves an AGENT-GIVER, but it may not involve an OBJECT-GENERIC because this may be implied, as shown on the second example.

3.8 COMITATIVE

None of the authors refers the COMITATIVE semantic role, though this is a "commonly recognized [case]" (Anderson, 1999: pp. 65) [20]: "The 'comitative' usually marks the animate (typically human) which is conceived of as accompanying the participation of some more centrally involved participant in a predication". In this sense, this role is to be used only in non-essential (adjunct) complements and is to be distinguished from symmetrical roles (see 3.17), for example, using paraphrases such as *na companhia de* (in the company of), and the like.

(32) *O Pedro/AGENT-GENERIC foi passear no jardim/LOCATIVE-PLACE com a Ana/COMITATIVE*
'Peter went for a walk in the park with Ana'

Table 3.1: Semantic Roles for Portuguese: a preliminary proposal. 1- Human arguments

Semantic Role	Definition	Example (translation)
AGENT-GENERIC (agent-generic)	The **human** entity that performs an **action**; implies **volition**, **motion**	***O Pedro*** *cantou o fado* (Peter sung the Fado)
AGENT-SPEAKER (agent-speaker)	AGENT, source of the MESSAGE in **communication** predicate	***O João*** *disse isso à Ana* (John said that to Ana)
AGENT-GIVER (agent-giver)	AGENT in **transfer** predicate: orientation: subj > OBJECT-GENERIC > indir.compl.	***A Ana*** *deu uma prenda ao João* (Ana gave John a present)
AGENT-TAKER (agent-taker)	AGENT in **transfer** predicate orientation: subj < OBJECT-GENERIC < indir.compl.	***O João*** *roubou a mala à senhora* (John stole the old lady's purse)
AGENT-CREATOR (agent-creator)	AGENT in **apparition** predicates	***O João*** *construiu uma casa* (John built a house)
EXPERIENCER-GENERIC (experiencer-generic)	The **human** entity that experiences a physical or psycological state; also the subject of mental perception predicates (pensar)	*Isso irritou* ***o Pedro*** (That irritated Peter)
EXPERIENCER-VOL (experiencer-vol)	The **human** entity that expresses an order or desire	***O Pedro*** *ordenou-a que fosse* (Peter told her to go)
PATIENT (patient)	The **human** entity directly *affected* by an **action**	*O Pedro magoou* ***a Ana*** (Peter has hurt Ana)
RECIPIENT (recipient)	The **human** entity that receives something from an AGENT-GIVER in a transfer predicate	*A Ana deu-**lhe** umas meias* (Ana gave him some socks)
ADDRESSEE (addressee)	The **human** entity that is indirect OBJECT-GENERIC of a **communication act**	*O Pedro disse que sim* ***à Ana*** (Pedro said yes to Ana)
BENEFICIARY (beneficiary)	The **human** entity benefited by an **action** (though it is not an essential argument)	*O João abriu caminho para* ***a Ana*** (John made way for Ana)
VICTIM (victim)	The **human** entity affected by an **action**, (though it is not an essential argument)	*O Pedro estragou-**lhe** o relvado* (Peter ruined her the lawn)
POSSESSOR (possessor)	The **human** entity that owns an OBJECT-GENERIC	***O Pedro*** *tem uma bicicleta* (Peter has a bycicle)

The next sections present the semantic roles that are used mostly for non-human arguments, and a synopsis is provided in Table 3.2.

3.9 OBJECT

This term is not used by any of the sources. Palmer *et al.* (2010) [56], Wikipedia and Allen (1994) [4] use the term THEME. However, the three definitions are very different from each other. Palmer *et al.* (2010) [56] define THEME as being, the entity moving or being located". We do not adopt these authors' definition, and rather use a broader concept of OBJECT-GENERIC, irrespective of the concept of **motion**.

Allen [4], defines THEME as "the thing affected by the event", which is a closer definition to the concept that we wish to capture by this semantic role. The term "thing" may denote a non-human concrete entity, which is exactly the semantic feature we want to capture with this semantic role. The term EVENT (3.12) may be too specific to capture the range of predicates that involve OBJECTs, namely, processes and states are often related to OBJECTs. On the other hand, the term "affected" is rather vague, as an OBJECT-GENERIC may not be necessarily affected in a process, but just be involved in it. For example, in *to grab a pen* the action does not affect *the pen*, though the OBJECT-GENERIC is manipulated, moved, etc. in the process. The term "affected" may be related to the concept of **change of state**, as used in the definition of THEME in the Wikipedia: "undergoes the action but does not change its state". However, it would be more convenient to use a broader definition of OBJECT-GENERIC that could also be applied to predicates where there is no change of state, as in **transfer** and **manipulation** predicates:

(33) *O sol/**CAUSE** derreteu o gelo/**OBJECT-GENERIC***
'The sun melted the ice'

(34) *A Ana/**AGENT-GIVER** deu uma camisa/**OBJECT-GENERIC** ao namorado/**RECIPIENT***
'Ana gave her boyfriend a shirt'

The semantic role of OBJECT-GENERIC will thus be applied to non-human concrete nominals usually functioning as direct complements of transitive verbs.

Certain classes of predicates allow for a more precise or specific concept of OBJECT-GENERIC.

OBJECT-CL

Body part nouns (*Nbp*) selected by verbs such as *suar* (to sweat) and *pentear* (to brush), such as the examples below:

(35) *Os pés/**OBJECT-CL** do João/**PATIENT** suam*
'John's feet sweat'

(36) *A Ana/**AGENT-GENERIC** penteou os seus cabelos/**OBJECT-CL***
'Ana brushed her hair'

In these cases, special restructuring of the *Nbp* NP's may yield a locative complement.

(37) *O João/**PATIENT** sua dos pés/**OBJECT-CL***
'John sweats from its feet'

OBJECT-Q

Quantified NP's with a concrete non-human head or measure units with verbs that express quantification, such as *medir, pesar, valer*, etc.

(38) *As batatas/**OBJECT-GENERIC** pesam dois quilos/**OBJECT-Q***
'The potatoes wheigh two quilograms'

OBJECT-F

The OBJECT of mental perception verbs involving a propositional content, such as *adorar, compreender*:

(39) *O João/**EXPERIENCER-GENERIC** adora fazer desporto/**OBJECT-F***
'John loves to make sports'

(40) *A Joana/**EXPERIENCER-GEN** compreende a razão de a Ana não poder vir/**OBJECT-F***
'Joana understands the reason that Ana can not come'

3.10 CAUSE

The semantic role CAUSE is sometimes merged with that of AGENT (3.1). Allen (1994) [4] calls it CAUSAL-AGENT and defines it as being, the OBJECT that caused the event. Ruppenhofer *et al.* (2010) [59] defines it as "a force, process or event that produces an effect". Both Wikipedia and Gildea and Jurafsky (2002) [33] have a SR called FORCE, but we chose not to use this term, because: (i) Wikipedia shows a definition very similar to the one of CAUSE presented above: "what caused the action to occur in the first place"; while (ii) Gildea and Jurafsky (2002) [33], offer no definition at all. This SR will then be used to represent a non-human force that causes an event to occur.

(41) *O tornado/CAUSE destruiu a cidade/LOCATIVE-PLACE*
'The tornado destroyed the city'

(42) *O raio/CAUSE atingiu o João/PATIENT*
'The lightning struck John'

The notion of CAUSE is also expressed by certain lexical (or full) verbs, such as *causar* (to cause), *provocar* (to provoke), as well as by structures that Maurice Gross (1981)[36] named *causative operator verbs* (*Vopc*). These constructions are characterized by imposing weak distributional constraints on the subject syntactic slot, which has been designated by this author as a "nom non-restraint" (unconstraint noun, *Nnr*). This explains why this type of verbs (and other with a CAUSE subject) allow for both an agentive and non-agentive reading when that syntactic slot is filled in by a human noun. Furthermore, this slot can often show not only noun phrases with the CAUSE semantic role, but also subclauses, and specially *factive* subclauses, that is, sentences introduced by *o facto de* (the fact that). For example:

(43) *O João/CAUSE destruiu a harmonia familiar/OBJECT-GENERIC*
'John destroyed the family harmony'

(44) *A atitude do João/CAUSE destruiu a sua reputação/OBJECT-GENERIC*
'John's attitude destroyed his reputation'

(45) *O facto de o João ter fugido de casa/CAUSE afectou a saúde da Ana/OBJECT-F*
'The fact that John ran away destroyed his relationship with Ana'

3.11 INSTRUMENT

The existence of a semantic role INSTRUMENT is unanimous. Ruppenhofer *et al.* (2010) [59] called it TOOL, but the concept is the same as the other authors, who called it INSTRUMENT. The unanimous definition is that it refers to an entity used to carry out an action. The following examples show the use of this role.

(46) *O Pedro/AGENT-GENERIC bateu no João/PATIENT com um martelo/INSTRUMENT*
'Peter has hit John with a hammer'

(47) *O João/AGENT-GENERIC pintou o carro/OBJECT-GENERIC à pistola/INSTRUMENT*
'John painted the car with a paint spray gun'

(48) *O João/AGENT-GENERIC matou o Pedro/PATIENT à facada/INSTRUMENT*
'John killed Peter by stabbing'

This SR is usually expressed by a prepositional phrase (PP) with concrete instrument nouns. They are introduced by preposition *com* (with), although other prepositions can be found, like *a* (to), in *à pistola*. Certain PPs with abstract, derived nouns, like *à facada*, are also to be given this semantic role. This last example is derived from a concrete instrument noun, *faca* (knife).

3.12 OCCURRENCE

An OCCURRENCE is something that "takes place at a place and time" as stated in Ruppenhofer *et al.* (2010) [59]. Examples of this semantic role are:

(49) *O jogo/OCCURRENCE vai decorrer no Porto/LOCATIVE-PLACE*
'The game will take place at Oporto'

(50) *O tornado/OCCURRENCE ocorreu ontem/TIME-CALENDAR*
'The tornado occurred yesterday'

This SR will be used for the subject of *occurrence support verbs* (Gross, 1986 [37]) such as *ocorrer* (to occur), *decorrer* (to take place), *suceder-se* (to succeed), etc. These are a finite list of verbs used mostly to introduce circumstantial complements in sentences.

The role of OCCURRENCE is also relevant for certain predicates such as *participar em* (to participate in), *assistir a* (to attend).

The term *occurrence* was preferred to the competing term *event*, as in the STRING this word as already a defined status (Baptista *et al.* 2010)[15].

3.13 MANNER

MANNER is, according to Wikipedia, "the way in which an action is carried out". This semantic role captures the adverbs in texts, as shown in the next examples:

(51) *O João/AGENT-GENERIC aterrou o avião/OBJECT-GENERIC de forma cuidadosa/MANNER*
'John landed carefully'

(52) *O João/AGENT-GENERIC aterrou o avião/OBJECT-GENERIC cuidadosamente/MANNER*
'John landed carefully'

(53) *O Pedro/AGENT-GENERIC sopra de forma furiosa/MANNER*
'Peter blows furiously'

(54) *O Pedro/AGENT-GENERIC sopra furiosamente/MANNER*
'Peter blows furiously'

This SR is normally expressed by a verb followed by a modifier (MOD), which is already captured by the STRING system. Besides, a finer classification of binary adverbs has already been made (Palma (2009)[54], Gaia (2011)[24]), and manner adverbs have been given a specific feature that enables their identification in texts (Baptista, 2012 [14])- The challenge will be to identify periphrasal constructions with operator-nouns, such as *modo* (mode), *maneira* (manner) and *forma* (form) [1].

3.14 Communication predicates: AGENT-SPEAKER, MESSAGE, TOPIC and ADDRESSEE

The predicates of **communication** involve an `AGENT-SPEAKER` and two arguments, the direct and indirect complements. These are given the semantic roles of `MESSAGE` ,`TOPIC` and `ADDRESSEE`, respectively. As stated in 3.1, the `MESSAGE` is typically the direct object of a **communication** act.

(55) *A Ana/**AGENT-SPEAKER** disse ao João/**ADDRESSEE** que não gostava dele/**MESSAGE***
'Ana told John that she doesn't love him'

`ADDRESSEE`, on the other hand, is the entity to whom a `MESSAGE` is directed.

(56) *O João/**AGENT-SPEAKER** gritou com a Ana/**ADDRESSEE***
'John shouted at Ana'

Ruppenhofer *et al.* (2010) [59] suggests the term READER be used for written messages, but we find this term and its use to be too specific, and so, we used the `ADDRESSEE` for any type of communication predicate, irrespective of the written or oral nature of the message.

Certain verbs of **mental perception** (v.g. *pensar em, pensar de, perceber de, neg. pescar de, saber de, tratar de*) require a `TOPIC` complement.

(57) *O Pedro/**AGENT-SPEAKER** falou desse assunto/**TOPIC** com a Ana/**CO-AGENT***
'John spoke about that topic with Ana'

This SR is not applicable to sentential arguments, and it does not involve the conveying of a content (see 3.14).

[1] In Brazilian Portuguese, there is also another operator noun, *jeito* (way/manner).

Table 3.2: Semantic Roles for Portuguese: a preliminary proposal. 2 - Non human arguments

Sem. Role	Definition	Example (translation)
OBJECT-GENERIC (object-generic)	**Non-human, concrete** entities, usually direct complements of transitive verbs	*O Pedro agarrou* ***a pedra*** (Peter caught the stone)
MESSAGE (message)	Typically, the direct object of a *communication* act	*O João disse* ***que não gostava dela*** (John said he didn't like her)
OCCURRENCE (occurrence)	The subject of **occurrence** support-verbs and other special predicates	***O jogo*** *ocorreu ontem* (The game occurred yesterday)
CAUSE (cause)	**Non-agentive** NP, PP or clause, interpreted as CAUSE; it can be a **human** NP	***O tornado*** *destruiu a cidade* (The tornado destroyed the city)
COMITATIVE (comitative)	Non essential **human** entity that accompanies a more centrally involved entity	*O Pedro foi à cidade com* ***a Ana*** (Peter went to the city with Ana)
MANNERSR (manner)	**Manner** adverbs and similar adverbials *como? (how)*	*O Pedro respira* ***cuidadosamente*** (Pedro breathes carefully)
INSTRUMENT (instrument)	OBJECT-GENERIC used to perform an **action**	*O João bateu-lhe com* ***o martelo*** (John has hit him with a hammer)
TOPICSR (topic)	Subject matter usually in NP's introduced by *de*, *em*, *sobre* or *acerca de*, and associated with communication acts	*O Pedro falou com a Ana sobre* ***a peça*** (Peter spoke to Ana about the play)
OBJECT-F (object-f)	The direct complement of **mental perception** verbs such as *adorar*. Usually **completive** sentences	*O Pedro percebe* ***que a Ana o ama*** (Peter understands **that Ana loves him**)
OBJECT-CL (object-cl)	**Human body parts**, usually functioning as direct complements of transitive verbs	*O Pedro bateu na parede com* ***a cabeça*** (Peter has hit the wall with the head)
OBJECT-Q (object-q)	Quantified NP's or PP's expressing a measurement, a value or an amount; involving measures units. The verb is 32NM or the like. Associated with *quanto(s)?* (how much/many?)	*O Pedro mede* ***2 metros de altura*** (Peter is 2 meters high)

3.15 LOCATIVE Semantic Roles

We consider four locative semantic roles associated with locative predicates: PLACE, PATH, SOURCE and DESTINATION. All of them are consensual among the authors previously referred to, even if the terminology differs.

PLACE, though it is sometimes named LOCATION, is unanimously defined as the place "where the event takes place" [59]:

(58) *O crime/OCCURRENCE ocorreu em Faro/LOCATIVE-PLACE*
'The crime occurred in Faro'

(59) *Chove muito em Lisboa/LOCATIVE-PLACE*
'It rains a lot in Lisbon'

In this project,we define this SR for **locative-static** predicates, that is, those not involving **motion**.

PATH is defined by both Allen (1994) [4] and Ruppenhofer *et al.* (2010) [59], as being "the ground over which someone/something travels". Examples of this role are:

(60) *A Ana/AGENT-GENERIC atravessou Lisboa/LOCATIVE-PATH*
'Ana crossed Lisbon'

(61) *O Pedro/**AGENT-GENERIC** passeou pelo pátio/**LOCATIVE-PATH***
'Pedro crossed the yard'

SOURCE is designated by the references above with three different terms. Allen (1994) [4] names it FROM-LOC, Palmer *et al.* (2010) [56] and Wikipedia named it ORIGIN and Ruppenhofer *et al.* (2010) [59] named it SOURCE. While the term may be disputed, the definition is clearly the same for all the authors. An example of this SR is:

(62) *O Pedro/**AGENT-GENERIC** partiu de Lisboa/**LOCATIVE-SOURCE***
'Peter departed from Lisbon'

We have adopted a definition where SOURCE is the original location prior to the motion.

The forth and last locative SR is DESTINATION, which is the opposite of SOURCE. The name DESTINATION is almost consensual, and so is its definition, among the authors researched. Allen (1994) [4] calls it TO-LOC, but gives it the same definition as the other authors.

(63) *O Pedro/**AGENT-GENERIC** chegou ontem/**TIME-CALENDAR** a Faro/**LOCATIVE-DESTINATION***
'Peter arrived yesterday to Faro'

As can be seen in the previous example, these semantic roles will be used to refer to locations involving **motion**, that is, in a **locative-dynamic** predicate.

Table 3.3: Semantic Roles for Portuguese: a preliminary proposal. 3 - Locative arguments

Sem. Role	definition	Example (translation)
LOCATIVE-PLACE (locative-place)	Static locative predicates ; "onde? (where)"	*Chove muito **em Lisboa*** (It rains a lot in Lisbon)
LOCATIVE-PATH (locative-path)	Dynamic locative predicates "por onde? (through where?)"	*A Ana atravessou **Faro*** (Ana crossed Faro)
LOCATIVE-SOURCE (locative-source)	The location prior to the **motion** with dynamic locative predicates	*O Pedro partiu **de Lisboa*** (Peter departed from Lisbon)
LOCATIVE-DESTINATION (locative-destination)	The location after the **action** with dynamic locative predicates	*A Ana chegou ontem **a Faro*** (Ana arrived in Faro yesterday)

3.16 TIME Semantic Roles

As we can observe from fig. 3.4, only the **temporal locative** semantic role is referred by the authors here revised. For the time SRL, we benefit from extensive previous work already done for STRING, particularly in the Named Entity Recognition task ([39], [40], [48]).

Hence, NE already captured by the STRING system will be awarded the SR of TIME. Since the system uses a finer ontology of time NE categories, the values of these NE will be propagated to the SR label. Thus, we consider the following TIME semantic roles: TIME-CALENDAR, TIME-FREQUENCY and TIME-DURATION ([48]).

TIME-CALENDAR stands for the sub-types DATE (absolute or referential), INTERVAL and HOUR. The following sentences from Hagège *et al.* (2008) [39] will exemplify each sub-type of TIME-CALENDAR:

DATE

(64) *Vou viajar no dia 19 de Outubro de 2007/*`TIME-CALENDAR`
'I will travel on October 19th, 2007'

(65) *Vou a Lisboa/*`LOCATIVE-PLACE` *no próximo dia 22/*`TIME-CALENDAR`
'I will go to Lisbon on 22nd of October'

INTERVAL

(66) *Trabalhei em Londres/*`LOCATIVE-PLACE` *durante entre 2000 e 2003/*`TIME-CALENDAR`
'I worked in London between 2000 and 2003'

To use this sub-type, the gap of time has to be explicit, as shown on the examples above.

HOUR

(67) *O Pedro está disponível às 15:00/*`TIME-CALENDAR`
'Peter is available at 15:00'

This sub-type is applied when the time expression has a smaller granularity than the unit day.

TIME-FREQUENCY

The `TIME-FREQUENCY` label stands for the time expressions that express a repetition in time. The following examples will exemplify the use of this label. As stated in Maurício (2011) [48], this sub-type answers the question *com que frequência?* (how often?)

(68) *Os meus pais/*`AGENT-GENERIC` *vêm cá diariamente/*`TIME-FREQUENCY`
'My parents come here daily'

TIME-DURATION

`TIME-DURATION` represents the amount of time an action takes to occur.

(69) *O jogo/*`OCCURRENCE` *dura 60 minutos/*`TIME-DURATION`
'The game lasts 60min'

This SR sub-type answers the question *(prep) quanto tempo?*((prep) *how long?*), as in the following QA example:

- *Q: Quanto tempo dura o jogo?* (How long does the game take?)
- *A: O jogo dura 60 minutes* (It takes 60 minutes)

Table 3.4: Semantic Roles for Portuguese: a preliminary proposal. 4 - Time arguments

Sem. Role	definition	Example (translation)
TIME-CALENDAR (time-calendar)	Absolute dates; *quando?* (when?)	*Vou viajar no* ***dia 19 de Outubro*** (I will travel on October 19th)
TIME-FREQUENCY (time-frequency)	Time expressions that express a repetition in time; *how often?* (quantas vezes?)	*Os meus pais vêm cá* ***diariamente*** (My parents come here daily)
TIME-DURATION (time-duration)	The amount of time an action takes to occur; *how long?* (quanto tempo?)	*O jogo dura* ***60 minutos*** (The game lasts 60 minutes)

3.17 Symmetric Semantic Roles

Symmetric predicates (Baptista, 2004 [8]) are built with two arguments that play the same semantic role. Because of this particular semantic property, the arguments can change places without changing the meaning of the sentence; they can also appear coordinated in the same syntactic slot; and, since these predicates are intrinsically reciprocal, the presence of an echo complement is facultative, *v.g. um com o outro* (with each other) or *entre si* (among each other/between them)[2]:

(70) *O Pedro conversou com a Ana = A Ana conversou com o Pedro*
'Peter chatted with Ana = Ana chatted with Peter'

(71) *= (O Pedro e a Ana = A Ana e o Pedro) conversaram (E + um com o outro + entre si)*
'Peter and Ana = Ana and Peter chatted with each each other'[3]

As far as verbs are concerned, four types of symmetric constructions exist in Portuguese (Baptista 2012), depending on the sentence structure and the symmetric arguments. When one of the symmetric arguments appears in a detached PP, it plays the same SR as its symmetric counterpart. To distinguish both, the CO- prefix is used. Hence, the following symmetric semantic roles are defined:

- [35S] O Pedro/AGENT-GENERIC casou com a Ana/CO-AGENT (Peter married with Ana)
- [36S1] O Pedro/AGENT-GENERIC misturou a farinha/OBJECT-GENERIC com o açúcar/CO-OBJECT (Peter mixed sugar and flour)
- [36S2] O Pedro/AGENT-GENERIC combinou com a Ana /CO-AGENT fazer isso/OBJECT-GENERIC (Peter arranged with Ana to do that)
- [42S] O Pedro/AGENT-GENERIC conspirou com o João/CO-AGENT contra a Ana/VICTIM (Peter conspired with João against Ana)

Verbs of class 35S can sometimes present a parallel causative constructions with a AGENT-CAUSE subject, yielding a structure that is formally similar to that of class 36S1. In this case, the symmetric arguments have the role of PATIENT (and CO-PATIENT):

[2]Unlike ordinary reciprocal constructions, which are to be derived from the coordination of two sentences with the arguments in reversed positions: *O Pedro conta com a Ana <para isso> e a Ana conta com o Pedro <para isso> = O Pedro e a Ana contam (*E + um com o outro) <para isso>* (Peter counts on Ana <for that> and Ana counts on Peter <for that> = Peter and Ana count on each other <for that>).

[3]Notation: the symbol 'E' stands for the empty string; elements between brackets and linked by '+' can vary in that given syntactic slot.

- [35S] O Padre João/`AGENT-GENERIC` foi quem casou o Pedro/`PATIENT` com a Ana/`CO-PATIENT` (Priest John married Peter with Ana)

In reflexive (or pronominal) passive constructions, the SR is kept:

- [36S1] A farinha/`OBJECT-GENERIC` misturou-se com o açúcar/`CO-OBJECT` (The flour mixed with the sugar)

In some cases, a facultative reflexive pronominal construction can be observed; the reflexive pronoun is then to be ignored as far as SRL is concerned, since it does not correspond to an argument proper for that predicate:

- [35S] O Pedro/`AGENT-GENERIC` casou(-se) com a Ana/`CO-AGENT` (Peter married himself with Ana)

The most common symmetrical SR are `CO-AGENT`, `CO-PATIENT`, `CO-OBJECT` and `CO-EXPERIENCER`, however some symmetrical predicates determine arguments of other semantic nature, like the verbs *coincidir* (coincide), involving the SR `OCCURRENCE` and *confronter* (face), with a `LOCATIVE-PLACE` argument:

35S *A vinda do Pedro/`OCCURRENCE` coincide com a visita da Ana/`CO-OCCURRENCE`* (Peter's coming coincide with Ana's visit)

35S *O terreno/`LOCATIVE-PLACE` A confronta com o terreno B/`CO-LOCATIVE`* (The terrain A faces/makes common boundaries with terrain B)

Finally, and unlike other SR, the symmetrical roles have not been subdivided into more specific subtypes. Table 3.5 resumes the symmetrical semantic roles defined so far.

Table 3.5: Semantic Roles for Portuguese: a preliminary proposal. 5 - Symmetric arguments

Sem. Role	**definition**	**Example (translation)**
`CO-AGENT`	Can change places with the `AGENT`	*O Pedro casou com* ***a Ana***
(`co-agent`)	without changing the meaning of the sentence	(Peter married Ana)
`CO-PATIENT`	Can change places with the `PATIENT`	*O padre casou o* ***Pedro e a Ana***
(`co-patient`)	without changing the meaning of the sentence	(The priest married Peter and Ana)
`CO-OBJECT`	Can change places with the `OBJECT-GENERIC`	*A farinha misturou-se com* ***o açucar***
(`co-object`)	without changing the meaning of the sentence	(The flour mixed with the sugar)
`CO-EXPERIENCER`	Can change places with the `EXPERIENCER`	*As opiniões do Pedro* ***e da Ana*** *convergem.*
(`co-experiencer`)	without changing the meaning of the sentence	(Peter and Ana's opinions converge)
`CO-OCCURRENCE`	Can change places with the `OCCURRENCE`	*A vinda do Pedro coincide com* ***a vinda dela***
(`co-occurrence`)	without changing the meaning of the sentence	(Peter's arrival coincides with her arrival)
`CO-LOCATIVE`	Can change places with the `LOCATIVEs`	*A casa do Rui confronta com* ***a casa da Ana***
(`co-locative`)	without changing the meaning of the sentence	(Rui's house borders Ana's house)

3.18 Summary

In this section, we went through the list of SR and the resulting 37 SRs for labeling the verbal arguments.

In the next chapter, we will present the natural language processing chain STRING (Mamede *et al.* 2012) in which we intend to implement the set of semantic roles described above, by building a rule-based SRL module for European Portuguese full (or lexical, or distributional) verbs.

Chapter 4

Semantic role labeling

4.1 STRING architecture

In this section we present the NLP processing chain STRING (Mamede *et al.*, 2012), in which the SRL module will be implemented. Figure 4.1 illustrates the STRING current architecture.

Figure 4.1: XIP architecture

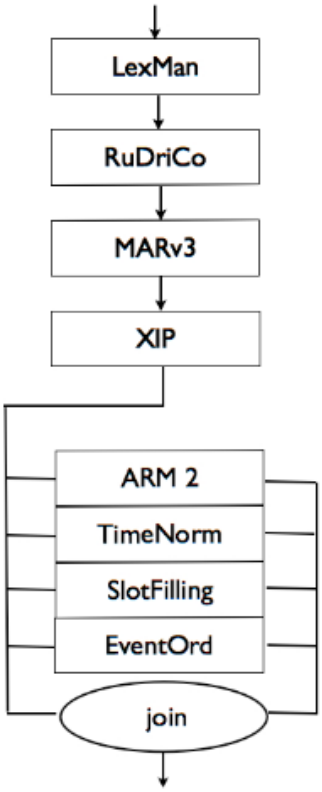
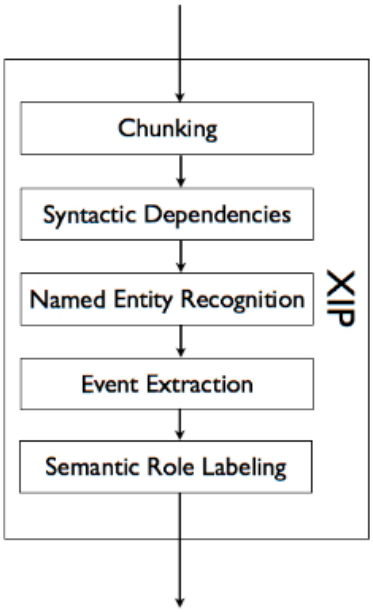


Figure 4.2: STRING architecture



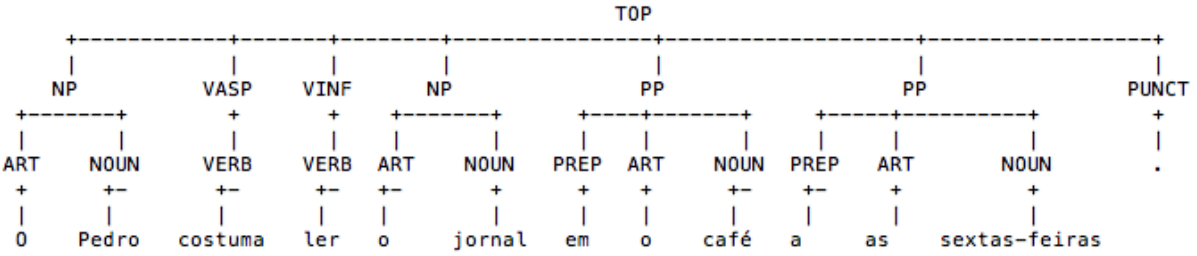
The STRING is composed of a POS tagger, LexMan (Vicente, 2013 [64]), a rule-based disambiguator, RuDriCo (Diniz *et al.*, 2010 [26]), and a machine learning-based disambiguator, MARv (Ribeiro, 2003 [58][1]). The parsing task runs after all the POS tagging and disambiguation is done, and it is performed by the Xerox Incremental Parser (XIP) (Ait-Mokhtar *et al.*, 2002 [3]). Its architecture is represented in figure 4.2. This is a finite-state, rule-based system, responsible for the extraction of elementary constituents (or chunks) and the syntactic dependencies between them. In other words, the system first identifies the elementary constituents (`NP` noun phrases; `PP` prepositional phrases; `AP` adjectival phrases, `ADVP` adverbial phrases, *etc.*) and then these are structured by binary dependencies between them, corresponding to the syntactic relations, such as `SUBJ` (subject), `CDIR` (direct complement), `MOD` (modifier), *etc.* Thus, for example, in the sentence:

(72) *O Pedro costuma ler o jornal no café às sextas-feiras*
'Peter normally reads the newspaper at the coffee shop on fridays'

the system begins by delimitating several chunks (or elementary constituents) and then it extracts the following dependency relationships which are sequentially structured in a shallow chunking tree (fig. 4.3). In the bottom line of fig. 4.4, we can see the various noun phrases (`NP`) and prepositional phrases (`PP`), as well as the two chunks that form the chain of the auxiliary aspectual verb (`VASP`) and the main verb, which is in the infinitive, forming an infinitive verbal chunk (`VINF`). The main syntactic dependencies extracted by the system consist of: the verbal chain (`VDOMAIN`) connecting the first auxiliary to the main verb; the relation between the subject and the main verb (`SUBJ`); the dependency between the verb and the direct complement (`CDIR`), and the two circumstantial complements (`MOD`); two named entities (`NE`) are also extracted and classified: *Pedro* as a `PEOPLE-INDIVIDUAL` and *às sextas-feiras* as a time expression, `TIME-CALENDAR`.

[1]Recently, Mamede et al. 2012 [45], the MARv has undergone extensive retraining, the training corpus was extensively revised and corrected and its results significantly improved.

Figure 4.3: STRING architecture




```
VDOMAIN(costuma,ler)
MOD-POST(ler,café)
MOD-POST(ler,sextas-feiras)
SUBJ-PRE(ler,Pedro)
CDIR-POST(ler,jornal)
NE-PEOPLE-INDIVIDUAL(Pedro)
NE-TIME-CALENDAR(a as sextas-feiras)
TOP{NP{O Pedro} VASP{costuma} VINF{ler} NP{o jornal} PP{em o café} PP{a as
sextas-feiras}.
```

Figure 4.4: Sintactic dependencies and chunking analysis performed by STRING

The semantic role labeling module to be developed in this project will be based on the syntactic dependency structure and linguistic information already available in the system. This information will be used to extract a new level of relations between constituents - the semantic roles. Considering the previous example, it is intended that the system is capable of extracting the semantic relationships between the various constituents that play the semantic roles of `AGENT`, `OBJECT`, `PLACE`, `TIME`, *etc.*, and associate them to the predicate on which they depend. Figure 4.5 illustrates one possible representation of this second set of dependencies to be extracted, adapting the notation proposed by Baptista *et al.* (2012, p. 4 [13])

```
EVENT-LEX(ler,outro)
EVENT_OTHER(ler)
EVENT_AGENT-GENERIC(ler,Pedro)
EVENT_OBJECT-GENERIC(ler,jornal)
EVENT_LOCATIVE-PLACE(ler,café)
EVENT_TIME-FREQUENCY(ler,a as sextas-feiras)
```

Figure 4.5: Semantic role extraction

In this representation, the semantic relationships are associated with events. These are captured by the dependency `EVENT`, to which several sub-dependencies of `EVENT` are then annotated. Each sub-dependency associates the core element of the event (`EVENT-LEX`), in this case the verb *ler* (to read), to the various elements of the sentence that depend on it, giving them their respective semantic role.

In the case of the core element, the feature `OTHER` captures the central element of the event and the

fact that this is not a *standardized* `EVENT` [2], respectively. Below, the semantic roles attributed are shown. First, the semantic relation of `AGENT-GENERIC` between the operator *ler* and the argument *Pedro*, and then the relations of `LOCATIVE-PLACE` between the operator *ler* and the argument *café* and between the operator *ler* and the argument *a_ as sextas-feiras*.

4.2 Semantic role annotation in ViPEr

ViPEr [11] is the lexicon-grammar of European Portuguese full (or lexical/distributional) verbs. It is a database in tabular format, that features 6,330 lines, corresponding to the full verbs' senses or constructions and 112 columns, indicating the corresponding syntactic, semantic and transformational properties.

The full list of 6,330 ViPEr verbal entries was manually annotated for the SR in each of their syntactic slots (subject and all other complement positions; only essential arguments are considered in ViPEr). A total number of 13,201 syntactic slots were manually classified for their SR.

For example, the ambiguous verb *cheirar* (to smell), appears in 4 ViPEr classes, and each verb entry may show different sets of SRs:

- in class 32C: *O Pedro cheirou a flor* (Peter smelled the flower), where it was given the SR features `experiencer-gen` (subject) and `object-gen` (direct complement);
- in verb class 33MV: *A flor cheira bem* (The flower smells good), where it has the features `object-gen` (subject) and `manner` (manner complement);
- in class 05: *Cheira-me que a Ana o sabe* (I think that Ana knows it), where it was given the SR features `object-f` (sentential subject) and `experiencer-gen` (indirect complement); and
- in class 33: *A Ana cheira a rosas* (Ana smells of roses), where it was given the SR features `object-gen` (subject) and `object-gen` (prepositional complement).

Table 4.1 shows the results of this classification procedure, namely the number of SRs per syntactic slot in all verb constructions of ViPEr. The distribution of the semantic roles per syntactic slot is very asymmetrical: 10 SR features cover almost 90% of the argument positions in ViPEr. The remaining 29 features cover the other 10%. It can also be seen that many SRs are residual (less than 5 occurrences) in ViPEr, such as `POSSESSOR` or `BENEFICIARY`. This does not mean that these SRs will not be found in real texts, eventually even with an expressive frequency.

[2] Standardized events are those that receive a specific treatment in view of several Information Retrieval/ Extraction tasks, such as `AGE`, `FAMILY`, *etc.* [13].

Table 4.1: Semantic Roles frequency in ViPEr verbs construction

Semantic Role	**N0**	**N1**	**N2**	**%**	**Cumul%**
agent-gen	4,248	0	0	0.3218	0.3218
object-gen	392	2,682	164	0.2453	0.5671
patient	0	899	18	0.0695	0.6365
experiencer-gen	522	373	0	0.0678	0.7043
object-f	37	400	178	0.0466	0.7509
cause	536	1	0	0.0407	0.7916
locative-dest	0	165	242	0.0342	0.8258
locative-place	15	305	13	0.0252	0.8510
agent-speaker	281	0	0	0.0213	0.8723
object-cl	20	208	0	0.0173	0.8896
addressee	0	34	170	0.0155	0.9050
message	0	178	13	0.0145	0.9195
locative-source	0	29	131	0.0121	0.9316
occurrence	34	99	2	0.0102	0.9418
co-agent	0	106	20	0.0095	0.9514
recipient	4	1	106	0.0084	0.9598
co-object	0	12	75	0.0066	0.9664
agent-giver	60	0	6	0.0050	0.9714
agent-cause	62	0	0	0.0047	0.9761
object-q	0	52	2	0.0041	0.9802
locative-path	0	47	0	0.0036	0.9837
agent-creator	36	0	0	0.0027	0.9864
patient-object	0	32	0	0.0024	0.9889
experiencer-vol	26	0	0	0.0020	0.9908
locative-source-locative-dest	0	22	0	0.0017	0.9925
agent-object	13	5	0	0.0014	0.9939
agent-taker	18	0	0	0.0014	0.9952
topic	0	12	5	0.0013	0.9965
manner	0	8	1	0.0007	0.9972
co-patient	0	0	7	0.0005	0.9977
co-agent-co-object	0	1	5	0.0005	0.9982
time-duration	0	6	0	0.0005	0.9986
instrument	0	4	1	0.0004	0.9990
victim	0	0	4	0.0003	0.9993
co-experiencer	0	3	0	0.0002	0.9995
time-calendar	1	0	1	0.0002	0.9997
beneficiary	0	1	0	0.0001	0.9998
co-locative	0	1	0	0.0001	0.9998
co-occurrence	0	1	0	0.0001	0.9999
possessor	0	1	0	0.0001	1.0000
Total	**6,305**	**5,688**	**1,164**	**1.0000**	**1.0000**

Some compound features like *agent-cause* or *agent-object* are used when a syntactic slot can be filled in by a human or non-human element, thus implying one of those two SRs, depending on the specific distributional class of that element. The semantic role attribution is made by rules that match the distributional class of the element to the appropriate SR.

An example of an ambiguous syntactic slot is the subject of the verb *irritar*, that has the feature `SR-N0-agent-cause` feature, in order to cover the two situation in the following examples:

(73) *O Pedro irrita a Ana*
'Peter irritates Ana'

(74) *O artigo do jornal irritou a Ana*
'The newspaper article irritates Ana'

A disambiguation rule applies the `AGENT-GENERIC` SR, to the first sentence (73), if it matches the subject distributional feature as a human. A similar rule exists for the opposite case.

As said before, a macro-SR is a construct that represents the set of all SRs of a certain type (*e.g.* `AGENT-x` includes all SRs that have an agentive nature). Table 4.2 shows the representativity of the macro-SRs in the entire set of 13,201 semantic roles that have been encoded in ViPEr.

Table 4.2: Macro semantic roles

Semantic Role	N0	N1	N2	total	% SR
agent-x	4,718	5	6	4,729	35.82
object-x	462	3,379	344	4,185	31.70
locative-x	15	568	386	969	7.34
patient-x	0	931	18	949	7.18
experiencer-x	548	373	0	921	6.98
co-x	0	124	107	231	1.74
time-x	1	6	1	8	0.06

The complex features `agent-cause`, `agent-object` and `patient-object`, were introduced in the `AGENT-x` and `PATIENT-x`, respectively. Symmetric complements were not introduced in the corresponding SR and were grouped together as `CO-x`.

The most frequent macro-SR in ViPEr is `AGENT-x`, immediately followed by `OBJECT-x`. The next most frequent macro-SR, with approximately 7% of the total SR tags are, in decreasing order: `LOCATIVE-x`, `PATIENT-x`, and `EXPERIENCER-x`. Symmetrical SR amount to little less than 2% of the entire set of tags.

In spite of the breakdown of the semantic role of `AGENT` into several subtypes, the more generic `AGENT` SR (`AGENT-GENERIC`) is significantly more representative on ViPEr (4,248 instances) than all the remaining, more specific `AGENT` SRs (481 instances).

This is not to say that the more specific SRs were not relevant to capture certain semantic relations expressed by those predicates that feature them, nor that their frequency in texts may prove to be significant.

Next, we present the syntactic-slot/semantic role combinations. Table 4.3 presents the 10 most frequent argument/SR combinations in ViPEr, from 177 different combinations (Full list of SR combinations in C.3).

In Table 4.3 it is possible to observe that the combination `SR-N0-agent-gen SR-N1-object-gen` is, by far, the most frequent construction, with 1,847 occurrences out of 6,632 (0,29%). In the top 10 constructions, the N0 position is 70% of the times an `AGENT`.

The next most productive configuration is `SR-N0-agent-gen-SR-N1-patient`. Notice the two intransitive construction with `OBJECT` (5%) or `AGENT` (4%) subjects, and the `LOCATIVE` structures. 331 SR combinations occur 10 or less times, and 67 only once.

Table 4.3: Semantic roles combinations frequencies

Frequency	**Semantic Roles combinations**	**%**
1847	SR-N0-agent-gen SR-N1-object-gen	0.29
568	SR-N0-agent-gen SR-N1-patient	0.09
343	SR-N0-experiencer-generic SR-N1-object-f	0.05
330	SR-N0-cause SR-N1-experiencer-generic	0.05
289	SR-N0-object-gen	0.05
271	SR-N0-agent-gen SR-N1-locative-place	0.05
233	SR-N0-agent-gen	0.04
198	SR-N0-agent-gen SR-N1-object-cl	0.04
169	SR-N0-agent-speaker SR-N1-message SR-N2-addressee	0.03
151	SR-N0-agent-gen SR-N1-object-gen SR-N2-location-destination	0.03

4.3 Semantic Role Labeling module

In this section we present the structure of the semantic role labeling module.

The SRL module takes place at the end of the parsing stage and it is performed by XIP.

By this stage:

- The verbs in the texts that are to be analyzed by the SRL module already have been previously disambiguated and given the information about the ViPEr class to which they belong ([63]). In other words, the word sense of the verb has been disambiguated, each sense of a polysemous verb corresponding to a different class in the Portuguese lexicon-grammar of full (or distributional) verbs. An external function (.kif), on XIP, performs this step. This verb sense disambiguator module combines a rule-based and a machine-learning based approach.

- The Named Entity Recognition (NER) module has already identified temporal expressions ([48]), in particular those of the type calendar (dates), duration and frequency. The generic temporal NE is to be treated as any other ordinary NP.

- A previous module for event identification extracts normalized events as described in [13]. These normalized events are domain and application-specific events while the SRL module here presented is to act upon a general-purpose event extraction module . This step is also performed in XIP.

4.3.1 Features

At this point, in order to correctly ascribe the corresponding semantic roles to their arguments, a set of features has to be attributed to the verbs that have been previously disambiguated. These features are automatically extracted from the information on the verbs' argument requirements, as they are defined on ViPEr.

This information consists of:

- A set of features that associate each syntactic slot (N0, N1, N2, N3) to a semantic role (`AGENT-GENERIC`, `OBJECT-GENERIC`, etc.). For example:

SR-N0-agent-gen for subject position
SR-N1-object-gen for 1st complement
SR-N2-locative-place for 2nd complement

(A complete list of these features concerning the SR of verb arguments is provided in appendix B)

- A feature indicating whether a N1 argument should have been parsed (and the corresponding syntactic dependency has been extracted) as a direct complement (CDIR) or a modifier (MOD) of a verb. This information is important because the previous information on the SR of a given syntactic slot is insufficient for the attribution of SR. In fact, the specific semantic role depends on the syntactic function of that constituent, and the N1 argument can either be a direct complement (parsed and identified by the dependency CDIR), or, if it is a prepositional phrase, be extracted by the dependency MOD.

For example, the verb *bater* (to hit) may have as its first complement either an `OBJECT-GENERIC`:

(75) *O Pedro/`AGENT-GENERIC` bateu as claras/`OBJECT-GENERIC`*
'Peter hit the egg whites'

(76) *O Pedro/`AGENT-GENERIC` bateu no assaltante/`PATIENT`*
'Peter hit the assailant'

The rules below extract the two different types of N1 shown in the examples above.

```
if (CDIR(#1[SR-N1-object-gen],#2) & EVENT[other](#1)
   & ~EVENT(#1,#2) & ~EVENT[object-generic](#1,?) )
         EVENT[object-generic=+](#1,#2).

if (MOD(#1[SR-N1-patient],#2[UMB-Human]) & PREPD(#2,?[lemma:em]) & EVENT[other](#1)
   & ~EVENT(#1,#2) & ~EVENT[patient](#1,?) )
         EVENT[patient=+](#1,#2).
```

The first rule reads:

If a verb (marked as variable #1) has the feature `SR-N1-object-gen`; and if it is an EVENT[other]; and if there is not already a relation of event between #1 and a variable #2; then, the semantic role `OBJECT-GENERIC` is attributed to the new dependency EVENT holding between the verb (#1) and the corresponding direct complement (#2).

The second rule reads:

If the second argument of the MOD (variable #2) has the feature UMB-Human (umbrella tag that represents any human noun in the grammar); if the verb has the feature `SR-N1-patient`; and if there is a preposition *em* introducing #2; if it is an EVENT[other]; and if there is not already a relation of EVENT between #1 and #2; then, the semantic role `PATIENT` is attributed to the dependency EVENT holding between the verb (#1) and the corresponding modifier (#2).

- The distributional features of the arguments (basically, human and non-human), in order to better identify the arguments of verbs. This information is relevant for the cases where two or more MODs are involved in the verb's construction and it is necessary to distinguish which is the constituent whose semantic role is determined by the verb.

(77) *O Pedro abdicou do trono*
'Peter declared that that was important'

```
"abdicar-34":"SR-N0-experiencer-gen SR-N1-de SR-N1-object-gen SR-N1-nhum ‘‘
```

(78) *O Pedro acertou na resposta*
'Peter got the answer right'

```
"acertar-35R":"SR-N0-agent-gen SR-N1-em SR-N1-object-gen SR-N1-hum SR-N1-nhum ‘‘
```

In this set of features, the prepositions introducing the argument PPs will also be encoded, in order to better identify which constituents extracted as MOD fulfill the distributional constraints of the verb.

```
if (MOD(#1[SR-N1-object-gen],#2) & PREPD(#2,?[lemma:de]) & EVENT[other](#1)
   & ~EVENT(#1,#2) & ~EVENT[object-generic](#1,?) )
        EVENT[object-generic=+](#1,#2).

if (MOD(#1[SR-N1-object-gen],#2) & PREPD(#2,?[lemma:em]) & EVENT[other](#1)
   & ~EVENT(#1,#2) & ~EVENT[object-generic](#1,?) )
        EVENT[object-generic=+](#1,#2).
```

Figure 4.6: Preposition information usage examples

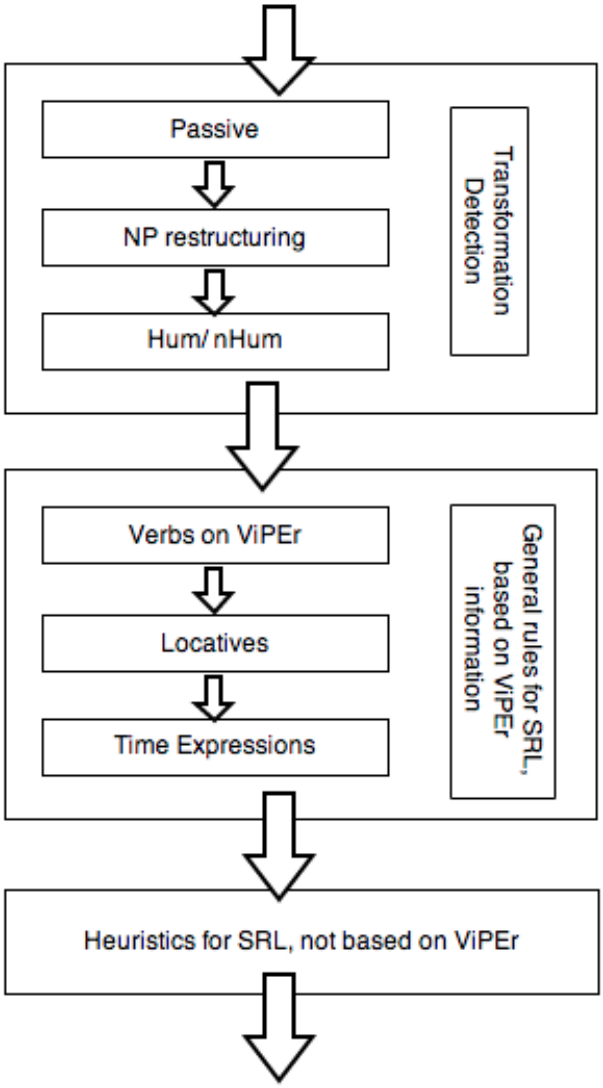
The set of features is automatically extracted from ViPEr and constitutes a lexicon file that is then ran by XIP on the input file. These features associate the set of properties to each verb instance, based on the ViPEr class to which all verb forms have been unambiguously classified.

The following examples illustrate some of the entries of the verb *abafar* (to suffocate[1], to make cozy[2], to steal[3], to make unnoticed[4]) this lexicon file, as defined for verb:

- `"abafar-31H":"SR-N0-experiencer-gen SR-N0-Hum",`
- `"abafar-32C":"SR-N0-agent-gen SR-N0-Hum SR-N0-nHum SR-N1-object-gen SR-N1-nHum SR-N1-cdir",`
- `"abafar-36DT":"SR-N0-agent-gen SR-N1-CDIR SR-N1-object-gen SR-N1-nhum SR-N2-a SR-N2-victim SR-N2-hum SR-pass-ser ",`
- `"abafar-32R":"SR-N0-agent-gen SR-N0-Hum SR-N1-object-gen SR-N1-nHum SR-N1-cdir SR-pass-ser",`

Once all this information has been associated to the verbal instances, the SRL module enters into action. This is to be done in several stages, which are briefly described and illustrated in figure 4.7:

Figure 4.7: Architecture of the SRL module



4.3.2 Transformations detection

Firstly, the SRL module recognizes some sentence constructions, resulting for the transformation of the canonic form represented in ViPEr, which may hinder the SRL task, namely:

`CAUSE/AGENT-GENERIC` alternation, depending on the Human and non-Human feature in a given syntactic slot (mostly subject), for some verbs, as in examples 79 and 80:

(79) *A Joana/`AGENT-GENERIC` partiu o vidro/`OBJECT-GENERIC`*

'Joana broke the glass'

(80) *O terramoto/**CAUSE** partiu o vidro/**OBJECT-GENERIC***
'The earthquake broke the glass'

In this case, the verb *partir* (to brake, to shatter) from class 32C had been attributed the features `SR-N0-agent-cause, N0-Hum and N0-nHum`. The following generic rules (figure 4.8) will extract the appropriate SR feature depending on the semantic features attributed to the head noun of the subject NP.

```
if (SUBJ(#1[SR-N0-agent-cause],#2[UMB-Human]) & EVENT[other](#1) & ~EVENT(#1,#2) )
  EVENT[agent-generic=+](#1,#2).

if (SUBJ(#1[SR-N0-agent-cause],#2) & EVENT[other](#1) & ~EVENT(#1,#2) )
  EVENT[cause=+](#1,#2).
```

Figure 4.8: Agent-Cause disambiguation

- Passive constructions: As the information encoded in ViPEr represents the semantic roles of the verbs' arguments in their canonical order, which corresponds to the active construction, the passive sentences must be identified, so that the adequate SR may be attributed to the arguments that have been moved from their basic syntactic position.

Rules, like the following, are ran in previous stages of the STRING, that identify the passive constructions:

```
|pastpart#2[pass-ser=+]|
if ( VLINK(?[lemma:ser],#2[pastpart]) )
```

Figure 4.9: Passive identification

Using this identification, it is possible to interchange the SRs from N0 to N1 and from N1 to N0 respectively, when the label `SR-transf-passiva` is found, as shown in the rule below. The rule reads: if the labels `SR-transf-passiva, SR-N1-object-gen` and `SR-N0-agent-gen` are found, then the labels `SR-N0-object-gen, SR-N1-agent-gen` are added, and the labels `SR-N1-object-gen, SR-N0-agent-gen` are taken out.

```
|#1[transf-passiva, SR-N1-object-gen, SR-N0-agent-gen, SR-N0-object-gen=+, SR-N1-agent-gen=+,
SR-N1-object-gen=~, SR-N0-agent-gen=~ ]|
  if(#1[SR-transf-passiva]) ~
```

After this rule, the `AGENT-GENERIC` and `OBJECT-GENERIC` are attributed do the N0 and N1 respectively, using the following rules:

```
if (SUBJ(#1[SR-N0-object-gen],#2) & EVENT[other](#1) & ~EVENT(#1,#2) )
  EVENT[object-generic=+](#1,#2).

if (^MOD[post=+](#1[SR-N1-agent-gen, SR-N1-por],#2) & EVENT[other](#1) & ~EVENT(#1,#2)
& PREPD(#2,#3[lemma:por]) )
  EVENT[agent-generic=+](#1,#2),
  COMPL[post=+](#1,#2).
```

Figure 4.10: Rules for SR changing

After this, the output of the STRING is:

```
MAIN(reservado)
DETD(carro,O)
DETD(Joao,o)
VDOMAIN(foi,reservado)
COMPL_POST(reservado,Joao)
SUBJ_PRE(reservado,carro)
NE_PEOPLE_INDIVIDUAL(Joao)
EVENT_LEX(reservado,outro)
EVENT_OTHER(reservado)
EVENT_OBJECT-GENERIC(reservado,carro)
EVENT_AGENT-GENERIC_POST(reservado,Joao)
0>TOP{NP{O carro} VCOP{foi} VCPART{reservado} PP{por o Joao} .}
```

Figure 4.11: Passive output

- Noun phrase restructuring ([38]) from a complex NP with a *Nbp* (body part noun) head and a determinative complement de *Hum*.

(81) *Os pés/OBJECT-CL do Pedro/PATIENT suam*
'Peter's feet transpire'

(82) *O Pedro/PATIENT sua dos pés/OBJECT-CL*
'Peter transpires from its feet'

This transformation leaves the *Hum* as the verb's subject and moves the *Nbp* to a PP complement with a locative value; the preposition is lexically determined by the verb. Since the restructured sentence (82) is regularly derived from (81), the later has been considered the base sentence (31CL). However, this implies that the restructured form be matched, so that the PP is not later marked as any ordinary MODifier.

The examples above were taken from class 31CL, of intransitive constructions. This transformation also takes place on class 32CL ([7],[11],[12]), where it affects the direct complement.

4.3.3 General rules for SRL, based on ViPEr information

On a third stage, general rules have to be made in order to mark the N0, N1 and N2 of sentences. An example of these, is the rule for `AGENT-GENERIC` as N0:

```
if (subj(#1[SR-N0-agent-gen],#2) & EVENT[other](#1) & ~EVENT(#1,#2) )
EVENT[agent-generic=+](#1,#2)
```

The explanation of this rule is: if the subject of the sentence is followed by a verb with the characteristic `SR-N0-agent-gen`; and if it is an OTHER EVENT; and if a semantic role has not been applied to the argument yet; then the output is `AGENT-GENERIC`.

Many verbs in ViPEr allow both *Hum* and *nHum*. In order to capture the correct role of the syntactic slots in ViPEr, complex features such as `SR-N1-patient-object` were created, which are then disambiguated using the distributional information of the arguments. This way, the correct SR of the constituent is captured using rules that disambiguate, for example, this features:

```
if (CDIR(#1[SR-N1-patient-object],#2[UMB-Human]) & EVENT[other](#1) & ~EVENT(#1,#2) )
  EVENT[patient=+](#1,#2).

if (CDIR(#1[SR-N1-patient-object],#2) & EVENT[other](#1) & ~EVENT(#1,#2) )
  EVENT[object-generic=+](#1,#2).
```

Figure 4.12: Rules for disambiguating SR-N1-patient-object

The first rule reads: if the second argument of the CDIR dependency has the feature UMB-Human; if the verb has the feature `SR-N1-patient-object` and; if it is an EVENT[other] and; if there is not already an EVENT relation of event between #1 and #2; then, the feature `PATIENT` is attributed to the dependency EVENT holding between the verb (#1) and the corresponding CDIR. (#2).

The second rule reads: if the head of the CDIR has no features associated; if the verb has the feature `SR-N1-patient-object` and; if it is an EVENT[other] and; if there is not already a relation of event between #1 and #2; then, the feature `OBJECT-GENERIC` is attributed to the dependency EVENT holding between the verb (#1) and the corresponding MOD (#2).

Since this rule is applied after the previous one, all cases with human CDIR have already been captured, so that only the non-human remain; this way the semantic role of `OBJECT-GENERIC` is attributed.

There are 8 such dichotomy cases in the rules, in a set of 158 rules.

The locative semantic roles can be derived from the preposition introducing the argument complements with the locative features `geo`, `UMB-Location` and `sem-cc`, as shown below:

de → `locative-source`
para → `locative-destination`
em → `locative-place`
por, através → `locative-path`

A list of most common portuguese prepositions and their semantic features exists already and all the prepositions are marked according to its semantic features. The features marked on each preposition (`preplocsource, preplocdestination, preplocplace, preplocpath`, etc.) allow us to build the following rules:

```
if ( (MOD(#1,#2[geo]) || MOD(#1,#2[UMB-location]) || MOD(#1,#2[sem-cc])) & PREPD(#2,?[preplocsource])
 & EVENT[other](#1) & ~EVENT(#1,#2) )
   EVENT[locative-source=+](#1,#2).

if ( (MOD(#1,#2[geo]) || MOD(#1,#2[UMB-location]) || MOD(#1,#2[sem-cc])) & PREPD(#2,?[preplocdest])
 & EVENT[other](#1) & ~EVENT(#1,#2) )
   EVENT[locative-destination=+](#1,#2).

if ( (MOD(#1,#2[geo]) || MOD(#1,#2[UMB-location]) || MOD(#1,#2[sem-cc])) & PREPD(#2,?[preplocpath])
 & EVENT[other](#1) & ~EVENT(#1,#2) )
   EVENT[locative-path=+](#1,#2).

if ( (MOD(#1,#2[geo]) || MOD(#1,#2[UMB-location]) || MOD(#1,#2[sem-cc])) & PREPD(#2,?[preplocplace])
 & EVENT[other](#1) & ~EVENT(#1,#2) )
   EVENT[locative-place=+](#1,#2).
```

Figure 4.13: General rules for locative semantic roles

where, in the `preplocsource` case: if the head of the MOD has the features `geo` for geographical locations like a city, `UMB-location` for place complements like an entrance, or a `sem-cc` for concrete countable objects like a bed; and if the head of the `PREPD` has the feature `preplocsource` and; if it is an EVENT[other]; and if there is not already a relation of event between #1 and #2; then, the feature `LOCATIVE-SOURCE` is attributed to newly created dependency EVENT holding between the verb (#1) and the corresponding modifier (#2).

The same occurs for the other types of locative prepositions.

4.3.4 Heuristics for SRL, not based on ViPEr

At last, a set of heuristic rules were written in order to capture the verbs that are not on ViPEr, therefore not having the linguistic information taken from ViPEr, needed to fire the rules before. These heuristics mark the arguments of such verbs with the corresponding semantic roles.

An example of these rules is:

```
   if ( SUBJ(#1,#2[human]) & EVENT[other](#1) & ~EVENT(#1,#2) )
EVENT[agent-generic=+](#1,#2)
```

This rule reads: if the subject (#2) of the verb (#1) has the human feature; if it the verb has already been captured as an EVENT[other]; and if a semantic role has not been associated to the EVENT relation between that verb and that argument yet; then, the feature `AGENT-GENERIC` is attributed to the dependency EVENT holding between the verb (#1) and the corresponding argument (#2).

In total, 11 pattern-matching rules were written for the passive identification and role changing, 133 for the semantic role labeling, 16 for the identification of time, location and manner complements and finally, 7 pattern-matching rules for the identification of arguments of verbs that are not on ViPEr, totaling 167 pattern-matching rules and heuristics.

The semantic role labeling module is a simple text file, easily editable, containing all the pattern matching rules, an example for each rule and the expected output. The code for this module can be found in appendix E.

Chapter 5

Evaluation

5.1 Corpus

To evaluate the system, a fragment of the PAROLE project [51] corpus was used. This fragment has been retrieved from a part of the full corpus PAROLE, containing about 250 thousand words, which had already been automatically annotated and manually corrected for parts-of-speech tagging [58]. Later on, it has undergone profound manual revision in view of the development of the new STRING POS-tagger [25] and further enriched with anaphora information [47] and the ViPEr verb class information [11].

This fragment (miscelanea) consists of 26 short texts of very diverse nature, including news on a large variety of topics: stock exchange, economy, society, national and international politics, minutes of meetings, medication leaflets, letters, opinion articles and scientific dissemination articles.

The fragment of the corpus selected for SR annotation contains 7,834 (2,587 different) words. Of these, 896 are verbs, distributed by 331 different lemmas. Table 5.1 presents the breakdown of these verbs in different types. From these, 511 (57%) are full (lexical, or distributional) verbs, and were provided with their ViPEr classification, manually corrected; 154 were support verbs (17%), 124 copula verbs (14%), and 82 verbal auxiliaries (9%). There were also 20 operator-verbs and 3 frozen sentences (idioms).[1]

Table 5.1: Frequencies of verbs constructions

Freq.	**Verb type**	**%**
511	Vfull	57%
154	Vsup	17%
124	Vcop	14%
82	AuxV	9%
20	Vop	2.2%
3	Idiom	0.3%
1	Noviperdata	0.1%
896	-	100%

[1] A single English verb form (*explaining*) was tagged as having no ViPEr data. For the definition of support and operator verbs, see [36].

Table 5.2 present the 10 most frequent lemmas and their count.

The auxiliary verbs represent the great majority of the most frequent verbs in the corpus used.

Table 5.2: Most frequent lemmas

Verb	**Count**	**%**
ser	101	0.11
ter	40	0.04
ir	27	0.03
estar	25	0.03
fazer	19	0.02
dar	13	0.01
realizar	13	0.01
dizer	12	0.01
ficar	11	0.01
haver	11	0.01

Table 5.3 presents the 10 most frequent ViPEr verb classes, their count, and their percentage considering the 511 full verbs on the corpus.

Table 5.3: Most frequent ViPEr verb classes

Class	**Count**	**%**
6	83	0.16
9	71	0.14
32C	64	0.13
36DT	25	0.05
38LD	22	0.04
01T	20	0.04
32R	20	0.04
35LD	20	0.04
32H	17	0.03
11	12	0.02

Table 5.4 present the lemmas' distribution. Notice that 177 verbs (lemmas) are *hapax legomena*, representing 19,75% of all verbs.

Table 5.4: Lemmas distribution

Freq.	Count	Total
101	1	101
40	1	40
27	1	27
25	1	25
19	1	19
13	2	26
12	1	12
11	3	33
10	3	30
7	2	14
6	1	6
5	10	50
4	24	96
3	34	102
2	69	138
1	177	177
Total	331	896

It is our conviction that the variety of textual genres, the number of verbal instances and particularly the number and the diversity of full verbs' lemmas makes this small fragment of the corpus sufficient for the evaluation of the SRL module.

5.2 Corpus annotation

Once the evaluation corpus had been defined, we set out to annotate the semantic roles holding between all full verbs and their arguments. Thus, idioms, (tense, modality and aspect) auxiliary verbs, support and operator verbs were excluded from the annotation campaign.

The annotation was carried out using the Glozz annotation tool [65], using the STRING tokenization output as annotation units (univocally indicated by an integer in each text). Figure 5.1 shows the appearance of a small portion of text being annotated. Different colors allows to distinguish each POS and facilitates the annotation process. The arrows are annotations indicating the relations between text units. In this case, two types of relations are already annotated in the corpus: anaphoric relations and semantic roles. In this small excerpt there is only one anaphoric relation, all other arrows represent semantic roles.

The semantic roles are defined as relation 'r' with a label (the name of the SR) between two arguments: the first is always the verb and the second is the noun holding that semantic role with that verb. These arguments of the relation are indicated by the integer that corresponds to the annotation unit. Finally, each relation of figure 5.1 gets a unique identifier (ID). Figure 5.2 shows part of the set of SR annotated in small text. Figure 5.3 shows the Glozz output of the manually annotated corpus. The set of SR used for annotation is the same presented in section 3.

In order to prepare the annotation process, a small sample of sentences was first manually annotated and the problems discussed until it was possible to define a working set of annotation directives, explaining

the most relevant decisions concerning what was the target of the current annotation campaign, the exclusions, and some particular cases requiring special decisions. The full text of the annotation directives is provided in Appendix D.

The annotation proper was carried out along several working sessions, during June/July 2013, by the author of this dissertation and another linguist. The two worked together, discussing each problem as it appeared, closely following the directives and the SR definitions. The information on semantic roles encoded in ViPEr was actively ignored, though the database was consulted punctually, to check whether a particular construction had been included. As the annotation was carried out directly on the text, as shown in Figure 5.1, the annotators were "unaware" of the ViPEr class a verb instance had been given, and had to rely exclusively in the sense the verb presented (or the sense that it was possible to infer) in its context. Very occasionally, it was not possible to define the precise SR for a given verb-argument relation. In this case, one of following solutions was adopted: (i) a generic SR was chosen, instead of a more precise subtype; (ii) the most appropriate SR was chosen, once all non-relevant roles had been discarded; (iii) the relation was left untagged. The decisions were taken by mutual agreement. The entire annotation process took around 50 person-hours.

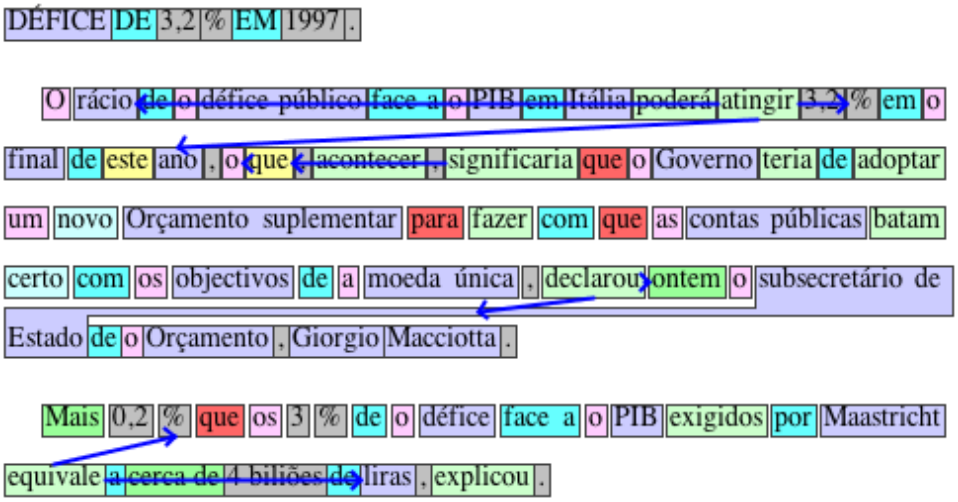


Figure 5.1: Glozz SR relations: the Gloss annotation layout, with different POS in different colors and relations as arrows linking annotation units.

```
r_OBJECT-Q(19,21) ID=96
r_OBJECT-GEN(19,9) ID=97
r_TIME-CALENDAR(19,27) ID=98
r_CAUSE(34,30) ID=99
r_AGENT-SPEAKER(59,62) ID=100
r_TIME-CALENDAR(59,60) ID=101
r_OBJECT-Q(86,91) ID=102
r_OBJECT-GEN(86,72) ID=103
```

Figure 5.2: Glozz SR relations: internal representation of the SR relations.

```
<relation type="OBJECT-Q">
  <participant>18</participant>
  <participant>20</participant>
</relation>
<relation type="OBJECT-GEN">
  <participant>18</participant>
  <participant>8</participant>
</relation>
<relation type="TIME-CALENDAR">
  <participant>18</participant>
  <participant>26</participant>
</relation>
<relation type="CAUSE">
  <participant>33</participant>
  <participant>29</participant>
</relation>
<relation type="AGENT-SPEAKER">
  <participant>58</participant>
  <participant>61</participant>
</relation>
<relation type="TIME-CALENDAR">
  <participant>58</participant>
  <participant>59</participant>
</relation>
<relation type="OBJECT-Q">
  <participant>85</participant>
  <participant>90</participant>
</relation>
<relation type="OBJECT-GEN">
  <participant>85</participant>
  <participant>71</participant>
</relation>
```

Figure 5.3: Glozz SR annotation: the XML output of the SR relations

In total, 655 SRs were annotated manually in the selected *miscelanea* corpus. Table 5.5 presents the breakdown of the semantic roles in the annotated corpus.

Many semantic roles, mostly symmetric SRs, were not found in the corpus. Some SR occur only once, such as `LOCATIVE-SOURCE` and `TIME-DURATION`. The most frequent SRs were, `OBJECT-F`, `OBJECT-GENERIC` and `AGENT-GENERIC`, with 115, 106 and 78 occurrences, respectively. `AGENT-SPEAKER` occurs 61 times, which can be explained by the fact that a large part of the corpus was journalistic, and this kind of text has many quotations and *verba dicendi* [9].

Table 5.5: Semantic role frequency in the annotated corpus

Semantic Role	**Occurrences**
ADDRESSEE	6
AGENT-CREATOR	3
AGENT-GENERIC	78
AGENT-GIVER	0
AGENT-SPEAKER	61
AGENT-TAKER	0
BENEFICIARY	3
CAUSE	24
CO-AGENT	3
CO-EXPERIENCER	0
CO-LOCATIVE	0
CO-OBJECT	3
CO-OCCURRENCE	0
CO-PATIENT	0
EXPERIENCER-GENERIC	19
EXPERIENCER-VOL	9
INSTRUMENT	9
LOCATIVE-DEST	7
LOCATIVE-PATH	0
LOCATIVE-PLACE	38
LOCATIVE-SOURCE	1
MANNER	7
MESSAGE	8
OBJECT-CL	0
OBJECT-F	114
OBJECT-GENERIC	105
OBJECT-Q	44
OCCURRENCE	25
PATIENT	15
POSSESSOR	6
RECIPIENT	1
TIME-CALENDAR	42
TIME-DURATION	1
TIME-FREQUENCY	3
TOPIC	19
VICTIM	1
Total	655

5.3 Results

The evaluation was performed automatically [2], and it compared the results from the automatic semantic role module with the manually annotated SRs. In a set of 655 SRs, the results retrieved on April 13th, 2013 are the following:

Correctly matched: 119;
Partially matched: 55;
Missed: 481.

Correctly matched instances correspond to the cases where the semantic role, the verb and its argument are exactly what has been encoded in the reference corpus. In partial matches, the verb is always correct, but either the SR or the verb argument is wrong. Missed instances correspond to the cases where a semantic role is in the reference but the system failed to assign it.

The evaluation took into account the evaluation metrics of *Precision* (P), *Recall* (R), *Accuracy* (A) and *F-measure* [57]. These can be defined as follows (TP: true positives; FP: false positives; TN: true negatives [3]; FN: false negatives):

- *Precision*,
$$P = \frac{TP}{TP + FP}$$
all the cases correctly marked (TP) divided by the sum of all marked cases (TP + FP);
- *Recall*,
$$R = \frac{TP}{TP + FN}$$
the sum of all the cases correctly marked divided by the sum of all the marked cases the system should have marked (TP + FN);
- *Accuracy*,
$$A = \frac{TP + TN}{TP + FP + FN + TN}$$
the sum of all cases cases correctly marked (TP) and the correctly unmarked cases (TN) divided by the sum of all cases under consideration.
- *F-measure*,
$$F = \frac{2.P.R}{P + R}$$
the harmonic mean between precision and recall, according to each one an equal weight.

Two types of evaluation can be performed with the results mentioned above: a *strict* or *relaxed* evaluation, depending on the weight accorded to the partial matches. In partial matches, three elements are in play: the semantic role, the verb and its argument; while the verb is always correct, either the relation name is wrong or the argument was incorrectly captured. We considered that, in a relaxed evaluation, at least two thirds of the solution were captured.

Therefore, in a strict evaluation, the partial matches count as false-positives (FP.) In the relaxed evaluation, the partial matches count as a true-positive (TP), but with a weight of two thirds of a correct match.

[2] The program used for the evaluation was built by Cláudio Diniz, as part of his work as research assistant at INESC-ID Lisboa. The author is grateful for his assistance in the use of the program.

[3] In this case, the true negatives were ignored (TN=0).

With this evaluation strategy, the results are those in Table 5.6:

Table 5.6: Results (on April 13th, 2014)

Measure	Strict	Relaxed
Precision	68.4%	76.4%
Recall	19.8%	19.8%
Accuracy	18.2%	18.7%
F-measure	30.1%	31.5%

5.4 Error analysis

The results are below what can be possible with this SRL module. As said before, this module runs at the very last steps of XIP. By that stage, the POS tagger, the POS disambiguator, the syntactic parser and the word (verb) sense disambiguator have been ran. Any errors that come from those previous steps impact on the results of the SRL module. We expect the results to improve specially when the verb sense disambiguator is perfected. Below, some examples of the errors found in the comparison between the manually and the automatically annotated corpora, are shown:

Consider the sentence:

(83) *O bar fecha às 3ªs (shortened for terças-feiras (tuesdays)*
'The bar closes on Tuesdays'

the expected result for *3ªs* is `TIME-CALENDAR`. Instead, the result is:

```
MAIN(fecha)
DETD(bar,O)
DETD(3as,as)
VDOMAIN(fecha,fecha)
MOD_POST(fecha,3as)
SUBJ_PRE(fecha,bar)
EVENT_LEX(fecha,outro)
EVENT_OTHER(fecha)
EVENT_AGENT-GENERIC(fecha,bar)
EVENT_OBJECT-GENERIC(fecha,3as)
```

Figure 5.4: XIP output

This happens because 3ªs (Tuesdays) was not identified as a time expression by the previous modules. Instead, it identified the constituent as a plain ordinal numeral. Hence, it was parsed as a ordinary PP, which made it impossible for the module to apply the rules that extract the semantic roles for time expressions.

When using the expressions *terças-feiras* or *3as* the output is:

```
EVENT_AGENT-GENERIC(fecha,bar)
EVENT_TIME-FREQUENCY(fecha,a as terças-feiras)

EVENT_AGENT-GENERIC(fecha,bar)
EVENT_TIME-CALENDAR(fecha,a as 3as)
```

The difference on the semantic role label depends on the information from the local grammar for time expressions which, in the future, will be perfected.

In the following example, *exposição* should have been captured as an `OBJECT-F`, and it is not captured at all:

(84) *A exposição nunca foi encarada com hostilidade*
'The exhibit was never seen with hostility'çã

```
MAIN(encarada) DETD(exposição,A) VDOMAIN(foi,encarada) MOD_POST(encarada,hostilidade)
MOD_PRE_NEG(encarada,nunca) SUBJ_PRE(encarada,exposição) PATIENT_PRE(encarada,exposição)
EVENT_LEX(encarada,outro) EVENT_OTHER(encarada) EVENT_LEX(exposicção,outro) EVENT_OTHER(exposição)
EVENT_LEX(hostilidade,outro) EVENT_OTHER(hostilidade) EVENT_MANNERSR(encarada,hostilidade)
```

The reason is that the verb has been wrongly disambiguated. The verb-sense disambiguator chose the ViPEr class 32H for the verb *encarar*, whereas the correct class in this context should have been the class 06. Notice, however, that the manner adverbial adjunct *com hostilidade* (with hostility) was adequately captured.

Another example is:

(85) *Esta foi capitaneada pelo Porto*
'This one was captained by Porto'

```
MAIN(capitaneada)
DETD(Porto,o)
VDOMAIN(seria,capitaneada)
COMPL_POST(capitaneada,Porto)
SUBJ_PRE(capitaneada,Esta)
NE_CITY_ADMIN_AREA_LOCATION(Porto)
EVENT_LEX(capitaneada,outro)
EVENT_OTHER(capitaneada)
```

where none of the arguments is labeled with the semantic roles. The XIP output for the verb is:

```
capitaneada(11-21)+[sr-n1-nhum:+,sr-n1-hum:+,sr-n1-por:+,sr-pass-ser:+,
sr-n1-agent-gen:+,sr-n0-patient:+,32h:+,markviper:+,pass-ser:+,
fem:+,sg:+,pastpart:+,hmmselection:+,last:+,first:+]
```

while the noun *Porto* (*id.* a football team), and the pronoun *Esta* (this$_{-fem.sg.}$):

```
Porto(28-32)+[enl2:+,city:+,admin_area:+,location:+,maj:+,proper:+,masc:+,
sg:+,noun:+,hmmselection:+,last:+]
```

```
Esta(0-3)+[proxdem:+,dem:+,maj:+,fem:+,sg:+,pron:+,hmmselection:+,start:+,last:+,first:+]
```

As can be seen, class 32H was chosen for the verb. Even though the passive has been recognized, since *Porto* was not labeled with the human feature, and the metonymy didn't work either [52],the `AGENT-GENERIC` semantic role was not attributed to N0. *Esta* was not attributed any semantic role because it is an anaphora that has not been previously solved by XIP. An anaphora resolution module is already being built (Marques 2013 [46]) to be integrated in STRING and that would be able to improve results in such cases.

During the whole process, ViPEr was perfected, with verb entries having been added, and with some changes in the ViPEr annotation.

5.5 Validation of the Semantic Role set

The semantic role set defined in 3 was validated by means of the annotation, by 6 annotators, of a small corpus of random sentences taken from the evaluation corpus (see 5.1). The corpus was divided into sentences (373) which were then randomized and filtered in order to retrieve those that were between 60 and 150 characters long, so that sentences were long enough, though not excessively long, to provide adequate context for the task of SRL. The arguments of the first full verb of each sentence were manually annotated by two linguists, to define a golden standard. At this stage we aimed at validating about 100–150 semantic roles' instances. All in all, 68 sentences (about 20%) of the corpus were selected, and a total of 113 semantic roles were annotated (68 in subject position, 34 in direct object/first complement position, 9 as second complement and 2 in third argument slot; 59 different verb lemmas and 16 ViPEr verbal classes. Table 5.7 shows the breakdown of SRs in this smaller semantic role set validation corpus, while the distribution of the semantic roles in the full corpus can be found in F.

Semantic Role	**Occurrences**	**%**
Addressee	2	0.02
Agent-generic	11	0.10
Agent-speaker	14	0.12
Agent-creator	1	0.01
Experiencer-generic	4	0.04
Experiencer-vol	3	0.03
Locative-destination	1	0.01
Locative-place	4	0.04
Manner	5	0.04
Object-generic	15	0.13
Object-f	31	0.27
Object-q	8	0.07
Occurrence	2	0.02
Patient	5	0.04
Time-Calendar	3	0.03
Topic	4	0.04
	113	**1.00**

Table 5.7: SR distribution in the small sample of sentences

	Classes	Occurrences	%
	01t	6	0.09
	06	10	0.15
	08	1	0.01
	09	21	0.31
	11	1	0.01
	13	1	0.01
	31R	1	0.01
	32A	1	0.01
	32C	6	0.09
	32H	5	0.07
	32R	1	0.01
	33MV	1	0.01
	33NM	7	0.10
	35LD	4	0.06
	36LD	1	0.01
	36TA	1	0.01
Total	**16**	**68**	**1.00**

Table 5.8: ViPEr classes occurrences in the evaluation corpus

The annotators were first instructed about the task to be performed with a simple sentence, such as the following example, and a link to the document defining the semantic roles set was provided.

(86) *Na frase "O jornalista do El País" vai buscar "A Capital" para dar a última estocada.": Indique a relação semântica entre os seguintes verbo e argumento: (buscar,jornalista).*
'In the sentence "The reporter of "El País" chooses "A Capital" to hit the final strike.": Indicate the semantic relation between the following verb and argument: (chooses,reporter).'

The validation process consisted in filling in a multiple choice form, where each annotator had to select, from a combo with all the 38 possible options, in alphabetical order, the correct semantic role for the arguments of a given verb. An extra option *Nenhuma das anteriores* (none of the above) was also provided. Figure 5.5, presents the form.

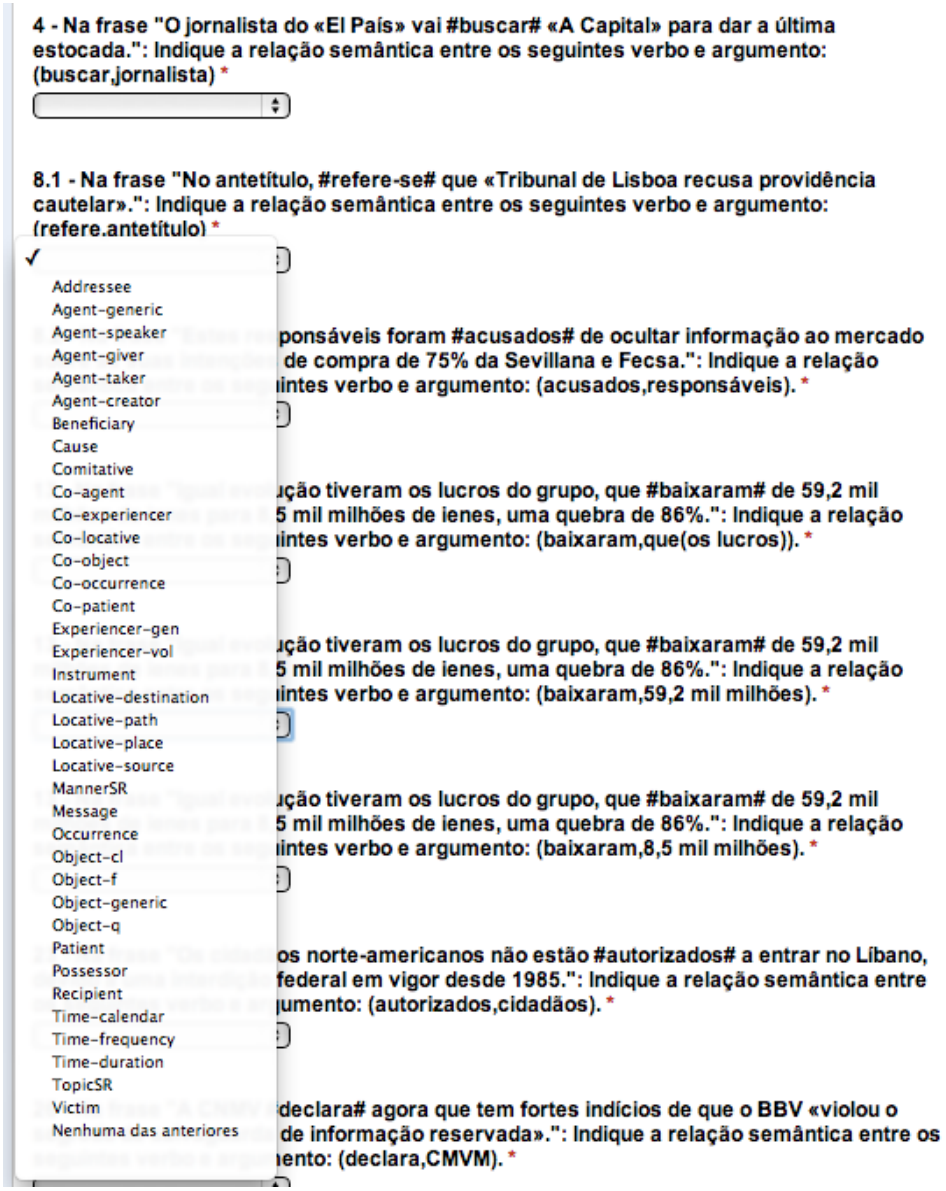


Figure 5.5: Semantic role labeling multiple choice form

Once all the answers have been collected, the results were analyzed and the answers of 2 of the annotators had to be discarded. The first one because some corrections had to be done to the form and the data from this annotator became incorrectly formatted and could not be retrieved; the second, because the answers were totally inconsistent with the guidelines. In the following case, a time complement is at play, so the correct answer should clearly have been `TIME-CALENDAR`. However, this annotator chose `AGENT-GENERIC`.

(87) *Na frase "A CNMV declara agora que tem fortes indícios de que o BBV "violou o segredo de salvaguarda de informação reservada".": Indique a relação semântica entre os seguintes verbo e argumento: (declara,agora).*

'In the sentence "CNVM now says they have strong evidence that BBV has "violated the safe confidential information secret"": Indicate the semantic relation between the following verb and argument: (says,now).'

The results were tabulated in a 4-column format and the ReCal 0.1 Alpha for 3+ Coders tool [4] was used to calculate the inter-annotator agreement.

Two experiments were considered:

- the semantic roles distinguishes all the subtypes (`AGENT-GENERIC, AGENT-SPEAKER, AGENT-TAKER, AGENT-CREATOR, AGENT-GIVER; OBJECT-GENERIC, OBJECT-F, OBJECT-Q, OBJECT-CL`, etc.)
- all the subtypes were collapsed into a single semantic role (`AGENT, OBJECT`).

Results are shown in tables 5.9 and 5.10.

As one can see, the average pairwise agreement is around 50%, but the collapsing of the subtypes of the semantic roles increases this value by 9 points, which can be due just to the fact that the number of potential alternatives has been reduced from 38 to 20 options. The best annotators' pair is 1 and 3, reaching an agreement of 73%.

According to Landis (1977)[43], the results are somewhere between "moderate" and "substantial" agreement.

The annotators' data was compared with the golden standard and their precision is presented in table 5.11. These results show the difficulty of the task, as the average precision was 34%. They may help to better frame the automatic results (5.6) as they can be seen as a ceiling to the task, as far as this set of semantic roles is concerned. When compared to the results of Zilio [67], where an inter-annotator agreement (involving 10 annotators and 25 sentences) of approximately 0.25 was achieved, our results proved to be much better.

Table 5.9: Average pairwise agreement

Experiment	Average pairwise % agreement	Pairwise pct. agr. cols 1 & 4	Pairwise pct. agr. cols 1 & 3	Pairwise pct. agr. cols 1 & 2	Pairwise pct. agr. cols 2 & 4	Pairwise pct. agr. cols 2 & 3	Pairwise pct. agr. cols 3 & 4
1	0.52	0.45	0.63	0.48	0.52	0.54	0.52
2	0.61	0.59	0.73	0.55	0.61	0.60	0.58

Table 5.10: Fleiss' Kappa results

Experiment	Fleiss' Kappa	Observed agreement	Expected agreement
1	0.47	0.53	0.11
2	0.52	0.61	0.20

[4] http://dfreelon.org/recal/recal3.php#result1

Table 5.11: Annotators' precision

Annotator 1	**Annotator 2**	**Annotator 3**	**Annotator 4**
0.38	0.34	0.42	0.27

Chapter 6

Conclusion and Future Work

The two main objectives on this project were achieved. Based on the related work, we have determined a general-purpose set of semantic roles and their corresponding, non-ambiguous, reproducible definitions (Appendix A); and we have built a rule-based, automatic semantic role labeling module for the XIP parser [2] (Appendix E), which was integrated in the STRING NLP chain [45]. This SRL module is based on the information derived from the lexicon-grammar for Portuguese full verbs, ViPEr [11]. This database which was enriched, specifically for this purpose, by providing the semantic roles information for every argument position (subject and essential complements) for more than 6,300 verb senses.

In order to achieve the first objective, an extensive review of the related work was done, in chapter 3, comparing the terms and concepts used by the most prominent authors who have worked on this topic. In the end, a set of 37 semantic roles, along with their definitions and examples, was presented. This set of semantic roles was only complete after the systematic encoding of the semantic roles on the full (or lexical/distributional) verbs of ViPEr. Appendix A presents the synopsis of all the semantic roles defined.

In the review of the related work, some attention was also given to the research on SRL in Portuguese. There are relatively few works in semantic role labeling for the Portuguese language. Most systems adopt a machine-learning approach, in contrast to the rule-based grammar approach that was to be adopted within this dissertation. Only Bick (2007) [19] approaches this difficult task of semantic role labeling using a rule-based system, within the constraint grammar framework, and his work has been a reference to this dissertation. He adopted a set of 35 semantic roles, which is closely similar to the set we came up with, despite some minor differences. However, Bick only tested his SRL module on a small journalistic text of about 2,500 words and 884 semantic roles. Therefore, his results, though very impressive, have not yet been scaled up to a larger nor more diverse corpus. Nevertheless, this work sets up a very high standard (0.86 recall, 0.90 precision and 0.89 F-measure), that is yet to match by any other extant Portuguese-oriented systems.

Another very interesting proposal was presented by Amâncio *et al.* (2011) [6] who set out to build an automatic question categorization system by assigning *wh*-question labels to verbal arguments in sentences. These labels correspond to the interrogative pronouns (*who*, *whom*, *whose*, *what*) or interrogative adverbs (*where*, *when*, *how*) that would replace any given constituents in a direct partial interrogative sentence. Though syntactically (*i.e.* formally) founded, these *pro*-forms are closely related to the semantic role those constituents play in the sentences, therefore, they can be considered as (almost) equivalent to semantic roles' labels. In fact, these type of label is much less ambiguous then most of the labels often used in the literature, which are prone to ambiguity and subject to idiosyncratic readings (*e.g.* *theme*, *goal*, *etc.*). A significantly sizable corpus of has been annotated with an impressive figure of 9,820

tags, from a set of 57 labels. This rather large tagset is a consequence of the fact that the authors consider every main preposition introducing the complement, and they also distinguish the subject from the complement positions (using left/right flags on the labels). The machine-learning approach yielded F-measure best scores from 0.74 to 0.79.

In spite of this impressive results, two remarks are in order. The first comes from the fact that the corpus used to annotate the *wh*-question labels has undergone a process of text simplification, adopting the PorSimples framework [32]. As a consequence, the SRL is fairly simplified, for simplified texts do not present the rich complexity that real discourse can convey, and that hinders natural language processing tasks. Secondly, while soundly based, *wh*-question labels are both too detailed to capture in a more concise or abstract way some semantic roles (different prepositions or even a direct object may convey the *locative* semantic role), so they are insufficient to capture several relevant semantic distinctions (under a direct object associated to the *wh*-form *what* one can subsume several semantic roles, such as *topic*, *event*, *etc.*).

Finally, comparison should be made with the work of Sequeira *et al.* (2012) [61], who also adopt a machine-learning approach, using a sizable, 4,416 sentences corpus, taken from Bosque, the revised part of the Floresta Sintá(c)tica Portuguese treebank, and converted it to the CONLL'2004 format. The authors used only the `Arg0`, `Arg1` and `P` (=predicate) tags, which correspond approximately to the *subject*, the *object* and the *verb*. While this is a simplification of the SRL task, it presupposes that some mapping would be done between the *Arg-n* and the semantic roles proper (*agent*, *patient*, *etc.*), which is not included in this work.

Having adopted a rule-based approach ourselves, one of the main contributions of this project was the manual annotation of the ViPEr database of distributional verbs with semantic role information. This means that for each verb sense, different sets of semantic roles can be found for all its syntactic arguments, subject and essential complements. Over 6,300 verb senses were annotated, which correspond to 13,201 semantic roles tags. After the verb-sense disambiguation module (Travanca 2013) [63] selects the adequate verb-sense for a particular instance of a given verb, this semantic role information is provided to the XIP parser, in order to the SRL rule-based module to perform its tasks.

The SRL module, presented in full in Section 4.3, was integrated in the last steps of the NLP processing chain, STRING. In particular, by the time the SRL module comes into play, the main steps of parsing have already taken place, namely, the chunking, named entity recognition and the extraction of the syntactic dependencies (subject, direct complement, modifier, *etc.*). Furthermore, the SRL is an integrant part of the `EVENT` extraction process (Baptista *et al.* 2012) [13], adopting the same generic formalism to associate events (in the sense of predicates) to their participants (or arguments), and tagging them according to their contribution to the sentence overall meaning. Thus, using the informations from ViPEr in association with the disambiguated verb sense, as well as the information from the parser, a total of 167 pattern-matching rules were built. These rules cover sentences where the verb is in its canonical, active form, but also their corresponding passive or passive-like constructions. A set of heuristics is used to also tag the arguments of verbs not yet included in ViPEr as well as several adjunct complements (*e.g.* *comitative*).

The evaluate the system's performance, a corpus annotated with the semantic roles was necessary. The corpus used is a fragment of PAROLE 5.2 , appropriately called *miscelanea*, that consists of 26 short texts of a large variety of topics, totaling 373 sentences, including 7.834 words, of which 896 are verbs. Of these, 511 were full verbs which were already manually annotated with the ViPEr information. A total of 655 semantic roles was manually annotated and the information merged with the parsed corpus. Even if the corpus has a large variety of topics, many semantic roles from our tag set were not found in it. It is of our belief that it would be of interest to annotate a larger corpus.

The evaluation was performed automatically, using a tool developed at the L2F/INESC-ID Lisboa. Two different evaluations were carried out: (i) a strict evaluation, where the semantic role, the governor

and its argument strictly match the *golden standard*; and (ii) a more relaxed evaluation, where the verb is the same but either the semantic role or the argument (but not both) match the reference.

A strict precision of 68.4% was obtained, while a relaxed evaluation reached a precision of 76.4%. In spite of the relatively high precision, the recall (19.8%) and F-measure (strict: 30.1%; relaxed: 31.5%) can not be considered satisfactory yet.

These poor results can be explained by a number of factors. Above all, the accumulation of errors in previous steps of the NLP processing chain. On the other hand, the corpus is extremely heterogenous, and no simplification has been performed, so these results can translate the extreme difficulty of the task.

In order to validate the set of semantic roles, a small corpus composed of 68 random sentences and about 113 semantic roles was taken from the evaluation corpus and presented to 4 annotators (section 5.5). The golden standard was independently defined by 2 linguists. Using the set of the 37 semantic roles, the average pairwise agreement is 52%. Collapsing the subtypes of semantic roles, such as *agent*, *object*, etc., this annotation experiment reaches the average pairwise agreement of 61%, which can be situated somewhere between *moderate* and *substantial* agreement [43]. Precision varied between 0.27 and 0.42, with an average of 35.25.

In view of these results, the actual results of the automatic SRL module can now be put in perspective: first of all, Semantic Role Labeling is a very difficult task, even for humans. Zílio *et al.* (2013) [67] report a 0.25 inter-annotator agreement for 10 annotators, 25 sentences and the VerbNet SR tagset. In a parallel study, Duran *et al.* (2010) [28] made another annotation experiment using *wh*-question tags on 25 simplified Brazilian Portuguese sentences (75 arguments), having reached 0.78 agreement rate.

While the two first experiments can not be directly compared, and the number of sentences is, in both cases, much to small to be generalized, they seem to indicate a low ceiling that NLP systems will struggle to reach. On the other hand, the use of *wh*-question tags, such as the ones used by [28] and later by [6] is a much more reproducible strategy, but they also have several limitations, as it was mentioned before. Still, a combination of the two strategies, probably confirming the semantic roles by associating them to *wh-* tags in the annotation guidelines may prove to be more effective. In our own guidelines, and depending on the specifics of each semantic role, the syntactic information is present, tough not exactly in the form of *pro*-forms equivalence.

In the future, we intend to improve the SRL module to make it able to analyze sentences that suffered different types of restructuring. The passive is already being treated and, in most cases, the semantic roles are correctly labeled. The major challenge seems to be to improve recall. This will probably be achieved by expanding the heuristic rules, at the end of the SRL module, compensating for the excessive precision of the ViPEr-derived information.

We would also like to encode the semantic roles into non-verbal predicates, such as the predicative nouns or the adjectives. This may prove to constitute a challenge, since these POS can involve different type of predicates, not previously covered in the verbal constructions, and because they allow for several sentence reformulations (or transformations) that will be hard to capture automatically.

One of the major challenges this work faced came from the semantic transfer phenomena (metaphor and metonymy), often found in texts; e.g. *Os preços caíram vertiginosamente*, *Aquela casa sempre o acolhera muito bem.* These tropes entail the presence in a syntactic slot of a lexical item that, in the literal word sense, would not be found there and there is no formal clue that this is actually occurring in the sentence. Previous attempts at capturing metonymy, in the frame of Named Entity Recognition (Oliveira 2010 [52]) can be used as a starting point for this task.

There is still a lot of work to be done in the Semantic Role Labeling module here presented, but as one of the first steps towards semantic analysis of Portuguese texts here presented, the results seem

promising.

Bibliography

[1] Afonso, Susana; Bick, Eckhard; Haber, Renato & Santos, Diana. 2002. Floresta sintá(c)tica: a treebank for Portuguese. *Pages 1678–1703 of:* Rodrigues, Manuel González & Araujo, Carmen Paz Suarez (eds), *Proceedings of the 3rd International Conference on Language Resources and Evaluation (LREC 2002).*

[2] Ait-Mokhtar, S.; Chanod, J. & Roux, C. 2002a. Robustness behyond shalowness: incremental dependency parsing. *Natural Language Engineering*, **8**(2/3), 121–144.

[3] Ait-Mokhtar, S.; Chanod, J. & Roux, C. 2002b. Robustness Beyond Shalowness: Incremental Dependency Parsing. *Natural Language Engineering*, **8**(2/3), 121–144.

[4] Allen, J. F. 1994. *Natural Language Understanding.* Benjamin Cummings.

[5] Alva-Manchego, Fernando & Rosa, João Luís. 2012. Towards Semi-supervised Brazilian Portuguese Semantic Role Labeling: Building a Benchmark. *Pages 210–217 of:* Caseli, H.; Villavicencio, A.; Teixeira, A. & Perdigão, F. (eds), *PROPOR 2012.* LNCS/LNAI, vol. 7243. Springer-Verlag Berlin Heidelberg.

[6] Amâncio, Marcelo Adriano; Duran, Magali Sanches & Aluísio, Sandra Maria. 2011 (March). Automatic question categorization: a new approach for text elaboration. *Pages 43–50 of: Procesamiento del Lenguaje Natural*, vol. 46.

[7] Baptista, Jorge. 1997. Conversão, nomes parte-do-corpo e restruturação dativa. *Pages 51–59 of:* Castro, Ivo (ed), *Actas do XII Encontro da Associação Portuguesa de Linguística (Braga - Guimarães, 30 de Setembro a 2 de Outubro de 1996). – Linguística*, vol. 1- Linguística. Lisboa: APL.

[8] Baptista, Jorge. 2004. *Construções Simétricas: Argumentos e Complementos.* Vol. 1. Faculdade de Letras da Universidade do Porto. Pages 353–367.

[9] Baptista, Jorge. 2010. *Les Tables. La grammaire du français par le menu. Mélanges en hommage à Christian Leclère.* Cahiers du CENTAL, no. 6. Louvain-la-Neuve: Presses Universitaires de Louvain. Chap. Verba dicendi: a structure looking for verbs, pages 11–20.

[10] Baptista, Jorge. 2012a. ViPEr: A Lexicon-Grammar of European Portuguese Verbs. *Pages 10–16 of:* Radimsky, Jan (ed), *Proceedings of the 31st International Conference on Lexis and Grammar.* University of South Bohemia in Nové Hrady (Czech Republic). Università degli Studi di Salerno (Italy)/University of South Bohemia in Nové Hrady (Czech Republic).

[11] Baptista, Jorge. 2012b. ViPEr: A Lexicon-Grammar of European Portuguese Verbs. *Pages 10–16 of: Proceedings of the 31st Conference on Lexis and Grammar, Novè Hrady (Czeck Republic).* Universitá degli Studi di Salerno (Italy)/University of South Bohemia in Nové Hrady (Czech Republic).

[12] Baptista, Jorge; Correia, Anabela & Fernandes, Graça. 2004. Frozen Sentences of Portuguese: Formal Descriptions for NLP. *Pages 72–79 of: Workshop on Multiword Expressions: Integrating Processing.* Barcelona, Spain: ACL.

[13] Baptista, Jorge; Mamede, Nuno & Cabarrão, Vera. 2012a. *Classification Directives for Events and Relations Extraction between Named Entities in Portuguese Texts.* Technical Report. INESC-ID Lisboa.

[14] Baptista, Jorge; Vieira, Lucas; Diniz, Cláudio & Mamede, Nuno. 2012b. Coordination of "-mente" ending Adverbs in Portuguese: an Integrated Solution. *Pages 24–34 of:* Caseli, H.; Villavicencio, A.; Teixeira, A. & Perdigão, F. (eds), *Computational Processing of the Portuguese Language, Proceedings of the 10th International Conference, PROPOR 2012.* LNCS/LNAI, vol. 7243. Coimbra, Portugal: Springer-Verlag Berlin Heidelberg.

[15] Baptista, Jorge; Vera Cabarrão, Nuno João Mamede. 2012c. *Classification directives for Events and Relations Extraction between Named Entities in Portuguese texts.* Technical Report. L2F-Spoken Language Laboratory.

[16] Bernhardi, August. 1805. *Anfangsgründe der Sprachwissenschaft.* Berlin: Heinrich Frölich.

[17] Bick, Eckard. 2000. *The Parsing System "PALAVRAS". Automatic Grammatical Analysis of Portuguese in a Constraint Grammar Framework.* Arhus University Press.

[18] Bick, Eckhard. 2006 (December 1-2, 2006). Noun Sense Tagging: Semantic Prototype Annotation of a Portuguese Treebank. *Pages 127–138 of:* Hajic, Jan & Nivre, Joakim (eds), *Proceedings of the 5th Workshop on Treebanks and Linguistic Theories.*

[19] Bick, Eckhard. 2007. Automatic Semantic Role Annotation for Portuguese. *Pages 1713–1716 of: Proceedings of TIL 2007 - 5th Workshop on Information and Human Language Technology / Anais do XXVII Congresso da SBC.*

[20] Brown, E. K. & Miller, Jim (eds). 1999. *Concise Encyclopedia of Grammatical Categories.* Elservier.

[21] Burchardt, Aljoscha; Erk, Katrin; Frank, Anette; Kowalski, Andrea & Pado, Sebastian. 2006. SALTO: A versatile multi-level annotation tool. *In: LREC-2006.*

[22] Caseli, H.M.; Pereira, T.F.and Specia, L.; Pardo, T.A.S.; Gasperin, C. & Aluísio, S.M. 2009. Building a Brazilian Portuguese parallel corpus of original and simplified texts. *Pages 59–70 of:* Gelbukh, Alexander (ed), *Advances in Computational Linguistics, Research in Computer Science*, vol. 41. 10th Conference on Intelligent Text Processing and Computational Linguistics.

[23] Collobert, R.; Weston, J.; Bottou, L.; Karlen, M.; Kavukcuoglu, K. & Kuksa, P.P. 2011. Natural language processing (almost) from scratch. *Journal of Machine Learning Research*, 2493–2537.

[24] da Paixão Fernandes, Gaia Olívia. 2011. *Classification and Word Sense Disambiguation: The case of –mente ending adverbs in Brazilian Portuguese.* Master thesis, Erasmus-Mundus International Master on Natural Language Processing and Human Language Technologies, Universitat Autónoma de Barcelona/Universidade do Algarve.

[25] Diniz, Cláudio & Mamede, Nuno. 2011. *LexMan - Lexical Morphological Analyser.* Technical Report. L2F/INESC-ID Lisboa.

[26] Diniz, Cláudio; Mamede, Nuno & Pereira, João D. 2010. RuDriCo2 - a faster disambiguator and segmentation modifier. *Pages 573–584 of: II Simpósio de Informática (INForum 2010).*

[27] Dowty, David. 1991. Thematic Proto-Roles and Argument Selection. *Pages 547–619 of: Language*, vol. 67. Linguistic Society of America.

[28] Duran, M. S.; Amâncio, M. A & Aluísio, S.M. 2010. Assigning Wh-Questions to Verbal Arguments in a Corpus of Simplified Texts. *Pages 1–6 of: Computational Processing of the Portuguese Language, Proceedings of PROPOR 2010, 9th International Conference on Computational Processing of the Portuguese Language, Extended Activities Proceedings, CD-ROM vol. 1.*

[29] Fillmore, Charles J. 1968. *Universals in Linguistic Theory.* Holt, Rinehart and Winston. Chap. The case of case, pages 1–88.

[30] Fonseca, Erick Rocha & Rosa, João Luís. 2012. An Architecture for Semantic Role Labeling on Portuguese. *Pages 204–209 of:* Caseli, H.; Villavicencio, A.; Teixeira, A. & Perdigão, F. (eds), *PROPOR 2012.* LNCS/LNAI, vol. 7243. Coimbra, Portugal: Springer-Verlag Berlin Heidelberg.

[31] Freitas, Cláudia; Santos, Diana; Oliveira, Hugo Gonçalo & Quental, Violeta. 2012 (January). VARRA: Validação, Avaliação e Revisão de Relações Semânticas no AC/DC. *In: Livro do IX Encontro de Linguística de Corpus.*

[32] Gasperin, C.; Maziero, E. & Aluísio, S.M. 2010. Challenging Choices for Text Simplification. *Pages 40–50 of:* Branco, António; Klautau, Aldebaro; Vieira, Renata & de Lima, Vera Lúcia Strube (eds), *Computational Processing of the Portuguese Language.* LNCS/LNAI, vol. 6001. Berlin: Springer, for PROPOR.

[33] Gildea, Daniel & Jurafsky, Daniel. 2002. Automatic Labeling of Semantic Roles. *Computational Linguistics*, 245–288.

[34] Gross, Gaston. 1989. *Les construction converses du français.* Genève: Droz.

[35] Gross, Maurice. 1975. *Méthodes en Syntaxe. Régime des constructions complétives.* Paris: Hermann.

[36] Gross, Maurice. 1981. Les bases empiriques de la notion de prédicat sémantique. *Langages*, 7–52.

[37] Gross, Maurice. 1986. *Grammaire transformationnelle du français: 3 – Syntaxe de l'adverbe.* Paris: ASSTRIL.

[38] Guillet, A. & Leclère, Ch. 1981. Restructuration du groupe nominal. *Langages*, 99–125.

[39] Hagège, Caroline; Baptista, Jorge & Mamede, Nuno. 2008. Proposta de Anotação e Normalização de Expressões Temporais da Categoria TEMPO para o HAREM II. *Pages 289–308 of:* Mota, Cristina & Santos, Diana (eds), *Desafios na avaliação conjunta do reconhecimento de entidades mencionadas: O Segundo HAREM.* Linguateca.

[40] Hagège, Caroline; Baptista, Jorge & Mamede, Nuno. 2010. Caracterização e Processamento de Expressões Temporais em Português. *Linguamática*, **2**(1), 63–76.

[41] Hall, Mark; Frank, Eibe; Holmes, Geoffrey; Pfahringer, Bernhard; Reutemann, Peter & Witten, Ian H. 2009. The WEKA Data Mining Software: An Update. *Pages 10–18 of: SIGKDD Explorations*, vol. 11.

[42] Harris, Zellig Sabettai. 1991. *A Theory of Language and Information: A Mathematical Approach.* Clarendon Press, Oxford.

[43] Landis, J. Richard & Koch, Gary G. 1977. The Measurement of Observer Agreement for Categorical Data. *Pages 159–174 of: Biometrics*, vol. 33. International Biometric Society.

[44] Levin, Beth. 1993. *English Verbs and Alternations: A Preliminary Investigation.* Chicago: The University of Chicago Press.

[45] Mamede, Nuno; Baptista, Jorge; Diniz, Cláudio & Cabarrão, Vera. 2012. STRING: An Hybrid Statistical and Rule-Based Natural Language Processing Chain for Portuguese. *In:* Caseli, H.; Villavicencio, A.; Teixeira, A. & Perdigão, F. (eds), *Computational Processing of the Portuguese Language, Proceedings of the 10th International Conference, PROPOR 2012 Demo Sessions*, vol. Demo Session.

[46] Marques, João. 2013. *Anaphora Resolution.* Master thesis, Engenharia Informática e de Computadores (MSc thesis), Instituto Superior Técnico, Universidade Técnica de Lisboa.

[47] Marques, João; Baptista, Jorge & Mamede, Nuno. 2013. *Anaphora Annotation Guidelines.* Technical Report. INESC-ID Lisboa.

[48] Maurício, Andreia. 2011. *Identificação, Classificação e Normalização de Expressões Temporais.* Master thesis, Instituto Superior Técnico, Universidade Técnica de Lisboa.

[49] Mendes, Amália. 2001. *Predicados Verbais Psicológicos do Português. Um Contributo para o Estudo da Polissemia Verbal.* Fundação Calouste Gulbenkian/ Fundação para a Ciência e a Tecnologia.

[50] Morante, R. & Busser, B. 2007. ILK2: Semantic Role Labelling for Catalan and Spanish using TiMBL. *In: 4th International Workshop on Semantic Evaluations, pp. 183–186.* Prague, Czech Republic: ACL.

[51] Nascimento, M.; Bettencourt, J.; Marrafa, P.; Ribeiro, R.; Veloso, R. & Wittman, L. 1998 (September). LE PAROLE – Do corpus à modelização da informação lexical num sistema multi-função Do corpus à modelização da informação lexical num sistema multi-função. *Pages 115–134 of: Actas do XIII Encontro da APL.*

[52] Oliveira, Diogo. 2010. *Extraction and Classification of Named Entities.* Master thesis, Instituto Superior Técnico, Universidade Técnica de Lisboa.

[53] Oliveira, Hugo Gonçalo; Santos, Diana & Gomes, Paulo. 2010. Extracção de Relações Semânticas entre Palavras a partir de um Dicionário: o PAPEL e sua Avaliação. *Linguamática*, **2**(1), 77–93.

[54] Palma, Cristina. 2009. *Estudo Contrastivo Português-Espanhol de Expressões Fixas Adverbiais.* Master thesis, FCHS-UALG.

[55] Palmer, Martha; Gildea, Daniel & Kingsbury, Paul. 2005. The Proposition Bank: An Annotated Corpus of Semantic Roles. *Computational Linguistics*, **31**(1), 71–106.

[56] Palmer, Martha; Gildea, Daniel & Xue, Nianwen. 2010. *Semantic Role Labeling.* Morgan and Claypool.

[57] Resnik, Philip & Lin, Jimmy. 2013. *The Handbook of Computational Linguistics and Natural Language Processing.* Wiley-Blackwell. Chap. Evaluation of NLP Systems, pages 271–295.

[58] Ribeiro, Ricardo. 2003. *Anotação Morfossintáctica Desambiguada do Português.* Master thesis, Instituto Superior Técnico, Lisboa.

[59] Ruppenhofer, Josef; Ellsworth, Michael; Petruck, Miriam R. L.; Johnson, Christopher R. & Scheffczyk, Jan. 2010. *FrameNet II: Extended Theory and Practice.* Technical Report. International Computer Science Institute.

[60] Schuler, Karin Kipper. 2005. *VerbNet: a broad-coverage, comprehensive verb lexicon.* Ph.D. thesis, University of Pennsylvania.

[61] Sequeira, João; Gonçalves, Teresa & Quaresma, Paulo. 2012. Semantic Role Labeling for Portuguese – A Preliminary Approach. *Pages 193–203 of:* Caseli, H.; Villavicencio, A.; Teixeira, A. & Perdigão, F. (eds), *PROPOR 2012*. LNCS/LNAI, vol. 7243. Springer-Verlag Berlin Heidelberg.

[62] Shen, Dan. 2007. Using Semantic Roles to Improve Question Answering. *Pages 12–21 of: In Proceedings of EMNLP 2007, Prague*. Association for Computational Linguistics.

[63] Travanca, Tiago. 2013. *Verb Sense Disambiguation*. Master thesis, IST/INESC-ID Lisboa.

[64] Vicente, Alexandre. 2013. *LexMan: um Segmentador e Analisador Morfológico com Transdutores*. Master thesis, Instituto Superior Técnico, Universidade Técnica de Lisboa, Lisboa, Portugal.

[65] Widlöcher, Antoine & Mathet, Yann. 2012. The Glozz Platform: A Corpus Annotation and Mining Tool. *Pages 171–180 of: Proceedings of the 2012 ACM Symposium on Document Engineering*.

[66] Zanette, Adriano; Scarton, Carolina & Zilio, Leonardo. 2012. Automatic xtraction of subcategorization frames from corpora: an approach to Portuguese. *In:* Caseli, H.; Villavicencio, A.; Teixeira, A. & Perdigão, F. (eds), *PROPOR 2012*. LNCS/LNAI, vol. 7243. Coimbra, Portugal: Springer-Verlag Berlin Heidelberg.

[67] Zilio, Leonardo; Ramisch, Carlos & Finatto, Maria. 2013 (December). Desenvolvimento de um recurso léxico com papéis semânticos para o português. *Pages 23–41 of: Linguamática*, vol. 5.

Appendix A

Synopsis of semantic roles

Tables A1 to A5 summarize the semantic roles that have been considered in this project.

Table A.1: Semantic Roles for Portuguese: 1- Human arguments

Semantic Role	**Definition**	**Example (translation)**
AGENT-GENERIC (agent-generic)	The **human** entity that performs an **action**; implies **volition**, **motion**	***O Pedro*** *cantou o fado* (Peter sung the Fado)
AGENT-SPEAKER (agent-speaker)	AGENT, source of the MESSAGE in **communication** predicate	***O João*** *disse isso à Ana* (John said that to Ana)
AGENT-GIVER (agent-giver)	AGENT in **transfer** predicate: orientation: subj > OBJECT-GENERIC > indir.compl.	***A Ana*** *deu uma prenda ao João* (Ana gave John a present)
AGENT-TAKER (agent-taker)	AGENT in **transfer** predicate orientation: subj < OBJECT-GENERIC < indir.compl.	***O João*** *roubou a mala à senhora* (John stole the old lady's purse)
AGENT-CREATOR (agent-creator)	AGENT in **apparition** predicates	***O João*** *construiu uma casa* (John built a house)
EXPERIENCER-GENERIC (experiencer-generic)	The **human** entity that experiences a physical or psycological state; also the subject of mental perception predicates (pensar)	*Isso irritou* ***o Pedro*** (That irritated Peter)
EXPERIENCER-VOL (experiencer-vol)	The **human** entity that expresses an order or desire	***O Pedro*** *ordenou-a que fosse* (Peter told her to go)
PATIENT (patient)	The **human** entity directly *affected* by an **action**	*O Pedro magoou* ***a Ana*** (Peter has hurt Ana)
RECIPIENT (recipient)	The **human** entity that receives something from an AGENT-GIVER in a transfer predicate	*A Ana deu-****lhe*** *umas meias* (Ana gave him some socks)
ADDRESSEE (addressee)	The **human** entity that is indirect OBJECT-GENERIC of a **communication act**	*O Pedro disse que sim* ***à Ana*** (Pedro said yes to Ana)
BENEFICIARY (beneficiary)	The **human** entity benefited by an **action** (though it is not an essential argument)	*O João abriu caminho para* ***a Ana*** (John made way for Ana)
VICTIM (victim)	The **human** entity affected by an **action**, (though it is not an essential argument)	*O Pedro estragou-****lhe*** *o relvado* (Peter ruined her the lawn)
POSSESSOR (possessor)	The **human** entity that owns an OBJECT-GENERIC	***O Pedro*** *tem uma bicicleta* (Peter has a bycicle)

Table A.2: Semantic Roles for Portuguese: 2- Non human arguments

Sem. Role	Definition	Example (translation)
OBJECT-GENERIC (object-generic)	**Non-human, concrete** entities, usually direct complements of transitive verbs	*O Pedro agarrou* ***a pedra*** (Peter caught the stone)
MESSAGE (message)	Typically, the direct object of a *communication* act	*O João disse* ***que não gostava dela*** (John said he didn't like her)
OCCURRENCE (occurrence)	The subject of **occurrence** support-verbs and other special predicates	***O jogo*** *ocorreu ontem* (The game occurred yesterday)
CAUSE (cause)	**Non-AGENTive** NP, PP or clause, interpreted as CAUSE; it can be a **human** NP	***O tornado*** *destruiu a cidade* (The tornado destroyed the city)
COMITATIVE (comitative)	Non essential **human** entity that accompanies a more centrally involved entity	*O Pedro foi à cidade com* ***a Ana*** (Peter went to the city with Ana)
MANNERSR (manner)	**Manner** adverbs and similar adverbials *como? (how)*	*O Pedro respira* ***cuidadosamente*** (Pedro breathes carefully)
INSTRUMENT (instrument)	OBJECT used to perform an **action**	*O João bateu-lhe com* ***o martelo*** (John has hit him with a hammer)
TOPICSR (topic)	Subject matter usually in NP's introduced by *de*, *em*, *sobre* or *acerca de*, and associated with communication acts	*O Pedro falou com a Ana sobre* ***a peça*** (Peter spoke to Ana about the play)
OBJECT-F (object-f)	The direct complement of **mental perception** verbs such as *adorar*. Usually **completive** sentences	*O Pedro adora* ***ouvir a Ana*** (O Pedro percebe **que a Ana o ama**) (Peter understands **that Ana loves him**)
OBJECT-CL (object-cl)	**Human body parts**, usually functioning as direct complements of transitive verbs	*O Pedro bateu na parede com* ***a cabeça*** (Peter has hit the wall with the head)
OBJECT-Q (object-q)	Quantified NP's or PP's expressing a measurement, a value or an amount; involving measures units. The verb is 32NM or the like. Associated with *quanto(s)?* (how much/many?)	*O Pedro mede* ***2 metros de altura*** (Peter is 2 meters high)

Table A.3: Semantic Roles for Portuguese: 3 - Locative arguments

Sem. Role	Definition	Example (translation)
LOCATIVE-PLACE (locative-place)	Static locative predicates ; "onde? (where)"	*Chove muito* ***em Lisboa*** (It rains a lot in Lisbon)
LOCATIVE-PATH (locative-path)	Dynamic locative predicates "por onde? (through where?)"	*A Ana atravessou* ***Faro*** (Ana crossed Faro)
LOCATIVE-SOURCE (locative-source)	The location prior to the **motion** with dynamic locative predicates	*O Pedro partiu* ***de Lisboa*** (Peter departed from Lisbon)
LOCATIVE-DESTINATION (locative-destination)	The location after the **action** with dynamic locative predicates	*A Ana chegou ontem* ***a Faro*** (Ana arrived in Faro yesterday)

Table A.4: Semantic Roles for Portuguese: 4 - Time arguments

Sem. Role	Definition	Example (translation)
TIME-CALENDAR (time-calendar)	Absolute dates; *quando?* (when?)	*Vou viajar no* ***dia 19 de Outubro*** (I will travel on October 19th)
TIME-FREQUENCY (time-frequency)	Time expressions that express a repetition in time; *how often?* (quantas vezes?)	*Os meus pais vêm cá* ***diariamente*** (My parents come here daily)
TIME-DURATION (time-duration)	The amount of time an action takes to occur; *how long?* (quanto tempo?)	*O jogo dura* ***60 minutos*** (The game lasts 60 minutes)

Note: Generic time expressions ([40]) are marked with the corresponding SR of the syntactic slot of the predicate they fill in, e.g. O **verão** aborrece-me. (**Summer** bothers me).

Table A.5: Semantic Roles for Portuguese: 5 - Symmetric arguments

Sem. Role	**Definition**	**Example (translation)**
CO-AGENT	Can change places with the AGENT	*O Pedro casou com* ***a Ana***
(co-agent)	without changing the meaning of the sentence	(Peter married Ana)
CO-PATIENT	Can change places with the PATIENT	*O padre casou o* ***Pedro e a Ana***
(co-patient)	without changing the meaning of the sentence	(The priest married Peter and Ana)
CO-OBJECT	Can change places with the OBJECT-GENERIC	*A farinha misturou-se com* ***o açucar***
(co-object)	without changing the meaning of the sentence	(The flour mixed with the sugar)
CO-EXPERIENCER	Can change places with the EXPERIENCER	*As opiniões do Pedro* ***e da Ana*** *convergem.*
(co-experiencer)	without changing the meaning of the sentence	(Peter and Ana's opinions converge)
CO-OCCURRENCE	Can change places with the OCCURRENCE	*A vinda do Pedro coincide com* ***a vinda dela***
(co-occurrence)	without changing the meaning of the sentence	(Peter's arrival coincides with her arrival)
CO-LOCATIVE	Can change places with the LOCATIVE	*A casa do Rui confronta com* ***a casa da Ana***
(co-locative)	without changing the meaning of the sentence	(Rui's house borders Ana's house)

Note: The symmetric SRs are only marked in the case that two arguments are not coordinated.

Appendix B

Semantic Roles' distribution by syntactic slot

Table B.1: Semantic Roles' distribution by syntactic slot

N0	**N1**	**N2**
SR-N0-agent-gen	SR-N1-addressee	SR-N2-addressee
SR-N0-agent-giver	SR-N1-beneficiary	SR-N2-agent-giver
SR-N0-agent-speaker	SR-N1-cause	SR-N2-co-agent
SR-N0-agent-taker	SR-N1-co-agent	SR-N2-locative-dest
SR-N0-cause	SR-N1-co-object	SR-N2-locative-place
SR-N0-experiencer-gen	SR-N1-experiencer-gen	SR-N2-locative-source
SR-N0-experiencer-vol	SR-N1-instrument	SR-N2-message
SR-N0-locative-place	SR-N1-locative-dest	SR-N2-object-f
SR-N0-object-cl	SR-N1-locative-path	SR-N2-object-gen
SR-N0-object-f	SR-N1-locative-place	SR-N2-object-q
SR-N0-object-gen	SR-N1-locative-source	SR-N2-occurrence
SR-N0-occurrence	SR-N1-manner	SR-N2-patient
SR-N0-recipient	SR-N1-message	SR-N2-recipient
SR-N0-time-calendar	SR-N1-object-cl	SR-N2-victim
SR-N0-agent-locative-place	SR-N1-object-f	SR-N2-topic
SR-N0-agent-object	SR-N1-object-gen	SR-N2-beneficiary
SR-N0-patient-object-cl	SR-N1-object-q	SR-N2-co-agent-co-object
	SR-N1-occurrence	SR-N2-co-object
	SR-N1-patient	SR-N2-co-patient
	SR-N1-recipient	SR-N2-instrument
	SR-N1-time-calendar	SR-N2-manner
	SR-N1-topic	SR-N2-time-calendar
	SR-N1-patient-object	
	SR-N1-locative-source-locative-dest	
	SR-N1-agent-object	
	SR-N1-co-agent-co-patient	
	SR-N1-patient-object-cl	
	SR-N1-object-occurrence	
	SR-N1-co-occurrence	
	SR-N1-co-locative	
	SR-N1-co-experiencer	

Appendix C

SR annotation on ViPEr

Table C.1: Semantic Roles frequency in ViPEr verbs construction

Semantic Role	**N0**	**N1**	**N2**	**%**	**Cumul%**
agent-gen	4248	0	0	0,3218	0,3218
object-gen	392	2682	164	0,2453	0,5671
patient	0	899	18	0,0695	0,6365
experiencer-gen	522	373	0	0,0678	0,7043
object-f	37	400	178	0,0466	0,7509
cause	536	1	0	0,0407	0,7916
locative-dest	0	165	242	0,0342	0,8258
locative-place	15	305	13	0,0252	0,8510
agent-speaker	281	0	0	0,0213	0,8723
object-cl	20	208	0	0,0173	0,8896
addressee	0	34	170	0,0155	0,9050
message	0	178	13	0,0145	0,9195
locative-source	0	29	131	0,0121	0,9316
occurrence	34	99	2	0,0102	0,9418
co-agent	0	106	20	0,0095	0,9514
recipient	4	1	106	0,0084	0,9598
co-object	0	12	75	0,0066	0,9664
agent-giver	60	0	6	0,0050	0,9714
agent-cause	62	0	0	0,0047	0,9761
object-q	0	52	2	0,0041	0,9802
locative-path	0	47	0	0,0036	0,9837
agent-creator	36	0	0	0,0027	0,9864
patient-object	0	32	0	0,0024	0,9889
experiencer-vol	26	0	0	0,0020	0,9908
locative-source-locative-dest	0	22	0	0,0017	0,9925
agent-object	13	5	0	0,0014	0,9939
agent-taker	18	0	0	0,0014	0,9952
topic	0	12	5	0,0013	0,9965
manner	0	8	1	0,0007	0,9972
co-patient	0	0	7	0,0005	0,9977
co-agent-co-object	0	1	5	0,0005	0,9982
time-duration	0	6	0	0,0005	0,9986
instrument	0	4	1	0,0004	0,9990
victim	0	0	4	0,0003	0,9993
co-experiencer	0	3	0	0,0002	0,9995
time-calendar	1	0	1	0,0002	0,9997
beneficiary	0	1	0	0,0001	0,9998
co-locative	0	1	0	0,0001	0,9998
co-occurrence	0	1	0	0,0001	0,9999
possessor	0	1	0	0,0001	1,0000
Total	**6305**	**5688**	**1164**	**1,0000**	**1,0000**

Table C.2: Macro semantic roles

Semantic Role	**N0**	**N1**	**N2**	**% SR**
agent-x	4718	5	6	0,36
object-x	462	3379	344	0,32
patient-x	0	931	18	0,07
experiencer-x	548	373	0	0,07
locative-x	15	568	386	0,07
co-x	0	124	107	0,02
time-x	1	6	1	$<$ 0,01

Table C.3: Semantic roles combinations frequencies I

Frequency	**Semantic Roles combinations**	%
1847	SR-N0-agent-gen SR-N1-object-gen	0,29
568	SR-N0-agent-gen SR-N1-patient	0,09
343	SR-N0-experiencer-generic SR-N1-object-f	0,05
330	SR-N0-cause SR-N1-experiencer-generic	0,05
289	SR-N0-object-gen	0,05
271	SR-N0-agent-gen SR-N1-locative-place	0,05
233	SR-N0-agent-gen	0,04
198	SR-N0-agent-gen SR-N1-object-cl	0,04
169	SR-N0-agent-speaker SR-N1-message SR-N2-addressee	0,03
151	SR-N0-agent-gen SR-N1-object-gen SR-N2-location-destination	0,03
1847	SR-N0-agent-gen SR-N1-object-gen	0,2930
568	SR-N0-agent-gen SR-N1-patient	0,0901
343	SR-N0-experiencer-generic SR-N1-object-f	0,0544
330	SR-N0-cause SR-N1-experiencer-generic	0,0524
289	SR-N0-object-gen	0,0459
271	SR-N0-agent-gen SR-N1-locative-place	0,0430
233	SR-N0-agent-gen	0,0370
198	SR-N0-agent-gen SR-N1-object-cl	0,0314
169	SR-N0-agent-speaker SR-N1-message SR-N2-addressee	0,0268
151	SR-N0-agent-gen SR-N1-object-gen SR-N2-location-destination	0,0240
144	SR-N0-agent-gen SR-N1-location-destination	0,0228
124	SR-N0-agent-gen SR-N1-object-gen SR-N2-object-gen	0,0197
101	SR-N0-agent-gen SR-N1-co-agent	0,0160
92	SR-N0-agent-gen SR-N1-object-gen SR-N2-location-source	0,0146
85	SR-N0-cause SR-N1-occurrence	0,0135
85	SR-N0-agent-gen SR-N1-patient SR-N2-location-destination	0,0135
80	SR-N0-cause SR-N1-object-gen	0,0127
71	SR-N0-agent-gen SR-N1-object-gen SR-N2-co-object	0,0113
61	SR-N0-experiencer-generic	0,0097
44	SR-N0-agent-gen SR-N1-object-gen SR-N2-recipient	0,0070
42	SR-N0-agent-gen SR-N1-location-path	0,0067
41	SR-N0-agent-gen SR-N1-patient SR-N2-object-f	0,0065
41	SR-N0-agent-giver SR-N1-object-gen SR-N2-recipient	0,0065
40	SR-N0-agent-cause SR-N1-object-gen	0,0063
35	SR-N0-agent-creator SR-N1-object-gen	0,0056
33	SR-N0-agent-speaker SR-N1-object-f	0,0052
29	SR-N0-agent-gen SR-N1-patient SR-N2-object-gen	0,0046
27	SR-N0-experiencer-generic SR-N1-object-gen	0,0043
26	SR-N0-agent-gen SR-N1-location-source	0,0041
26	SR-N0-occurrence SR-N1-experiencer-generic	0,0041
24	SR-N0-agent-gen SR-N1-patient-object SR-N2-location-source	0,0038
24	SR-N0-experiencer-generic SR-N1-patient	0,0038
23	SR-N0-experiencer-generic SR-N1-patient SR-N2-object-f	0,0036
22	SR-N0-experiencer-vol SR-N1-patient SR-N2-object-f	0,0035

Table C.4: Semantic roles combinations frequencies II

Frequency	**Semantic Roles combinations**	%
22	SR-N0-object-gen SR-N1-locative-place	0,0035
21	SR-N0-agent-gen SR-N1-location-source-locative-dest	0,0033
20	SR-N0-object-gen SR-N1-object-gen	0,0032
20	SR-N0-object-cl	0,0032
19	SR-N0-agent-speaker SR-N1-addressee	0,0030
19	SR-N0-object-gen SR-N1-object-q	0,0030
19	SR-N0-agent-speaker SR-N1-patient SR-N2-object-f	0,0030
17	SR-N0-agent-gen SR-N1-object-q	0,0027
16	SR-N0-agent-gen SR-N1-object-gen SR-N2-co-agent	0,0025
16	SR-N0-object-f	0,0025
15	SR-N0-agent-cause SR-N1-patient SR-N2-object-f	0,0024
14	SR-N0-object-gen SR-N1-location-destination	0,0022
13	SR-N0-agent-gen SR-N1-patient SR-N2-location-source	0,0021
11	SR-N0-object-gen SR-N1-co-object	0,0017
11	SR-N0-object-f SR-N1-experiencer-generic SR-N2-object-f	0,0017
10	SR-N0-agent-gen SR-N1-object-gen SR-N2-locative-place	0,0016
10	SR-N0-cause SR-N1-patient	0,0016
10	SR-N0-cause SR-N1-object-cl	0,0016
10	SR-N0-agent-speaker SR-N1-addressee SR-N2-message	0,0016
9	SR-N0-agent-gen SR-N1-object-f	0,0014
9	SR-N0-experiencer-generic SR-N1-object-f SR-N2-object-f	0,0014
9	SR-N0-agent-speaker SR-N1-object-gen	0,0014
8	SR-N0-agent-taker SR-N1-object-gen SR-N2-recipient	0,0013
7	SR-N0-agent-gen SR-N1-patient SR-N2-co-patient	0,0011
7	SR-N0-experiencer-generic SR-N1-topic	0,0011
7	SR-N0-agent-object SR-N1-object-gen	0,0011
7	SR-N0-agent-gen SR-N1-occurrence	0,0011
7	SR-N0-locative-place SR-N1-object-gen	0,0011
6	SR-N0-cause SR-N1-locative-place	0,0010
6	SR-N0-agent-speaker SR-N1-patient	0,0010
6	SR-N0-agent-speaker SR-N1-object-gen SR-N2-object-f	0,0010
5	SR-N0-object-gen SR-N1-patient	0,0008
5	SR-N0-experiencer-generic SR-N1-manner	0,0008
5	SR-N0-agent-giver SR-N1-object-gen SR-N2-patient	0,0008
4	SR-N0-cause SR-N1-patient SR-N2-object-f	0,0006
4	SR-N0-experiencer-generic SR-N1-message	0,0006
4	SR-N0-agent-gen SR-N1-object-gen SR-N2-object-f	0,0006
4	SR-N0-agent-gen SR-N1-patient SR-N2-recipient	0,0006
4	SR-N0-agent-gen SR-N1-agent-object SR-N2-co-agent-co-object	0,0006
4	SR-N0-agent-gen SR-N1-object-gen SR-N2-patient	0,0006
4	SR-N0-agent-giver SR-N1-object-q SR-N2-object-gen	0,0006
4	SR-N0-experiencer-generic SR-N1-object-gen SR-N2-object-f	0,0006
4	SR-N0-object-gen SR-N1-location-path	0,0006
4	SR-N0-agent-giver SR-N1-patient SR-N2-recipient	0,0006
4	SR-N0-locative-place SR-N1-location-destination	0,0006
4	SR-N0-recipient SR-N1-object-gen SR-N2-agent-giver	0,0006

Table C.5: Semantic roles combinations frequencies III

Frequency	Semantic Roles combinations	%
3	SR-N0-agent-gen SR-N1-patient SR-N2-locative-place	0,0005
3	SR-N0-agent-gen SR-N1-message SR-N2-co-agent	0,0005
3	SR-N0-agent-gen SR-N1-co-agent SR-N2-object-f	0,0005
3	SR-N0-cause SR-N1-object-gen SR-N2-co-object	0,0005
3	SR-N0-agent-cause SR-N1-patient	0,0005
3	SR-N0-object-gen SR-N1-location-source	0,0005
3	SR-N0-agent-taker SR-N1-object-gen SR-N2-patient	0,0005
3	SR-N0-cause SR-N1-experiencer-generic SR-N2-object-f	0,0005
3	SR-N0-experiencer-generic SR-N1-co-experiencer	0,0005
3	SR-N0-occurrence SR-N1-occurrence	0,0005
3	SR-N0-agent-object SR-N1-locative-place	0,0005
3	SR-N0-agent-gen SR-N1-object-q SR-N2-object-gen	0,0005
3	SR-N0-agent-taker SR-N1-object-gen SR-N2-victim	0,0005
3	SR-N0-experiencer-generic SR-N1-time-duration SR-N2-object-f	0,0005
2	SR-N0-agent-taker SR-N1-object-gen SR-N2-agent-giver	0,0003
2	SR-N0-locative-place SR-N1-patient	0,0003
2	SR-N0-agent-giver SR-N1-object-q SR-N2-recipient	0,0003
2	SR-N0-agent-gen SR-N1-instrument SR-N2-patient	0,0003
2	SR-N0-agent-taker SR-N1-object-q SR-N2-object-f	0,0003
2	SR-N0-occurrence SR-N1-manner	0,0003
2	SR-N0-agent-gen SR-N1-topic	0,0003
2	SR-N0-experiencer-generic SR-N1-occurrence	0,0003
2	SR-N0-agent-giver SR-N1-object-gen SR-N2-location-destination	0,0003
2	SR-N0-agent-gen SR-N1-patient-object SR-N2-location-destination	0,0003
2	SR-N0-agent-gen SR-N1-time-duration SR-N2-object-f	0,0003
2	SR-N0-agent-gen SR-N1-addressee SR-N2-message	0,0003
2	SR-N0-experiencer-vol SR-N1-object-f	0,0003
2	SR-N0-agent-speaker SR-N1-patient-object SR-N2-object-f	0,0003
2	SR-N0-experiencer-generic SR-N1-locative-place	0,0003
1	SR-N0-agent-gen SR-N1-object-q SR-N2-object-q	0,0002
1	SR-N0-occurrence SR-N1-beneficiary	0,0002
1	SR-N0-cause SR-N1-object-gen SR-N2-recipient	0,0002
1	SR-N0-agent-cause SR-N1-locative-place	0,0002
1	SR-N0-cause SR-N1-object-gen SR-N2-location-source	0,0002
1	SR-N0-cause SR-N1-location-path	0,0002
1	SR-N0-agent-object SR-N1-agent-object SR-N2-co-agent-co-object	0,0002
1	SR-N0-agent-gen SR-N1-patient-object	0,0002
1	SR-N0-agent-cause SR-N1-location-destination	0,0002
1	SR-N0-time-calendar SR-N1-occurrence	0,0002
1	SR-N0-agent-gen SR-N1-object-gen SR-N2-instrument	0,0002
1	SR-N0-agent-speaker SR-N1-message SR-N2-object-f	0,0002
1	SR-N0-agent-gen SR-N1-topic SR-N2-object-f	0,0002
1	SR-N0-agent-speaker SR-N1-object-gen SR-N2-patient	0,0002
1	SR-N0-experiencer-generic SR-N1-patient SR-N2-object-gen	0,0002
1	SR-N0-object-gen SR-N1-manner	0,0002
1	SR-N0-occurrence SR-N1-co-occurrence	0,0002
1	SR-N0-agent-object SR-N1-co-agent-co-object	0,0002
1	SR-N0-agent-creator SR-N1-patient	0,0002
1	SR-N0-agent-gen SR-N1-occurrence SR-N2-occurrence	0,0002
1	SR-N0-object-gen SR-N1-co-object SR-N2-object-f	0,0002
1	SR-N0-agent-gen SR-N1-co-agent SR-N2-object-gen	0,0002

Table C.6: Semantic roles combinations frequencies IV

Frequency	**Semantic Roles combinations**	%
1	SR-N0-locative-place SR-N1-co-locative	0,0002
1	SR-N0-agent-gen SR-N1-co-agent SR-N2-patient	0,0002
1	SR-N0-agent-speaker SR-N1-object-gen SR-N2-occurrence	0,0002
1	SR-N0-object-f SR-N1-object-gen SR-N2-object-gen	0,0002
1	SR-N0-object-gen SR-N1-time-duration	0,0002
1	SR-N0-experiencer-generic SR-N1-experiencer-generic	0,0002
1	SR-N0-object-f SR-N1-patient SR-N2-object-f	0,0002
1	SR-N0-agent-gen SR-N1-experiencer-generic	0,0002
1	SR-N0-agent-speaker SR-N1-topic	0,0002
1	SR-N0-agent-gen SR-N1-topic SR-N2-co-agent	0,0002
1	SR-N0-agent-object SR-N1-location-destination	0,0002
1	SR-N0-object-f SR-N1-experiencer-generic SR-N2-topic	0,0002
1	SR-N0-agent-giver SR-N1-patient SR-N2-location-destination	0,0002
1	SR-N0-agent-gen SR-N1-message SR-N2-patient	0,0002
1	SR-N0-agent-speaker SR-N1-location-destination	0,0002
1	SR-N0-agent-speaker SR-N1-instrument	0,0002
1	SR-N0-object-f SR-N1-a	0,0002
1	SR-N0-object-f SR-N1-patient SR-N2-object-gen	0,0002
1	SR-N0-object-gen SR-N1-instrument	0,0002
1	SR-N0-object-f SR-N1-object-gen SR-N2-co-object	0,0002
1	SR-N0-experiencer-generic SR-N1-patient SR-N2-topic	0,0002
1	SR-N0-locative-place SR-N1-object-q	0,0002
1	SR-N0-experiencer-vol SR-N1-addressee SR-N2-message	0,0002
1	SR-N0-agent-cause SR-N1-occurrence	0,0002
1	SR-N0-agent-speaker SR-N1-addressee SR-N2-topic	0,0002
1	SR-N0-agent-gen SR-N1-patient SR-N2-addressee	0,0002
1	SR-N0-object-f SR-N1-object-gen	0,0002
1	SR-N0-experiencer-generic SR-N1-object-f SR-N2-topic	0,0002
1	SR-N0-experiencer-generic SR-N1-possessor	0,0002
1	SR-N0-object-f SR-N1-object-f	0,0002
1	SR-N0-agent-gen SR-N1-object-gen SR-N2-time-date	0,0002
1	SR-N0-agent-gen SR-N1-patient SR-N2-patient	0,0002
1	SR-N0-experiencer-vol SR-N1-patient SR-N2-topic	0,0002
1	SR-N0-experiencer-generic SR-N1-patient-object SR-N2-object-f	0,0002
1	SR-N0-cause SR-N1-object-q SR-N2-recipient	0,0002
1	SR-N0-cause SR-N1-patient SR-N2-location-source	0,0002
1	SR-N0-object-gen SR-N1-cause	0,0002
1	SR-N0-object-gen SR-N1-recipient	0,0002
1	SR-N0-agent-gen SR-N1-addressee	0,0002
1	SR-N0-agent-giver SR-N1-object-gen SR-N2-victim	0,0002
1	SR-N0-occurrence SR-N1-locative-place	0,0002
1	SR-N0-object-f SR-N1-patient	0,0002
1	SR-N0-agent-gen SR-N1-patient-object SR-N2-manner	0,0002
1	SR-N0-object-f SR-N1-patient SR-N2-object-q	0,0002
1	SR-N0-agent-cause SR-N1-patient-object SR-N2-recipient	0,0002
1	SR-N0-agent-object SR-N1-location-source-locative-dest	0,0002

Appendix D

Annotation Guidelines

Presentation

This document presents the general guidelines to annotate the semantic roles (SR) in a corpus.

The document assumes basic linguistic knowledge on semantic roles and their use in linguistic description, although a succinct definition is provided for each SR here used in (Talhadas, in progress). The purpose of the annotation task is to show whether the SR's as defined here are adequate and if they are sufficiently explicit.

D.1 Annotation process

- A semantic role (SR) is a binary dependency between a predicate and a argument or adjunct:

  O Pedro costuma ler o jornal na sala durante a tarde
  'Peter usually reads the newspaper in the living room during the afternoon'
  AGENT-GENERIC (ler → Pedro)
  OBJECT-GENERIC (ler → jornal)
  LOCATIVE-PLACE (ler → sala)
  TIME-CALENDAR (ler → tarde)

- It is assumed that the annotation tool is the Glozz annotation software[1], currently being used for this and other annotation tasks at L2F/INESC-ID Lisboa. The *annotation units* are the *tokens* as shown in the Glozz. This is the tokenization performed by STRING ([45]). Because of the tokenization underlying the tokens as shown in Glozz does not take into consideration longer strings of words that form named entities, and since the SR dependencies must hinge on two node tokens (the verb and another word), it is necessary to define general principles for selecting the node word functioning as the verb's SR second argument:

- The SR **first argument** is always a verb; arguments of nouns, adjectives or other categories are to be left untagged: *O Pedro leu a entrevista do PM ao jornal Público*

[1] http://www.glozz.org/

'Peter read the the PM's interview to the Público newspaper'
AGENT-GENERIC (ler → Pedro)
OBJECT-GENERIC (ler → entrevista) ← CORRECT
PATIENT-GEN (entrevista → PM) ← INCORRECT
AGENT-GENERIC (entrevista → jornal) ← INCORRECT

- The SR **second argument** is the head noun of a NP or a PP (see example above); in the case of complex NPs and PPs, formed by strings of PPs, the head of the first NP or PP, on which every other constituent hinges, is to be selected:
  O Pedro leu os jornais do Porto de sábado à noite
  'Peter read the friday night newspapers from Oporto'
  AGENT-GENERIC (ler → Pedro)
  OBJECT-GENERIC (ler → jornais) ← CORRECT
  LOCATIVE-PLACE (ler → Porto) ← INCORRECT
  TIME-CALENDAR (ler → sábado) ← INCORRECT
  TIME-CALENDAR (ler → noite) ← INCORRECT

  In this annotation campaign, only verb arguments and complements are to be annotated. That leaves out any argument or complement of predicative nouns and adjectives, including past participles modifying a noun. The passive constructions with *ser* or *estar* are to be treated.

- Only full verbs are to be analyzed. Operator verbs, copulative verbs and support verbs are not to be analyzed. In these case of verbal chains with auxiliary verbs, the semantic role relation is to be marked from the main verb of the VP.

- Essential arguments and adjunct are not distinguished in the annotation scheme. The SR tagging is meant to help distinguish essential arguments, treated as complements (and noted `COMPL`) from mere adjunct complements/modifiers (noted `MOD`).

- In the case of complex NPs and PPs, formed by strings of NPs in apposition and without commas (e.g. *o jornal Público* 'the newspaper *Público*'), usually forming named entities (NE), the first token of the NP is chosen;

- In long named entities (NE), the first token of the NE is chosen as the node for the SR second argument.

- In the case of nominal determiners, the determined noun (and not the determiner) is to be selected:
  O Pedro leu uma dúzia de jornais
  'Peter read a dozen newspapers'
  AGENT-GENERIC (ler → Pedro)
  OBJECT-GENERIC (ler → jornais) ← CORRECT
  OBJECT-Q (ler → dúzia) ← INCORRECT

  O Pedro perdeu 9,5% da empresa
  'Peter lost 9,5% of the company'
  AGENT-GENERIC (perdeu → Pedro)
  OBJECT-GENERIC (perdeu → empresa) ← CORRECT
  OBJECT-GENERIC (perdeu → %) ← INCORRECT

  In case of headless NPs with quantifiers, the same `OBJECT-GENERIC` semantic role is extracted, even if the head has been zeroed or topicalized, as in the example below (the head of the NP is the % symbol):

  Desta empresa, o Pedro perdeu 95%
  'From this company, Peter lost 95%'
  AGENT-GENERIC (perdeu → Pedro)

OBJECT-GENERIC (perdeu → %) ← CORRECT
OBJECT-GENERIC (perdeu → empresa) ← INCORRECT

Certain event nouns can be quantified by time measure nouns:

A empresa festejou 5 anos de atividade
'The company celebrated 5 years of activity'
AGENT-GENERIC (festejou → empresa)
OCCURRENCE (festejou → atividade) ← CORRECT
TIME-CALENDAR (festejou → anos) ← INCORRECT

- the SR second argument can also be an adverb:
Ontem, o Pedro leu os jornais atentamente
'Yesterday, Peter read attentively the newspapers'
AGENT-GENERIC (ler → Pedro)
OBJECT-GENERIC (ler → jornais)
TIME-CALENDAR (ler → Ontem)
MANER (ler → atentamente)

- Pronominal anaphors are to be marked according to the role of their unreduced form in the source sentence, including relative pronouns:

O Pedro leu os jornais na sala
'Peter read the newspapers in the living room'
→ *Ele leu-os ali*
'He read them there'
AGENT-GENERIC (ler → Ele)
OBJECT-GENERIC (ler → os)
LOCATIVE-PLACE (ler → ali)

- Indefinite pronoun *se* is not to be marked .

E viu-se a terra inteira/ de repente/ surgir redonda/ do azul profundo (Fernando Pessoa, *Mensagem*)
'And one has seen the entire earth/ suddenly/ appear round/ from the deep blue'
EXPERIENCER-GENERIC (viu → se) ← INCORRECT

- The reflex pronouns are marked with the corresponding role of their unreduced form.

O Pedro lavou-se
'Peter washed himself'
OBJECT-GENERIC (lavou → se) ← CORRECT

In intrinsically reflexive constructions, however, the reflex pronoun is not marked.

O Pedro queixou-se.
'Peter complained [himself]'
OBJECT-GENERIC (queixou → se) ← INCORRECT

The reflex pronouns in pronominal passive constructions are also not to be marked; instead, the direct OBJECT-GENERIC NP that has been moved to subject position, is marked with the SR it had in the corresponding active sentence; in many cases, the subject of these pronominal passive constructions appears in a post-verbal position:

Vendem-se apartamentos
'Apartments for sale'
OBJECT-GENERIC (vendem → apartamentos) ← CORRECT
EXPERIENCER-GENERIC (vendem → se) ← INCORRECT

- Echo arguments are to be marked.

  O Pedro e a Ana deram uma prenda um ao outro. 'Peter and Ana gave a present to one another.'
  AGENT-GENERIC (deram → Pedro) ← CORRECT
  AGENT-GENERIC (deram → Ana) ← CORRECT
  RECIPIENT (deram → um) ← CORRECT
  RECIPIENT (deram → outro) ← CORRECT

- Zeroed anaphoric arguments (subject and complements) of verbs are not to be marked, even if their antecedents are within the same sentence:

  O Pedro reuniu-se com o João para 0[o Pedro e o João] discutirem este assunto
  'Peter meet with John to [Peter and John] discuss that '
  AGENT-GENERIC (reuniu → Pedro)
  CO-AGENT-GENERIC (reuniu → João)
  AGENT-GENERIC (discutirem → Pedro) ← INCORRECT
  AGENT-GENERIC (discutirem → João) ← INCORRECT
  TOPIC (discutirem → assunto)

- Relations between sentences are not to be taken into account.

- Semantic roles across appositions are not to be marked, including reduced relative clauses, gerundives,etc.:

  O Pedro, lendo isso, declarou que não iria.
  'Peter reading it, declared he wouldn't go.'
  AGENT-SPEAKER (declarou → Pedro) ← CORRECT
  OBJECT-GENERIC (lendo → isso) ← CORRECT
  AGENT-GENERIC (lendo → Pedro) ← INCORRECT

  O Pedro, conhecido como o terror do bairro, aproximou-se.
  'Peter, known as the hood badboy, came near.'
  OBJECT-GENERIC (conhecido → Pedro) ← INCORRECT
  AGENT-GENERIC (aproximou → Pedro) ← CORRECT

- Relative pronouns are to be marked:

  *A operação, **que** respeita a 4,945 milhões de acções ao preço de 259 francos cada (...)*

  'The operation, **that** respects 4.945 million stocks at the price of 259 francs each (...)'
  OBJECT-F (respeita → que) ← CORRECT

- The arguments of idiomatic expressions are not to be marked.

  O Pedro tem a cabeça nas nuvens. 'Peter has his head on the clouds.'
  EXPERIENCER-GENERIC (tem → cabeça) ← INCORRECT
  LOCATIVE-PLACE (tem → nuvens) ← INCORRECT

- The metonymy is not to be marked.

  O local não respeita as normas de segurança. 'The site doesn't respect the security norms.'
  AGENT-GENERIC (respeita → local) ← INCORRECT
  OBJECT-F (respeita → normas) ← CORRECT

- Most adverbial phrases (ADVP) are to be marked, often as INSTRUMENT or MANNER.

  O local não respeita as normas de segurança. 'The site doesn't respect the security norms.'
  AGENT-GENERIC (respeita → local) ← INCORRECT
  OBJECT-F (respeita → normas) ← CORRECT

- Coordinated NPs are to be marked.

  Os manifestantes aceitaram o abandono do local e a sua identificação. 'The protesters accepted the abandoning of the place and their identification.'
  AGENT-GENERIC (aceitaram → abandono) ← CORRECT
  OBJECT-F (aceitaram → identificação) ← CORRECT

D.2 Special or problematic cases

There are cases where the distinction between SR's can be thin or complex. This section will present some alternatives to solve them.

1. AGENT-GENERIC/ CAUSE

   These SR's are of close interpretation, for the Agent can be interpreted as the cause of an action. The difference between both is the human or non-human feature of the syntactic subject of the sentence, as shown in the examples below:

   (a) *O Pedro destruiu o telhado* 'Peter destroyed the roof'
   (b) *O tornado destruiu o telhado* 'The tornado destroyed the roof'

   In (a), Pedro has the human feature, and so, this is classified as AGENT-GENERIC. In (b), tornado has the non-human feature, and so, this is classified as CAUSE.

2. AGENT-GENERIC/ OBJECT-GENERIC and CO-AGENT/ CO-OBJECT

   In some rarer cases, two different roles can be used in the same slot, as in the following examples:

   (c) *O Pedro e a Ana complementam-se* 'Peter and Ana complement'
   (d) *A carne e o molho complementam-se* 'Meat and sauce complement'

   As can be seen, the verb meaning is the same, only the arguments' features are different.
   In (c) the subject is human, and in (d) the subject is non-human.

3. PATIENT/ OBJECT-GENERIC

   This case is very similar to the dicotomy AGENT-GENERIC/ OBJECT-GENERIC. The only feature that distinguishes the argument's semantic role is the human/ non human features.

   (i) *O Pedro alvejou a Ana* 'Peter shot Ana'

   (j) *O Pedro alvejou a garrafa* 'Peter shot the bottle'

4. BENEFICIARY/ VICTIM

   SR tagged nouns are always facultative, mostly the so-called ethical or benefactive dative complements.

5. Quantification

   Quantification structures with copula verbs are not marked.

   (k) *O valor por ação foi de 2 dólares* 'The stock value was 2 dollars'

   In the case of movement verbs, the source and destination complements are reinterpreted as quantification objects (OBJECT-Q), in the same way as *atingir* (to reach):

   (l) *Os lucros baixaram de 5 milhões de euros para 2,3 milhões de euros* 'The profits lowered from 5 million to 2,3 million dollars'

   (m) *Os lucros atingiram os 10 milhões de dólares* 'The profits have reached 10 million dollars'

6. "Part of" relations

   Constructions involving part of relations are not marked for now.

   (n) *Esta operação insere-se na/ é parte da estratégia definida* 'This operation is part of the defined strategy'

Appendix E

Semantic Role Labeling Module in XIP Portuguese Grammar

```
//!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
//! L2F - Spoken Language Systems Laboratory / INESC-ID Lisboa                 !
//!                                                                            !
//! Copyright (c) 2006-2013 INESC-ID Lisboa.  All Rights Reserved.             !
//!                                                                            !
//! --- CONTENTS ---                                                           !
//!   Project    : Portuguese XIP grammar                                      !
//!   filename   : dependencyEventSRole.xip                                    !
//!   Date       : 2013/February                                               !
//!   Author     : Nuno Mamede & Jorge Baptista & Rui Talhadas                 !
//!                                                                            !
//!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
//!                                                                            !
//!   Dependency make some reorganization of dependencies                      !
//!                                                                            !
//!   Dependencies handled in this file:  AGENT                                !
//!                                       PATIENT                              !
//!                                       ATTRIB (enlarged)                    !
//!                                       MAIN                                 !
//!                                       PP attachment heuristics             !
//!                                       Missing SUBJ handled                 !
//!                                                                            !
//!                                       SUBJ (Anaphora 0)                    !
//!                                                                            !
//!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!

DependencyRules:

//******************************************************************************
//******************************************************************************
//***                                                                       ****
```

```
//***                              FIRST PART                                ****
//***                        (transformations: passive)                      ****
//***                                                                         ****
//*****************************************************************************
//*****************************************************************************


//=========================================================
// Identification of Passive
// Takes into account the distributional features of canonical N1 (assigned by ViPER)
//=========================================================

//Se passive
// O Pedro partiu a janela -> A janela partiu-se
|verb\#1[SR-pass-pron,pass-pron=+]|
if ( CLITIC(\#1,\#2[pron,ref,3p:~]) & \#1[person]:\#2[person] & \#1[number]:\#2[number] &
     ( SUBJ(\#1[SR-N1-hum],\#2[UMB-Human]) || SUBJ(\#1[SR-N1-nhum],\#2[UMB-Human:~]) )
   )
~

|verb\#1[SR-pass-pron,pass-pron:~,pass-pron=+]|
if ( CLITIC(\#1[3p],\#2[pron,ref,3p]) &
     ( (\#1[pl] & \#2[sg:~]) | (\#1[sg] & \#2[pl:~]) ) &
     ( SUBJ(\#1[SR-N1-hum],\#2[UMB-Human]) | SUBJ(\#1[SR-N1-nhum],\#2[UMB-Human:~]) )
   )
~

//=================================================================
// Change Semantic Roles when in presence of instrument raising
//=================================================================

// Ex: A faca corta o pao.
//|\#1[InstrumRaise, SR-N1-object-gen, SR-N0-agent-gen, SR-N0-instrument=+, SR-N0-agent-gen=~ ] |
//  if(\#1[SR-InstrumRaise]) ~

//=========================================================
// Change Semantic Roles when in presence of a passive
//=========================================================
// Quando um verbo esta na forma passiva sao removidas as restri\c{c}\~oes ao N0 ser
// ou nao ser humano

// RT 2013-07-15 regras para a transforma\c{c}ao da passiva

|\#1[transf-passiva, SR-N1-patient, SR-N0-agent-gen, SR-N0-patient=+, SR-N1-agent-gen=+,
    SR-N1-patient=~, SR-N0-agent-gen=~ ]|
  if(\#1[SR-transf-passiva]) ~

|\#1[transf-passiva, SR-N1-object-gen, SR-N0-agent-gen, SR-N0-object-gen=+, SR-N1-agent-gen=+,
    SR-N1-object-gen=~, SR-N0-agent-gen=~ ]|
  if(\#1[SR-transf-passiva]) ~

|\#1[transf-passiva, SR-N1-object-gen, SR-N0-agent-creator, SR-N0-object-gen=+,
```

```
    SR-N1-agent-creator=+, SR-N1-object-gen=~, SR-N0-agent-creator=~ ]|
  if(\#1[SR-transf-passiva]) ~

|\#1[transf-passiva, SR-N1-message, SR-N0-agent-speaker, SR-N0-message=+, SR-N1-agent-speaker=+,
    SR-N1-message=~, SR-N0-agent-speaker=~ ]|
  if(\#1[SR-transf-passiva]) ~

|\#1[transf-passiva, SR-N1-object-gen, SR-N0-agent-giver, SR-N0-object-gen=+, SR-N1-agent-giver=+,
    SR-N1-object-gen=~, SR-N0-agent-giver=~ ]|
  if(\#1[SR-transf-passiva]) ~

|\#1[transf-passiva, SR-N1-object-gen, SR-N0-agent-taker, SR-N0-object-gen=+, SR-N1-agent-taker=+,
    SR-N1-object-gen=~, SR-N0-agent-taker=~ ]|
  if(\#1[SR-transf-passiva]) ~

|\#1[transf-passiva, SR-N1-object-f, SR-N0-experiencer-gen, SR-N0-object-f=+,
    SR-N1-experiencer-gen=+, SR-N1-object-f=~, SR-N0-experiencer-gen=~ ]|
  if(\#1[SR-transf-passiva]) ~

|\#1[transf-passiva, SR-N1-object-gen, SR-N0-recipient, SR-N0-object-gen=+, SR-N1-recipient=+,
    SR-N1-object-gen=~, SR-N0-recipient=~ ]|
  if(\#1[SR-transf-passiva]) ~

|\#1[transf-passiva, SR-N1-occurrence, SR-N0-cause, SR-N0-occurrence=+, SR-N1-cause=+,
    SR-N1-occurrence=~, SR-N0-cause=~ ]|
  if(\#1[SR-transf-passiva]) ~


// O Joao fez isso/ isso foi feito pelo Joao.
|\#1[pass-ser, SR-N1-CDIR=~, SR-N1-por=+]| ~

|\#1[pass-estar, SR-N1-CDIR=~, SR-N1-por=+]| ~

//*****************************************************************************
//*****************************************************************************
//***                                                                       ****
//***                              SECOND PART                              ****
//***                    (apply rules associated with ViPEr)                ****
//*****************************************************************************
//*****************************************************************************

//=============================================================================
//=============================================================================
// Converting N0 Syntactic dependencies into Semantic Roles
//=============================================================================
//=============================================================================

// Exemplo: O Pedro construiu a casa. -> EVENT_AGENT-CREATOR(construiu,Pedro)

if (SUBJ(\#1[SR-N0-agent-creator],\#2[UMB-Human]) & EVENT[other](\#1) & ~EVENT(\#1,\#2) )
  EVENT[agent-creator=+](\#1,\#2).
```

```
// Exemplo: O Pedro partiu a janela. -> EVENT_AGENT-GEN(partiu,Pedro)

if (SUBJ(\#1[SR-N0-agent-gen],\#2[UMB-Human]) & EVENT[other](\#1) & ~EVENT(\#1,\#2) )
  EVENT[agent-generic=+](\#1,\#2).


// Exemplo: O Pedro partiu a janela. -> EVENT_AGENT-GEN(partiu,Pedro)

//RT+JB 2013-07-27 Probably to be removed later- Captures SR-N0-agent-gen
but relaxes the  previous rule in order not to impose the human feature on the subject.
ex. ‘‘O c~ao mordeu o carteiro"

if (SUBJ(\#1[SR-N0-agent-gen],\#2) & EVENT[other](\#1) & ~EVENT(\#1,\#2) )
  EVENT[agent-generic=+](\#1,\#2).


// Exemplo: O Pedro deu a moldura. -> EVENT_AGENT-GIVER(deu,Pedro)

if (SUBJ(\#1[SR-N0-agent-giver],\#2[UMB-Human]) & EVENT[other](\#1) & ~EVENT(\#1,\#2) )
  EVENT[agent-giver=+](\#1,\#2).


// Exemplo: O Pedro falou com a Ana. -> EVENT_AGENT-SPEAKER(falou,Pedro)

if (SUBJ(\#1[SR-N0-agent-speaker],\#2[UMB-Human]) & EVENT[other](\#1) & ~EVENT(\#1,\#2) )
  EVENT[agent-speaker=+](\#1,\#2).


// Exemplo: O Pedro tirou a moldura ao Rui. -> EVENT_AGENT-TAKER(tirou,Pedro)

if (SUBJ(\#1[SR-N0-agent-taker],\#2[UMB-Human]) & EVENT[other](\#1) & ~EVENT(\#1,\#2) )
  EVENT[agent-taker=+](\#1,\#2).


// Exemplo: O tornado partiu a janela. -> EVENT_CAUSE(partiu,tornado)

if (SUBJ(\#1[SR-N0-cause],\#2) & EVENT[other](\#1) & ~EVENT(\#1,\#2) )
  EVENT[cause=+](\#1,\#2).


// Exemplo: O Pedro pensou no assunto. -> EVENT_EXPERIENCER-GEN(pensou,Pedro)

if (SUBJ(\#1[SR-N0-experiencer-gen],\#2[UMB-Human]) & EVENT[other](\#1) & ~EVENT(\#1,\#2) )
  EVENT[experiencer-generic=+](\#1,\#2).


// Exemplo: O Pedro deseja a Ana. -> EVENT_EXPERIENCER-VOL(deseja,Pedro)

if (SUBJ(\#1[SR-N0-experiencer-vol],\#2[UMB-Human]) & EVENT[other](\#1) & ~EVENT(\#1,\#2) )
  EVENT[experiencer-vol=+](\#1,\#2).
```

```
// Exemplo: O martelo partiu a janela. -> EVENT_INSTRUMENT(partiu,martelo)

if (SUBJ(\#1[SR-N0-instrument],\#2) & EVENT[other](\#1) & ~EVENT(\#1,\#2) )
  EVENT[instrument=+](\#1,\#2).


// Exemplo: A estrada termina num beco. -> EVENT_LOCATIVE-PLACE(termina,estrada)

if (SUBJ(\#1[SR-N0-locative-place],\#2) & EVENT[other](\#1) & ~EVENT(\#1,\#2) )
  EVENT[locative-place=+](\#1,\#2).


// Exemplo: A mensagem foi passada pelo Pedro. -> EVENT_MESSAGE(passada,mensagem)

if (SUBJ(\#1[SR-N0-message],\#2) & EVENT[other](\#1) & ~EVENT(\#1,\#2) )
  EVENT[message=+](\#1,\#2).


// Exemplo: O cora\c{c}ao palpitava. -> EVENT_OBJECT-CL(palpitava,cora\c{c}~ao)

if (SUBJ(\#1[SR-N0-object-cl],\#2) & EVENT[other](\#1) & ~EVENT(\#1,\#2) )
  EVENT[object-cl=+](\#1,\#2).


// Exemplo: O Pedro sofreu maus-tratos. -> EVENT_PATIENT(sofreu,Pedro)

if (SUBJ(\#1[SR-N0-patient],\#2[UMB-Human]) & EVENT[other](\#1) & ~EVENT(\#1,\#2) )
  EVENT[patient=+](\#1,\#2).

// Exemplo: O Pedro sua dos pes -> EVENT_PATIENT(sua,Pedro)

if (SUBJ(\#1[SR-N0-patient-object-cl],\#2[UMB-Human]) & EVENT[other](\#1) & ~EVENT(\#1,\#2) )
  EVENT[patient=+](\#1,\#2).

// Exemplo: Os pes do Jo~ao suam -> EVENT_OBJECT-CL(suam,pes)

if (SUBJ(\#1[SR-N0-patient-object-cl],\#2[UMB-Anatomical]) & EVENT[other](\#1) & ~EVENT(\#1,\#2) )
  EVENT[object-cl=+](\#1,\#2).


// Exemplo: O contrato fidelizou o Pedro. -> EVENT_OBJECT-F(fidelizou,contrato)

if (SUBJ(\#1[SR-N0-object-f],\#2) & EVENT[other](\#1) & ~EVENT(\#1,\#2) )
  EVENT[object-f=+](\#1,\#2).


// Exemplo: O p\'o assentou. -> EVENT_OBJECT-GEN(assentou,p\'o)

if (SUBJ(\#1[SR-N0-object-gen],\#2) & EVENT[other](\#1) & ~EVENT(\#1,\#2) )
  EVENT[object-generic=+](\#1,\#2).
```

```
//if (SUBJ(\#1[SR-N0-object-locative-place],\#2[UMB-Location]) & EVENT[other](\#1) & ~EVENT(\#1,\#2) )
//  EVENT[locative-place=+](\#1,\#2).

//if (SUBJ(\#1[SR-N0-object-locative-place],\#2) & EVENT[other](\#1) & ~EVENT(\#1,\#2) )
//  EVENT[object-generic=+](\#1,\#2).


// Exemplo: O concerto come\c{c}a as 3. -> EVENT_OCCURRENCE(come\c{c}a,concerto)

if (SUBJ(\#1[SR-N0-occurrence],\#2) & EVENT[other](\#1) & ~EVENT(\#1,\#2) )
  EVENT[occurrence=+](\#1,\#2).


// Exemplo: O Pedro recebeu uma carta. -> EVENT_RECIPIENT(recebeu,Pedro)

if (SUBJ(\#1[SR-N0-recipient],\#2[UMB-Human]) & EVENT[other](\#1) & ~EVENT(\#1,\#2) )
  EVENT[recipient=+](\#1,\#2).


// Exemplo: O Natal aproxima-se. -> EVENT_TIME-CALENDAR(aproxima,Natal)

if (SUBJ(\#1[SR-N0-time-calendar],\#2) & EVENT[other](\#1) & ~EVENT(\#1,\#2) )
  EVENT[time-calendar=+](\#1,\#2).


// Exemplo: O Pedro partiu a janela. -> EVENT_AGENT-GEN(partiu,Pedro)

if (SUBJ(\#1[SR-N0-agent-cause],\#2[UMB-Human]) & EVENT[other](\#1) & ~EVENT(\#1,\#2) )
  EVENT[agent-generic=+](\#1,\#2).


// Exemplo: O vento partiu a janela. -> EVENT_CAUSE(partiu,vento)

if (SUBJ(\#1[SR-N0-agent-cause],\#2) & EVENT[other](\#1) & ~EVENT(\#1,\#2) )
  EVENT[cause=+](\#1,\#2).


// Exemplo: O Pedro amparou a queda da Ana. -> EVENT_AGENT-GEN(amparou,Pedro)

if (SUBJ(\#1[SR-N0-agent-object],\#2[UMB-Human]) & EVENT[other](\#1) & ~EVENT(\#1,\#2) )
  EVENT[agent-generic=+](\#1,\#2).


// Exemplo: A relva amparou a queda da Ana. -> EVENT_OBJECT-GEN(amparou,relva)

if (SUBJ(\#1[SR-N0-agent-object],\#2) & EVENT[other](\#1) & ~EVENT(\#1,\#2) )
  EVENT[object-generic=+](\#1,\#2).


// Exemplo: A gruta abrigava a Ana. -> EVENT_LOCATIVE-PLACE(abrigava,gruta)

if (SUBJ(\#1[SR-N0-agent-locative-place],\#2[UMB-Human]) & EVENT[other](\#1) & ~EVENT(\#1,\#2) )
```

```
  EVENT[locative-place=+](\#1,\#2).

// Exemplo: O Pedro abrigava a Ana. -> EVENT_AGENT-GEN(abrigava,Pedro)

if (SUBJ(\#1[SR-N0-agent-locative-place],\#2) & EVENT[other](\#1) & ~EVENT(\#1,\#2) )
  EVENT[agent-gen=+](\#1,\#2).

//=============================================================================
// Converting N1 Syntactic dependencies into Semantic Roles
//=============================================================================

//=============================================
// Passive complements
//=============================================

// RT 2013-07-15 regras para a agentes complementos da passiva

//com o tra\c{c}o ‘‘SR-N1-por" introduzido em ‘‘|\#1[pass-ser, SR-N1-CDIR=~, SR-N1-por=+]| ~"
//e com os SRs do suj convertidos em N1

if (^MOD[post=+](\#1[SR-N1-agent-gen, SR-N1-por],\#2) & EVENT[other](\#1) & ~EVENT(\#1,\#2)
    & PREPD(\#2,\#3[lemma:por]) )
  EVENT[agent-generic=+](\#1,\#2),
  COMPL[post=+](\#1,\#2).

if (^MOD[post=+](\#1[SR-N1-agent-creator, SR-N1-por],\#2) & EVENT[other](\#1) & ~EVENT(\#1,\#2)
    & PREPD(\#2,\#3[lemma:por]))
  EVENT[agent-creator=+](\#1,\#2),
  COMPL[post=+](\#1,\#2).

if (^MOD[post=+](\#1[SR-N1-agent-giver, SR-N1-por],\#2) & EVENT[other](\#1) & ~EVENT(\#1,\#2)
    & PREPD(\#2,\#3[lemma:por]))
  EVENT[agent-giver=+](\#1,\#2),
  COMPL[post=+](\#1,\#2).

if (^MOD[post=+](\#1[SR-N1-agent-speaker, SR-N1-por],\#2) & EVENT[other](\#1) & ~EVENT(\#1,\#2)
    & PREPD(\#2,\#3[lemma:por]))
  EVENT[agent-speaker=+](\#1,\#2),
  COMPL[post=+](\#1,\#2).

if (^MOD[post=+](\#1[SR-N1-agent-taker, SR-N1-por],\#2) & EVENT[other](\#1) & ~EVENT(\#1,\#2)
    & PREPD(\#2,\#3[lemma:por]))
  EVENT[agent-taker=+](\#1,\#2),
  COMPL[post=+](\#1,\#2).

if (^MOD[post=+](\#1[SR-N1-cause, SR-N1-por],\#2) & EVENT[other](\#1) & ~EVENT(\#1,\#2)
    & PREPD(\#2,\#3[lemma:por]))
  EVENT[cause=+](\#1,\#2),
  COMPL[post=+](\#1,\#2).

if (^MOD[post=+](\#1[SR-N1-experiencer-gen, SR-N1-por],\#2)
& EVENT[other](\#1) & ~EVENT(\#1,\#2)
```

```
    & PREPD(\#2,\#3[lemma:por]))
  EVENT[experiencer-generic=+](\#1,\#2),
  COMPL[post=+](\#1,\#2).

if (^MOD[post=+](\#1[SR-N1-recipient, SR-N1-por],\#2) & EVENT[other](\#1) & ~EVENT(\#1,\#2)
    & PREPD(\#2,\#3[lemma:por]))
  EVENT[recipient=+](\#1,\#2),
  COMPL[post=+](\#1,\#2).




//================================================
// Agent-giver (N2)
//================================================

//RT 2013-07-23 added
//moved rules from N2 because of rules order

// Exemplo: O Pedro arrebatou o premio a Santa Casa.
if (MOD(\#1[SR-N2-agent-giver],\#2[time:~,location:~]) & (PREPD(\#2,?[lemma:a]) |
    PREPD(\#2,?[lemma:de])) & EVENT[other](\#1) & ~EVENT(\#1,\#2) )
  EVENT[agent-giver=+](\#1,\#2).

//RT 2013-07-23 added in all SRs that have CDIR and MOD ‘‘& ~EVENT[SR](\#1,?)"
// so that it doesn't find 2 equal SRs as N0 or N1 or N2

//================================================
// Addressee
//================================================

// Exemplo: O Pedro chamou a Ana  -->  EVENT_addressee(chamar,Ana)
// Exemplo: O Pedro chamou-a      -->  EVENT_addressee(chamar,a)
//
if ( (CDIR(\#1[SR-N1-addressee],\#2[UMB-Human]) | CDIR(\#1[SR-N1-addressee],\#2[acc])) &
     EVENT[other](\#1) & ~EVENT(\#1,\#2)
   )
  EVENT[addressee=+](\#1,\#2).

//Exemplo: O Pedro telefonou a Ana -> EVENT_addressee(telefonar,Ana)
//
if (CINDIR(\#1[SR-N1-addressee],\#2[UMB-Human]) & PREPD(\#2,?[lemma:a]) & EVENT[other](\#1)
    & ~EVENT(\#1,\#2) & ~EVENT[addressee](\#1,?) )
  EVENT[addressee=+](\#1,\#2).

// Exemplo: O Pedro abriu-se a Ana -> EVENT_addressee(abrir,Ana)
// Exemplo: Os candidatos apelaram aos cidadaos para que fizessem isso.
//
if ( MOD(\#1[SR-N1-addressee],\#2[UMB-Human]) & PREPD(\#2,?[lemma:a]) & EVENT[other](\#1)
    & ~EVENT(\#1,\#2) & ~EVENT[addressee](\#1,?) )
  EVENT[addressee=+](\#1,\#2).
```

```
// Exemplo: Os candidatos apelaram-lhes para que fizessem isso ->
//
if (MOD[dat](\#1[SR-N1-addressee],\#2[time:~,location:~]) & EVENT[other](\#1) & ~EVENT(\#1,\#2)
    & ~EVENT[addressee](\#1,?) )
  EVENT[addressee=+](\#1,\#2).

// Exemplo: O Pedro falou com a Ana  -->  EVENT_addressee(falar,Ana)
//
if (MOD(\#1[SR-N1-addressee],\#2[UMB-Human]) & PREPD(\#2,?[lemma:com]) & EVENT[other](\#1)
    & ~EVENT(\#1,\#2) & ~EVENT[addressee](\#1,?) )
  EVENT[addressee=+](\#1,\#2).

// Exemplo: O Pedro lamentou-se junto da Ana -> EVENT_addressee(lamentar,Ana)
//
if (MOD(\#1[SR-N1-addressee],\#2[UMB-Human]) & PREPD(\#2,?[lemma:"junto de"]) & EVENT[other](\#1)
    & ~EVENT(\#1,\#2) & ~EVENT[addressee](\#1,?) )
  EVENT[addressee=+](\#1,\#2).

//================================================
// Beneficiary
//================================================

// Exemplo: A tempestade abonou-lhes agua -> EVENT_BENEFICIARY(abonar,lhes)
//
if (MOD[dat](\#1[SR-N1-beneficiary],\#2) & EVENT[other](\#1) & ~EVENT(\#1,\#2)
    & ~EVENT[beneficiary](\#1,?) )
  EVENT[beneficiary=+](\#1,\#2).

// Exemplo: A tempestade abona em favor da Ana -> EVENT_BENEFICIARY(abonar,Ana)
//
if (MOD(\#1[SR-N1-beneficiary],\#2) & PREPD(\#2,?[lemma:"em favor de"]) & EVENT[other](\#1)
    & ~EVENT(\#1,\#2) & ~EVENT[beneficiary](\#1,?) )
  EVENT[beneficiary=+](\#1,\#2).

// Exemplo: A tempestade abona a favor da Ana -> EVENT_BENEFICIARY(abonar,Ana)
//
if (MOD(\#1[SR-N1-beneficiary],\#2) & PREPD(\#2,?[lemma:"a favor de"]) & EVENT[other](\#1)
    & ~EVENT(\#1,\#2) & ~EVENT[beneficiary](\#1,?) )
  EVENT[beneficiary=+](\#1,\#2).

//================================================
// Cause
//================================================

// Exemplo: A casa resistiu a tempestade - EVENT_cause(resistir,tempestade)
//
if (MOD(\#1[SR-N1-cause],\#2) & PREPD(\#2,?[lemma:a]) & EVENT[other](\#1) & ~EVENT(\#1,\#2) )
  EVENT[cause=+](\#1,\#2).

//================================================
// Co-agent
//================================================
```

```
// Exemplo: O Ze casou com a Ana - EVENT_co-agent(casou,Ana)
//
if (MOD(\#1[SR-N1-co-agent],\#2) & PREPD(\#2,?[lemma:com])
& EVENT[other](\#1) & ~EVENT(\#1,\#2) )
  EVENT[co-agent=+](\#1,\#2).

// Exemplo: O Ze separou-se da Ana - EVENT_co-agent(separar,Ana)
//
if (MOD(\#1[SR-N1-co-agent],\#2) & PREPD(\#2,?[lemma:de])
& EVENT[other](\#1) & ~EVENT(\#1,\#2)
    & ~EVENT[co-agent](\#1,?) )
  EVENT[co-agent=+](\#1,\#2).

//================================================
// co-agent-co-object
//================================================

// Exemplo: O Pedro complementa a Ana -> EVENT_co-agent(complementar,Ana)
//
if (CDIR(\#1[SR-N1-co-agent-co-object],\#2[UMB-Human]) & EVENT[other](\#1) & ~EVENT(\#1,\#2)
    & ~EVENT[co-agent](\#1,?) )
  EVENT[co-agent=+](\#1,\#2).

// Exemplo: O vinagre complementa o azeite -> EVENT_co-object(complementar,azeite)
//
if (CDIR(\#1[SR-N1-co-agent-co-object],\#2) & EVENT[other](\#1) & ~EVENT(\#1,\#2)
    & ~EVENT[co-object](\#1,?) )
  EVENT[co-object=+](\#1,\#2).

//================================================
// co-experiencer
//================================================

// Exemplo: O Pedro da-se com a Ana -> EVENT_co-experiencer(dar,Ana)

if (MOD(\#1[SR-N1-co-experiencer],\#2[UMB-Human]) & PREPD(\#2,?[lemma:com]) & EVENT[other](\#1)
    & ~EVENT(\#1,\#2) )
  EVENT[co-experiencer=+](\#1,\#2).

//================================================
// co-locative
//================================================

// Exemplo: O terreno do Pedro confronta com o terreno do Joao
-> EVENT_co-locative(confrontar,terreno)

if (MOD(\#1[SR-N1-co-locative],\#2) & PREPD(\#2,?[lemma:com])
& EVENT[other](\#1) & ~EVENT(\#1,\#2) )
  EVENT[co-locative=+](\#1,\#2).

//================================================
```

```
// co-object
//================================================

// Exemplo: Esta cor assemelha-se aquela cor -> EVENT_co-object(assemelhar,cor)
//
if (MOD(\#1[SR-N1-co-object],\#2) & PREPD(\#2,?[lemma:a]) & EVENT[other](\#1) & ~EVENT(\#1,\#2)
    & ~EVENT[co-object](\#1,?) )
  EVENT[co-object=+](\#1,\#2).

// Exemplo:
//
if (MOD[dat](\#1[SR-N1-co-object],\#2) & EVENT[other](\#1)
& ~EVENT(\#1,\#2) & ~EVENT[co-object](\#1,?) )
  EVENT[co-object=+](\#1,\#2).

// Exemplo: Esta comida liga bem com aquele vinho -> EVENT_co-object(ligar,vinho)
//
if (MOD(\#1[SR-N1-co-object],\#2) & PREPD(\#2,?[lemma:com]) & EVENT[other](\#1) & ~EVENT(\#1,\#2)
    & ~EVENT[co-object](\#1,?) )
  EVENT[co-object=+](\#1,\#2).

// Exemplo: Este quadro difere do outro quadro -> EVENT_co-object(diferir,quadro)
//
if (MOD(\#1[SR-N1-co-object],\#2) & PREPD(\#2,?[lemma:de]) & EVENT[other](\#1) & ~EVENT(\#1,\#2)
    & ~EVENT[co-object](\#1,?) )
  EVENT[co-object=+](\#1,\#2).

//================================================
// co-occurrence
//================================================

// Exemplo: A vinda da Ana coincidiu com a vinda do Ze -> EVENT_co-occurrence(coincidir,vinda)
//
if (MOD(\#1[SR-N1-co-occurrence],\#2) & PREPD(\#2,?[lemma:com])
& EVENT[other](\#1) & ~EVENT(\#1,\#2) )
  EVENT[co-occurrence=+](\#1,\#2).


//================================================
// experiencer-gen
//================================================

// Exemplo: A not\'icia abalou o Pedro/abalou-o -> EVENT_experiencer-gen(abalar,Pedro)
//
if ( (CDIR(\#1[SR-N1-experiencer-gen],\#2[UMB-Human]) | CDIR(\#1[SR-N1-experiencer-gen],\#2[acc])) &
     EVENT[other](\#1) & ~EVENT(\#1,\#2)
   )
  EVENT[experiencer-generic=+](\#1,\#2).

// Exemplo: A novidade apraz ao Pedro -> EVENT_experiencer-gen(apraz,Pedro)
//
if (CINDIR(\#1[SR-N1-experiencer-gen],\#2[UMB-Human]) & PREPD(\#2,?[lemma:a]) & EVENT[other](\#1)
```

```
    & ~EVENT(\#1,\#2) & ~EVENT[experiencer-generic](\#1,?) )
  EVENT[experiencer-generic=+](\#1,\#2).

// Exemplo: As despesas pesam sobre o Pedro -> EVENT_experiencer-gen(pesar,Pedro)
//
if (MOD(\#1[SR-N1-experiencer-gen],\#2[UMB-Human]) & PREPD(\#2,?[lemma:sobre]) & EVENT[other](\#1)
    & ~EVENT(\#1,\#2) & ~EVENT[experiencer-generic](\#1,?) )
  EVENT[experiencer-generic=+](\#1,\#2).

// Exemplo: Fazer isso coube ao Pedro -> EVENT_experiencer-gen(caber,Pedro)
//
if (MOD(\#1[SR-N1-experiencer-gen],\#2[UMB-Human]) & PREPD(\#2,?[lemma:a]) & EVENT[other](\#1)
    & ~EVENT(\#1,\#2) & ~EVENT[experiencer-generic](\#1,?) )
  EVENT[experiencer-generic=+](\#1,\#2).

//================================================
// instrument
//================================================

// Exemplo: O Pedro percutiu com a pedra de encontro ao muro -> EVENT_instrument(apontar,faca)
//
if (MOD(\#1,\#2[SR-N1-instrument]) & PREPD(\#2,?[lemma:com]) & EVENT[other](\#1) & ~EVENT(\#1,\#2) )
  EVENT[instrument=+](\#1,\#2).

// Exemplo: O Pedro falou ao telefone com a Ana -> EVENT_instrument(apontar,faca)
//
if (MOD(\#1,\#2[SR-N1-instrument]) & PREPD(\#2,?[lemma:a]) & EVENT[other](\#1) & ~EVENT(\#1,\#2) )
  EVENT[instrument=+](\#1,\#2).

// Exemplo: O Pedro apontou uma faca ao Joao  -->  EVENT_instrument(apontar,faca)
//
if (CDIR(\#1[SR-N1-instrument],\#2) & EVENT[other](\#1) & ~EVENT(\#1,\#2) )
  EVENT[instrument=+](\#1,\#2).

//================================================
// locative-place
//================================================

// Exemplo: O Pedro ocupou o terreno -> EVENT_locative-place(ocupar,terreno)
//
if (CDIR(\#1[SR-N1-locative-place],\#2) & EVENT[other](\#1) & ~EVENT(\#1,\#2) )
  EVENT[locative-place=+](\#1,\#2).

//================================================
// message
//================================================

// Exemplo: O Pedro ordenou a morte do plebeu -> EVENT_message(ordenou,morte)
//
if (CDIR(\#1[SR-N1-message],\#2) & EVENT[other](\#1) & ~EVENT(\#1,\#2) )
  EVENT[message=+](\#1,\#2).
```

```
//================================================
// object-gen
//================================================

// Se um verbo tiver tra\c{c}o SR-N1-object-gen, e se houver um CDIR  ...
//
// Exemplo: O Pedro abalroou outro carro -> EVENT_object-gen(abalroou,carro)
//
if (CDIR(\#1[SR-N1-object-gen],\#2[time:~,location:~]) & EVENT[other](\#1) & ~EVENT(\#1,\#2)
    & ~EVENT[object-generic](\#1,?) )
  EVENT[object-generic=+](\#1,\#2).

// Exemplo: O Pedro jogou as damas -> EVENT_object-gen(jogou,damas)
//
if (MOD(\#1[SR-N1-object-gen],\#2[time:~,location:~]) & PREPD(\#2,?[lemma:a]) & EVENT[other](\#1)
    & ~EVENT(\#1,\#2) & ~EVENT[object-generic](\#1,?) )
  EVENT[object-generic=+](\#1,\#2).

// Exemplo: O Pedro acabou com o pudim -> EVENT_object-gen(acabar,pudim)
//
if (MOD(\#1[SR-N1-object-gen],\#2[time:~,location:~]) & PREPD(\#2,?[lemma:com]) & EVENT[other](\#1)
    & ~EVENT(\#1,\#2) & ~EVENT[object-generic](\#1,?) )
  EVENT[object-generic=+](\#1,\#2).

// Exemplo: O Pedro abdicou do trono -> EVENT_object-gen(abdicar,trono)
//
if (MOD(\#1[SR-N1-object-gen],\#2[time:~,location:~]) & PREPD(\#2,?[lemma:de]) & EVENT[other](\#1)
    & ~EVENT(\#1,\#2) & ~EVENT[object-generic](\#1,?) )
  EVENT[object-generic=+](\#1,\#2).

// Exemplo: O Pedro acertou na resposta -> EVENT_object-gen(acertar,resposta)
//
if (MOD(\#1[SR-N1-object-gen],\#2[time:~,location:~]) & PREPD(\#2,?[lemma:em]) & EVENT[other](\#1)
    & ~EVENT(\#1,\#2) & ~EVENT[object-generic](\#1,?) )
  EVENT[object-generic=+](\#1,\#2).

// Exemplo: O Pedro apurou-se para o Mundial -> EVENT_object-gen(apurar,Mundial)
//
if (MOD(\#1[SR-N1-object-gen],\#2[time:~,location:~]) & PREPD(\#2,?[lemma:para]) & EVENT[other](\#1)
    & ~EVENT(\#1,\#2) & ~EVENT[object-generic](\#1,?) )
  EVENT[object-generic=+](\#1,\#2).

// Exemplo: O Pedro age contra a natureza -> EVENT_object-gen(agir,natureza)
//
if (MOD(\#1[SR-N1-object-gen],\#2[time:~,location:~]) & PREPD(\#2,?[lemma:contra]) & EVENT[other](\#1)
    & ~EVENT(\#1,\#2) & ~EVENT[object-generic](\#1,?) )
  EVENT[object-generic=+](\#1,\#2).

// Exemplo: O Pedro conta-se entre um dos alunos mais brilhantes -> EVENT_object-gen(contar,alunos)
//
if (MOD(\#1[SR-N1-object-gen],\#2[time:~,location:~]) & PREPD(\#2,?[lemma:entre]) & EVENT[other](\#1)
    & ~EVENT(\#1,\#2) & ~EVENT[object-generic](\#1,?) )
```

```
  EVENT[object-generic=+](\#1,\#2).

//================================================
// object-cl
//================================================

// Exemplo: O cao abana a cauda -> EVENT_object-cl(abana,cauda)
//
if (CDIR(\#1[SR-N1-object-cl],\#2[UMB-Anatomical]) & EVENT[other](\#1) & ~EVENT(\#1,\#2) )
  EVENT[object-cl=+](\#1,\#2).


//================================================
// object-f
//================================================

// Se um verbo tiver tra\c{c}o SR-N1-object-f, e se houver um CDIR  ...
//
// Exemplo: O Pedro adivinhou a idade da Ana -> EVENT_object-f(adivinhou,idade)
//
if (CDIR(\#1[SR-N1-object-f],\#2) & EVENT[other](\#1) & ~EVENT(\#1,\#2) )
  EVENT[object-f=+](\#1,\#2).


// Exemplo: O Pedro apegou-se a companhia da Ana -> EVENT_object-f(apegou,companhia)
//
if (MOD(\#1[SR-N1-object-f],\#2) & PREPD(\#2,?[lemma:a]) & EVENT[other](\#1) & ~EVENT(\#1,\#2)
    & ~EVENT[object-f](\#1,?) )
  EVENT[object-f=+](\#1,\#2).

// Exemplo: O Pedro conformou-se com a decisao -> EVENT_object-f(conformou,decisao)
//
if (MOD(\#1[SR-N1-object-f],\#2) & PREPD(\#2,?[lemma:com]) & EVENT[other](\#1) & ~EVENT(\#1,\#2)
    & ~EVENT[object-f](\#1,?) )
  EVENT[object-f=+](\#1,\#2).

// Exemplo: O Pedro absteve-se de comentar -> EVENT_object-f(absteve,comentar)
//
if (MOD(\#1[SR-N1-object-f],\#2) & PREPD(\#2,?[lemma:de]) & EVENT[other](\#1) & ~EVENT(\#1,\#2)
    & ~EVENT[object-f](\#1,?) )
  EVENT[object-f=+](\#1,\#2).

// Exemplo: O Pedro acredita em extraterrestres -> EVENT_object-f(acredita,extraterrestres)
//
if (MOD(\#1[SR-N1-object-f],\#2) & PREPD(\#2,?[lemma:em]) & EVENT[other](\#1) & ~EVENT(\#1,\#2)
    & ~EVENT[object-f](\#1,?) )
  EVENT[object-f=+](\#1,\#2).

// Exemplo: O Pedro anseia pela reforma -> EVENT_object-f(anseia,reforma)
//
if (MOD(\#1[SR-N1-object-f],\#2) & PREPD(\#2,?[lemma:por]) & EVENT[other](\#1) & ~EVENT(\#1,\#2)
    & ~EVENT[object-f](\#1,?) )
```

```
  EVENT[object-f=+](\#1,\#2).

// Exemplo: O Pedro apareceu como o salvador da Patria -> EVENT_object-f(apareceu,salvador)
//
if (MOD(\#1[SR-N1-object-f],\#2) & PREPD(\#2,?[lemma:como]) & EVENT[other](\#1) & ~EVENT(\#1,\#2)
    & ~EVENT[object-f](\#1,?) )
  EVENT[object-f=+](\#1,\#2).

// Exemplo: Os manifestantes reclamaram contra o aumento do custo de vida
-> EVENT_object-f(reclamaram,aumento)
//
if (MOD(\#1[SR-N1-object-f],\#2) & PREPD(\#2,?[lemma:contra]) & EVENT[other](\#1) & ~EVENT(\#1,\#2)
    & ~EVENT[object-f](\#1,?) )
  EVENT[object-f=+](\#1,\#2).

// Exemplo: Os manifestantes reclamaram a favor do aumento do custo de vida
-> EVENT_object-f(reclamaram,aumento)
//
if (MOD(\#1[SR-N1-object-f],\#2) & PREPD(\#2,?[lemma:"a favor de"]) & EVENT[other](\#1) & ~EVENT(\#1,\#2)
    & ~EVENT[object-f](\#1,?) )
  EVENT[object-f=+](\#1,\#2).

// Exemplo: O Ze curvou-se perante a amea\c{c}a da Ana -> EVENT_object-f(curvou,amea\c{c}a)
//
if (MOD(\#1[SR-N1-object-f],\#2) & PREPD(\#2,?[lemma:perante]) & EVENT[other](\#1) & ~EVENT(\#1,\#2)
    & ~EVENT[object-f](\#1,?) )
  EVENT[object-f=+](\#1,\#2).

// Exemplo: O Pedro regula-se segundo o C\'odigo da Estrada -> EVENT_object-f(regula,C\'odigo)
//
if (MOD(\#1[SR-N1-object-f],\#2) & PREPD(\#2,?[lemma:segundo]) & EVENT[other](\#1) & ~EVENT(\#1,\#2)
    & ~EVENT[object-f](\#1,?) )
  EVENT[object-f=+](\#1,\#2).

//================================================
// object-q
//================================================

// Exemplo: O premio atingiu os 2 milh\~oes de euros -> EVENT_object-q(atingiu,euros)

if (CDIR(\#1[SR-N1-object-q],\#2) & EVENT[other](\#1) & ~EVENT(\#1,\#2) )
  EVENT[object-q=+](\#1,\#2).


// Exemplo: A d\'ivida ascende a 10 mil euros -> EVENT_object-q(ascende,euros)

if (MOD(\#1[SR-N1-object-q],\#2) & PREPD(\#2,?[lemma:a]) & EVENT[other](\#1) & ~EVENT(\#1,\#2)
    & ~EVENT[object-q](\#1,?) )
  EVENT[object-q=+](\#1,\#2).

// Exemplo: O Estado comparticipa com 100$00 neste medicamento
-> EVENT_object-q(comparticipa,euros)
```

```
if (MOD(\#1[SR-N1-object-f],\#2) & PREPD(\#2,?[lemma:com]) & EVENT[other](\#1) & ~EVENT(\#1,\#2)
    & ~EVENT[object-q](\#1,?) )
  EVENT[object-q=+](\#1,\#2).

// Exemplo: O \'indice acabou nos 1000 pontos -> EVENT_object-q(acabou,pontos)

if (MOD(\#1[SR-N1-object-f],\#2) & PREPD(\#2,?[lemma:em]) & EVENT[other](\#1) & ~EVENT(\#1,\#2)
    & ~EVENT[object-q](\#1,?) )
  EVENT[object-q=+](\#1,\#2).

//================================================
// occurrence
//================================================

// Exemplo: A relva amparou a queda -> EVENT_occurrence(amparou,queda)

if (CDIR(\#1[SR-N1-occurrence],\#2) & EVENT[other](\#1) & ~EVENT(\#1,\#2) )
  EVENT[occurrence=+](\#1,\#2).

// Exemplo: O aumento do petr\'oleo conduz a crise -> EVENT_occurrence(conduz,crise)

if (MOD(\#1[SR-N1-occurrence],\#2) & PREPD(\#2,?[lemma:a]) & EVENT[other](\#1) & ~EVENT(\#1,\#2)
    & ~EVENT[occurrence](\#1,?) )
  EVENT[occurrence=+](\#1,\#2).

// Exemplo: A vinda do Joao prende-se com a necessidade de fazer isso
-> EVENT_occurrence(prende,necessidade)

if (MOD(\#1[SR-N1-occurrence],\#2) & PREPD(\#2,?[lemma:com]) & EVENT[other](\#1) & ~EVENT(\#1,\#2)
    & ~EVENT[occurrence](\#1,?) )
  EVENT[occurrence=+](\#1,\#2).

// Exemplo: Esse problema advem do facto do Pedro nao lavar as maos
-> EVENT_occurrence(advem,facto)

if (MOD(\#1[SR-N1-occurrence],\#2) & PREPD(\#2,?[lemma:de]) & EVENT[other](\#1) & ~EVENT(\#1,\#2)
    & ~EVENT[occurrence](\#1,?) )
  EVENT[occurrence=+](\#1,\#2).

// Exemplo: A celebra\c{c}ao da cerim\'onia dos \'oscares culminou na entrega do premio ao melhor filme
//          -> EVENT_occurrence(culminar,entrega)

if (MOD(\#1[SR-N1-occurrence],\#2) & PREPD(\#2,?[lemma:em]) & EVENT[other](\#1) & ~EVENT(\#1,\#2)
    & ~EVENT[occurrence](\#1,?) )
  EVENT[occurrence=+](\#1,\#2).

// Exemplo: O Pedro contribuiu para isso -> EVENT_occurrence(contribuir,isso)

if (MOD(\#1[SR-N1-occurrence],\#2) & PREPD(\#2,?[lemma:para]) & EVENT[other](\#1) & ~EVENT(\#1,\#2)
    & ~EVENT[occurrence](\#1,?) )
  EVENT[occurrence=+](\#1,\#2).
```

```
//================================================
// patient
//================================================

// Exemplo: O Joao abordou a Ana -> EVENT_patient(abordar,Ana)

if (CDIR(\#1[SR-N1-patient],\#2[UMB-Human]) & EVENT[other](\#1)
    & ~EVENT(\#1,\#2) & ~EVENT[patient](\#1,?) )
  EVENT[patient=+](\#1,\#2).

// RT 31/07/2013 nao sei se pode ser
// Exemplo: O Pedro acorrentou o cao ao poste -> EVENT_patient(acorrentar,cao)
if (CDIR(\#1[SR-N1-patient],\#2) & EVENT[other](\#1) & ~EVENT(\#1,\#2) & ~EVENT[patient](\#1,?) )
  EVENT[patient=+](\#1,\#2).

//Exemplo: O Pedro telefonou a Ana -> EVENT_addressee(telefonar,Ana)
//
if (CINDIR(\#1[SR-N1-patient],\#2[UMB-Human]) & PREPD(\#2,?[lemma:a]) & EVENT[other](\#1)
    & ~EVENT(\#1,\#2) & ~EVENT[patient](\#1,?) )
  EVENT[patient=+](\#1,\#2).

// Exemplo: O Pedro acenou ao Joao - EVENT_patient(acenar,Joao)

if (MOD(\#1[SR-N1-patient],\#2[UMB-Human]) & PREPD(\#2,?[lemma:a]) & EVENT[other](\#1)
& ~EVENT(\#1,\#2) & ~EVENT[patient](\#1,?) )
  EVENT[patient=+](\#1,\#2).

// Exemplo: O Pedro aconselhou-se com o seu av\^o -> EVENT_patient(aconselhar,av\^o)

if (MOD(\#1[SR-N1-patient],\#2[UMB-Human]) & PREPD(\#2,?[lemma:com]) & EVENT[other](\#1)
& ~EVENT(\#1,\#2) & ~EVENT[patient](\#1,?) )
  EVENT[patient=+](\#1,\#2).

// Exemplo: O Pedro acorrentou o cao ao poste -> EVENT_patient(acorrentar,cao)

if (CDIR(\#1[SR-N1-patient],\#2) & EVENT[other](\#1) & ~EVENT(\#1,\#2) & ~EVENT[patient](\#1,?) )
  EVENT[patient=+](\#1,\#2).

// Exemplo: O Pedro abusou da Ana -> EVENT_patient(abusar,Ana)

if (MOD(\#1[SR-N1-patient],\#2[UMB-Human])
& PREPD(\#2,?[lemma:de]) & EVENT[other](\#1) & ~EVENT(\#1,\#2)
    & ~EVENT[patient](\#1,?) )
  EVENT[patient=+](\#1,\#2).

// Exemplo: O Pedro bateu no ladrao -> EVENT_patient(bateu,ladrao)

if (MOD(\#1[SR-N1-patient],\#2[UMB-Human])
& PREPD(\#2,?[lemma:em]) & EVENT[other](\#1) & ~EVENT(\#1,\#2)
    & ~EVENT[patient](\#1,?) )
  EVENT[patient=+](\#1,\#2).
```

```
// Exemplo: O Pedro mostrou-se da Ana -> EVENT_patient(mostrar,Ana)

if (MOD(\#1[SR-N1-patient],\#2[UMB-Human])
& PREPD(\#2,?[lemma:para]) & EVENT[other](\#1) & ~EVENT(\#1,\#2)
    & ~EVENT[patient](\#1,?) )
  EVENT[patient=+](\#1,\#2).

// Exemplo: O Pedro remou pela Ana -> EVENT_patient(remar,Ana)

if (MOD(\#1[SR-N1-patient],\#2[UMB-Human])
& PREPD(\#2,?[lemma:por]) & EVENT[other](\#1) & ~EVENT(\#1,\#2)
    & ~EVENT[patient](\#1,?) )
  EVENT[patient=+](\#1,\#2).

// Exemplo: O exercito portugu\^es carregou sobre os inimigos - EVENT_patient(carregar,inimigos)

 if (MOD(\#1[SR-N1-patient],\#2[UMB-Human])
 & PREPD(\#2,?[lemma:sobre]) & EVENT[other](\#1)
    & ~EVENT(\#1,\#2) & ~EVENT[patient](\#1,?) )
  EVENT[patient=+](\#1,\#2).

// Exemplo: O exercito portugu\^es carregou contra os inimigos - EVENT_patient(carregar,inimigos)

if (MOD(\#1[SR-N1-patient],\#2[UMB-Human]) & PREPD(\#2,?[lemma:contra]) & EVENT[other](\#1)
    & ~EVENT(\#1,\#2) & ~EVENT[patient](\#1,?) )
  EVENT[patient=+](\#1,\#2).

// Exemplo: O Pedro testemunhou a favor do Joao -> EVENT_patient(testemunhar,Joao)

if (MOD(\#1[SR-N1-patient],\#2[UMB-Human]) & PREPD(\#2,?[lemma:"a favor de"]) & EVENT[other](\#1)
    & ~EVENT(\#1,\#2) & ~EVENT[patient](\#1,?) )
  EVENT[patient=+](\#1,\#2).

// Exemplo: O Pedro confessou-se junto da Ana -> EVENT_patient(confessar,Ana)

if (MOD(\#1[SR-N1-patient],\#2[UMB-Human]) & PREPD(\#2,?[lemma:"junto de"]) & EVENT[other](\#1)
    & ~EVENT(\#1,\#2) & ~EVENT[patient](\#1,?) )
  EVENT[patient=+](\#1,\#2).

// Exemplo: O Pedro confessou-se junto a Ana -> EVENT_patient(confessar,Ana)

if (MOD(\#1[SR-N1-patient],\#2[UMB-Human]) & PREPD(\#2,?[lemma:"junto a"]) & EVENT[other](\#1)
    & ~EVENT(\#1,\#2) & ~EVENT[patient](\#1,?) )
  EVENT[patient=+](\#1,\#2).

// Exemplo: O Ze desculpou-se perante a Ana - EVENT_patient(desculpar,Ana)

if (MOD(\#1[SR-N1-patient],\#2[UMB-Human]) & PREPD(\#2,?[lemma:perante]) & EVENT[other](\#1)
    & ~EVENT(\#1,\#2) & ~EVENT[patient](\#1,?) )
  EVENT[patient=+](\#1,\#2).
```

```
// Exemplo: O Ze desculpou-se diante da Ana - EVENT_patient(desculpar,Ana)

if (MOD(\#1[SR-N1-patient],\#2[UMB-Human]) & PREPD(\#2,?[lemma:"diante de"]) & EVENT[other](\#1)
    & ~EVENT(\#1,\#2) & ~EVENT[patient](\#1,?) )
  EVENT[patient=+](\#1,\#2).

//================================================
// patient-object
//================================================

// Exemplo: O Pedro alvejou o Joao -> EVENT_patient(alvejar,Joao)

if (CDIR(\#1[SR-N1-patient-object],\#2[UMB-Human]) & EVENT[other](\#1) & ~EVENT(\#1,\#2) )
  EVENT[patient=+](\#1,\#2).

// Exemplo: O Joao alvejou a pedra -> EVENT_object(alvejou,pedra)

if (CDIR(\#1[SR-N1-patient-object],\#2) & EVENT[other](\#1) & ~EVENT(\#1,\#2) )
  EVENT[object-generic=+](\#1,\#2).




//================================================
// patient-object-cl
//================================================

// Falta atualizar o ViPEr
// Exemplo: O Pedro penteou o Joao -> EVENT_patient(penteou,Joao)

if (CDIR(\#1[SR-N1-patient-object-cl],\#2[UMB-Human]) & EVENT[other](\#1) & ~EVENT(\#1,\#2) )
  EVENT[patient=+](\#1,\#2).

// Exemplo: O Joao penteou o cabelo -> EVENT_object-cl(penteou,cabelo)

if (CDIR(\#1[SR-N1-patient-object-cl],\#2[UMB-Anatomical]) & EVENT[other](\#1) & ~EVENT(\#1,\#2) )
  EVENT[object-cl=+](\#1,\#2).


//================================================
// possessor
//================================================

// Exemplo: O Joao pertence a Comissao Organizadora
//            -> EVENT_possessor(pertence,Comissao Organizadora)
//
if (MOD(\#1[SR-N1-possessor],\#2) & PREPD(\#2,?[lemma:a]) & EVENT[other](\#1) & ~EVENT(\#1,\#2) )
  EVENT[possessor=+](\#1,\#2).

// Falta regra para POSS inalienavel.  Exemplo: As canelas do Joao.
// Nao se pode marcar em ``a cor da pedra", ``pedra" como possessor.
```

```
//================================================
// recipient
//================================================

// Exemplo: As recolhas revertem a Ana -> EVENT_recipient(reverter,Banco Alimentar)

if (MOD(\#1[SR-N1-recipient],\#2[UMB-Human]) & PREPD(\#2,?[lemma:a]) & EVENT[other](\#1)
    & ~EVENT(\#1,\#2) )
  EVENT[recipient=+](\#1,\#2).

// Exemplo: Os fundos revertem para o Banco Alimentar
-> EVENT_recipient(reverter,Banco Alimentar)

if (MOD(\#1[SR-N1-recipient],\#2[UMB-Human]) & PREPD(\#2,?[lemma:para]) & EVENT[other](\#1)
    & ~EVENT(\#1,\#2) )
  EVENT[recipient=+](\#1,\#2).

// Exemplo: Os fundos revertem a favor do Banco Alimentar
-> EVENT_recipient(reverter,Banco Alimentar)

if (MOD(\#1[SR-N1-recipient],\#2[UMB-Human]) & PREPD(\#2,?[lemma:"a favor de"]) & EVENT[other](\#1)
    & ~EVENT(\#1,\#2) )
  EVENT[recipient=+](\#1,\#2).

//================================================
// topic
//================================================

// Exemplo: O Joao discutiu esse assunto com a Ana -> EVENT_topic(discutir,assunto)

if (CDIR(\#1[SR-N1-topic],\#2) & EVENT[other](\#1) & ~EVENT(\#1,\#2) )
  EVENT[topicsr=+](\#1,\#2).

// Exemplo: O Paulo percebe de vinhos -> EVENT_topic(perceber,vinhos)

if (MOD(\#1[SR-N1-topic],\#2) & PREPD(\#2,?[lemma:de]) & EVENT[other](\#1) & ~EVENT(\#1,\#2)
    & ~EVENT[topic](\#1,?) )
  EVENT[topicsr=+](\#1,\#2).

// Exemplo: Os educadores influem na educa\c{c}ao das crian\c{c}as -> EVENT_topic(influir,educa\c{c}ao)

if (MOD(\#1[SR-N1-topic],\#2) & PREPD(\#2,?[lemma:em]) & EVENT[other](\#1) & ~EVENT(\#1,\#2)
    & ~EVENT[topic](\#1,?) )
  EVENT[topicsr=+](\#1,\#2).

// Exemplo: O Paulo fala sobre vinhos -> EVENT_topic(falar,vinhos)

if (MOD(\#1[SR-N1-topic],\#2) & PREPD(\#2,?[lemma:sobre]) & EVENT[other](\#1) & ~EVENT(\#1,\#2)
    & ~EVENT[topic](\#1,?) )
  EVENT[topicsr=+](\#1,\#2).

//================================================================================
```

```
//================================================================================
// Converting N2 Syntactic dependencies into Semantic Roles
//================================================================================
//================================================================================

//RT 2013-07-23 added all

//================================================
// Addressee
//================================================

//if (MOD(\#1[SR-N2-addressee],\#2[UMB-Human]) & PREPD(\#2,?[lemma:a])) &
// EVENT[other](\#1) & ~EVENT(\#1,\#2) & ~EVENT[addressee](\#1,?) )
// EVENT[addressee=+](\#1,\#2).

if (CINDIR(\#1[SR-N2-addressee],\#2[UMB-Human]) & EVENT[other](\#1) & ~EVENT(\#1,\#2)
    & ~EVENT[addressee](\#1,?) )
  EVENT[addressee=+](\#1,\#2).

if (MOD(\#1[SR-N2-addressee],\#2[UMB-Human]) & PREPD(\#2,?[lemma:a]) & EVENT[other](\#1) & ~EVENT(\#1,\#2)
    & ~EVENT[addressee](\#1,?) )
  EVENT[addressee=+](\#1,\#2).

//================================================
// Agent-giver
//================================================

//if (MOD(\#1[SR-N2-agent-giver],\#2[UMB-Human]) & (PREPD(\#2,?[lemma:a]) | PREPD(\#2,?[lemma:de])) &
// EVENT[other](\#1) & ~EVENT(\#1,\#2) & ~EVENT[agent-giver](\#1,?) )
// EVENT[agent-giver=+](\#1,\#2).

//foi mudada para antes dos N1

//================================================
// Beneficiary
//================================================

if (MOD(\#1[SR-N2-beneficiary],\#2[UMB-Human,time:~,location:~]) & PREPD(\#2,?[lemma:a])
    & EVENT[other](\#1) & ~EVENT(\#1,\#2) & ~EVENT[beneficiary](\#1,?) )
EVENT[beneficiary=+](\#1,\#2).


//================================================
// Co-agent
//================================================

// Exemplo: O Pedro apostou isso com o Joao -> EVENT_co-agent(apostou,Joao)
//
if (MOD(\#1[SR-N2-co-agent],\#2[UMB-Human]) & PREPD(\#2,?[lemma:a]) & EVENT[other](\#1) & ~EVENT(\#1,\#2)
    & ~EVENT[co-agent](\#1,?) )
  EVENT[co-agent=+](\#1,\#2).
```

```
//=============================================
// co-agent-co-object
//=============================================

// Exemplo: O Pedro equiparou o Joao a Hitler -> EVENT_co-agent(equiparou,Hitler)
//
if (MOD(\#1[SR-N2-co-agent-co-object],\#2[UMB-Human]) & ( PREPD(\#2,?[lemma:a]) ||
    PREPD(\#2,?[lemma:de]) || PREPD(\#2,?[lemma:com]) ) & EVENT[other](\#1) & ~EVENT(\#1,\#2)
    & ~EVENT[co-agent](\#1,?) )
  EVENT[co-agent=+](\#1,\#2).

// Exemplo: O Pedro uniu uma metade a outra metade -> EVENT_co-object(uniu,metade)
//
if (MOD(\#1[SR-N2-co-agent-co-object],\#2) & ( PREPD(\#2,?[lemma:a]) || PREPD(\#2,?[lemma:de]) ||
    PREPD(\#2,?[lemma:com]) ) & EVENT[other](\#1) & ~EVENT(\#1,\#2) & ~EVENT[co-object](\#1,?) )
  EVENT[co-object=+](\#1,\#2).


//=============================================
// Co-patient
//=============================================

// Exemplo: O padre casou o Pedro e a Ana  -->  EVENT_co-patient(casar,Ana)
//
if (MOD(\#1[SR-N2-co-patient],\#2[UMB-Human]) & PREPD(\#2,?[lemma:com]) & EVENT[other](\#1)
    & ~EVENT(\#1,\#2) )
  EVENT[co-patient=+](\#1,\#2).

// Exemplo: O padre divorciou o Pedro da Ana  -->  EVENT_co-patient(divorciar,Ana)
//
if (MOD(\#1[SR-N2-co-patient],\#2[UMB-Human]) & PREPD(\#2,?[lemma:de]) & EVENT[other](\#1)
    & ~EVENT(\#1,\#2) )
  EVENT[co-patient=+](\#1,\#2).

// Exemplo: O padre trocou a Joana pela Ana  -->  EVENT_co-patient(trocar,Ana)

if (MOD(\#1[SR-N2-co-patient],\#2[UMB-Human]) & PREPD(\#2,?[lemma:por]) & EVENT[other](\#1)
    & ~EVENT(\#1,\#2) )
  EVENT[co-patient=+](\#1,\#2).

// Exemplo: O Ze destacou a Ana entre os outros alunos da turma
//          -->  EVENT_co-patient(destacar,alunos)
//
if (MOD(\#1[SR-N2-co-patient],\#2[UMB-Human]) & PREPD(\#2,?[lemma:entre]) & EVENT[other](\#1)
    & ~EVENT(\#1,\#2) )
  EVENT[co-patient=+](\#1,\#2).

// Exemplo: O Ze destacou a Ana de entre os outros alunos da turma
//          -->  EVENT_co-patient(destacar,alunos)
//
if (MOD(\#1[SR-N2-co-patient],\#2[UMB-Human]) & PREPD(\#2,?[lemma:"de entre"]) & EVENT[other](\#1)
    & ~EVENT(\#1,\#2) )
```

```
  EVENT[co-patient=+](\#1,\#2).


//================================================
// Co-object
//================================================

// Exemplo: O Pedro juntou agua aos ovos -> EVENT_co-object(juntou,ovos)
//
if (MOD(\#1[SR-N2-co-object],\#2) & PREPD(\#2,?[lemma:a]) & EVENT[other](\#1) & ~EVENT(\#1,\#2)
    & ~EVENT[co-object](\#1,?) )
  EVENT[co-object=+](\#1,\#2).

// Exemplo: O Joao juntou a farinha com os ovos  -->  EVENT_co-object(juntar,ovos)
//
if (MOD(\#1[SR-N2-co-object],\#2[time:~,location:~]) & PREPD(\#2,?[lemma:com]) & EVENT[other](\#1)
    & ~EVENT(\#1,\#2) )
  EVENT[co-object=+](\#1,\#2).

// Exemplo: O Joao separou a farinha dos ovos  -->  EVENT_co-object(adicionar,ovos)
//
if (MOD(\#1[SR-N2-co-object],\#2[time:~,location:~]) & PREPD(\#2,?[lemma:de]) & EVENT[other](\#1)
    & ~EVENT(\#1,\#2) )
  EVENT[co-object=+](\#1,\#2).

// Exemplo: O Joao adicionou a farinha nos ovos  -->  EVENT_co-object(adicionar,ovos)
//
if (MOD(\#1[SR-N2-co-object],\#2[time:~,location:~]) & PREPD(\#2,?[lemma:em]) & EVENT[other](\#1)
    & ~EVENT(\#1,\#2) )
  EVENT[co-object=+](\#1,\#2).

// Exemplo: O Joao adicionou atirou farinha para os ovos  -->  EVENT_co-object(adicionar,ovos)
//
if (MOD(\#1[SR-N2-co-object],\#2[time:~,location:~]) & PREPD(\#2,?[lemma:para]) & EVENT[other](\#1)
    & ~EVENT(\#1,\#2) )
  EVENT[co-object=+](\#1,\#2).


//================================================
// Message
//================================================
// Sai varias vezes indevidamente/ Tem de ser revista
// Exemplo: Os candidatos apelaram aos cidadaos para que fossem votar.
-> EVENT_message(apelaram,para)
//
if (MOD(\#1[SR-N2-message],\#2[time:~,location:~]) & ( PREPD(\#2,?[lemma:em]) ||
    PREPD(\#2,?[lemma:sobre]) || PREPD(\#2,?[lemma:para]) || PREPD(\#2,?[lemma:"em como"]) ||
    PREPD(\#2,?[lemma:de]) ) & EVENT[other](\#1) & ~EVENT(\#1,\#2) & ~EVENT[message](\#1,?) )
  EVENT[message=+](\#1,\#2).


//================================================
```

```
// Object-gen
//================================================

// Exemplo: O Pedro trocou euros por d\'olares -> EVENT_object-gen(trocar,d\'olares)
//
if (MOD(\#1[SR-N2-object-gen],\#2[time:~,location:~]) & PREPD(\#2,?[lemma:por]) & EVENT[other](\#1)
    & ~EVENT(\#1,\#2) & ~EVENT[object-generic](\#1,?) )
  EVENT[object-generic=+](\#1,\#2).

//================================================
// Object-f
//================================================

if (MOD(\#1[SR-N2-object-f],\#2[time:~,location:~]) & EVENT[other](\#1) & ~EVENT(\#1,\#2)
    & ~EVENT[object-f](\#1,?) )
  EVENT[object-f=+](\#1,\#2).

if (MOD(\#1[SR-N2-object-f],\#2[time:~,location:~]) & ( PREPD(\#2,?[lemma:de]) ||
    PREPD(\#2,?[lemma:a]) || PREPD(\#2,?[lemma:com]) || PREPD(\#2,?[lemma:em]) ||
    PREPD(\#2,?[lemma:para]) || PREPD(\#2,?[lemma:por]) || PREPD(\#2,?[lemma:"a partir de"]) ||
    PREPD(\#2,?[lemma:como]) ) & EVENT[other](\#1) & ~EVENT(\#1,\#2) & ~EVENT[topic](\#1,?)
    & ~EVENT[object-f](\#1,?) )
  EVENT[object-f=+](\#1,\#2).



//================================================
// Object-occurrence
//================================================

// Exemplo: A crise abrandou a economia -> EVENT_occurrence(abrandou,economia)

if (CDIR(\#1[SR-N1-object-occurrence],\#2[SEM-am]) || CDIR(\#1[SR-N1-object-occurrence],\#2[SEM-ac])
    & EVENT[other](\#1) & ~EVENT(\#1,\#2) & ~EVENT[occurrence](\#1,?) )
  EVENT[occurrence=+](\#1,\#2).


// Exemplo: Isso abrandou o Joao/carro -> EVENT_object-gen(abrandou,Joao/carro)

if (CDIR(\#1[SR-N1-object-occurrence],\#2[SEM-cc]) || CDIR(\#1[SR-N1-object-occurrence],\#2[UMB-Human])
    & EVENT[other](\#1) & ~EVENT(\#1,\#2) & ~EVENT[object-generic](\#1,?) )
  EVENT[object-generic=+](\#1,\#2).

//================================================
// Object-q
//================================================

// Exemplo: O vendedor abateu 50 euros ao pre\c{c}o inicial do carro -> EVENT_object-q(abateu,pre\c{c}o)

if (MOD(\#1[SR-N2-object-q],\#2) & ( PREPD(\#2,?[lemma:a]) ||  PREPD(\#2,?[lemma:em]) ||
    PREPD(\#2,?[lemma:com]) ) & EVENT[other](\#1) & ~EVENT(\#1,\#2) & ~EVENT[object-q](\#2,?) )
  EVENT[object-q=+](\#1,\#2).
```

```
//================================================
// Occurrence
//================================================

// Exemplo: O Joao culpabilizou a crise pelo seu div\'orcio -> EVENT_occurrence(culpabilizou,div\'orcio)
// Exemplo: O Joao conciliou o facto de ter um encontro com o facto de sair tarde
//          -->  EVENT_occurrence(conciliar,facto)

if (MOD(\#1[SR-N2-occurrence],\#2[sem-cc:~]) & ( PREPD(\#2,?[lemma:por]) ||
    PREPD(\#2,?[lemma:com]) ) & EVENT[other](\#1) & ~EVENT(\#1,\#2) & ~EVENT[occurrence](\#1,?) )
  EVENT[occurrence=+](\#1,\#2).

//================================================
// Patient
//================================================

// Exemplo: O Joao apontou uma faca ao pol\'icia -> EVENT_patient(apontou,pol\'icia)
// Exemplo: O Pedro remeteu as d\'uvidas para o seu acessor  --> EVENT_patient(remeter,assessor)
// Exemplo: O Pedro distribuiu panfletos pelos transeuntes --> EVENT_patient(distribuir,transeuntes)
// Exemplo: O Pedro atirar com um tijolo contra o Joao     --> EVENT_patient(atirar,Joao)

if (MOD(\#1[SR-N2-patient],\#2) & ( PREPD(\#2,?[lemma:a]) || PREPD(\#2,?[lemma:em])||
    PREPD(\#2,?[lemma:para])|| PREPD(\#2,?[lemma:por])|| PREPD(\#2,?[lemma:contra]) )
    & EVENT[other](\#1) & ~EVENT(\#1,\#2) & ~EVENT[patient](\#1,?) )
  EVENT[patient=+](\#1,\#2).

// Exemplo: O Pedro chamou mentiroso ao Joao  -->  EVENT_patient(chamar,Joao)
//
if (CINDIR(\#1[SR-N2-patient],\#2) & EVENT[other](\#1) & ~EVENT(\#1,\#2) )
  EVENT[patient=+](\#1,\#2).
//================================================
// Recipient
//================================================

// Exemplo: O Pedro deu o livro ao Joao -> EVENT_recipient(deu,Joao)

if (CINDIR(\#1[SR-N2-recipient],\#2) & PREPD(\#2,?[lemma:a]) & EVENT[other](\#1) & ~EVENT(\#1,\#2)
    & ~EVENT[recipient](\#1,?) )
  EVENT[recipient=+](\#1,\#2).

// Exemplo: O Pedro vinculou-se ao banco A -> EVENT_recipient(vinculou,banco A)

if (MOD(\#1[SR-N2-recipient],\#2) & ( PREPD(\#2,?[lemma:a]) || PREPD(\#2,?[lemma:para]) ||
    PREPD(\#2,?[lemma:em]) ) & EVENT[other](\#1) & ~EVENT(\#1,\#2) & ~EVENT[recipient](\#1,?) )
  EVENT[recipient=+](\#1,\#2).

//================================================
// Topic
//================================================

// Exemplo: Os investigadores questionaram a Ana acerca dos seus habitos ->
```

```
 EVENT_topic(questionaram,habitos)

if (MOD(\#1[SR-N2-topic],\#2) & ( PREPD(\#2,?[lemma:de]) || PREPD(\#2,?[lemma:sobre]) ||
    PREPD(\#2,?[lemma:"acerca de"]) ) & EVENT[other](\#1) & ~EVENT(\#1,\#2) & ~EVENT[topic](\#1,?) )
  EVENT[topicsr=+](\#1,\#2).


//================================================
// Victim
//================================================

if (MOD(\#1[SR-N2-victim],\#2[human:+]) & ( PREPD(\#2,?[lemma:contra]) || PREPD(\#2,?[lemma:a]) )
    & EVENT[other](\#1) & ~EVENT(\#1,\#2) )
  EVENT[victim=+](\#1,\#2).


if (MOD(\#1,\#2[human:+]) & ( PREPD(\#2,?[lemma:contra]) || PREPD(\#2,?[lemma:a]) ) & EVENT[other](\#1)
    & ~EVENT(\#1,\#2) )
  EVENT[victim=+](\#1,\#2).

//if (CDIR(\#1[SR-N2-victim],\#2[[human:+]]) & EVENT[other](\#1) & ~EVENT(\#1,\#2) )
//  EVENT[victim=+](\#1,\#2).

// Exemplo: O Pedro sonegou as provas a/da pol\'icia  -->  EVENT_victim(sonegar,pol\'icia)
//
if (CINDIR(\#1[SR-N2-victim],\#2) & EVENT[other](\#1) & ~EVENT(\#1,\#2) )
  EVENT[victim=+](\#1,\#2).

//*****************************************************************************
//*****************************************************************************
//***                                                                       ****
//***                             THIRD  PART                               ****
//***               (apply rules for LOCATIVE, TIME and MANNER)             ****
//***                                                                       ****
//*****************************************************************************
//*****************************************************************************

//Locative exceptions

// RT+JB 2013-07-28 ``aproximar-se de" : The Prep ``de" has feature preplocsource but expresses
// loc-destination with these verbs.
if (MOD(\#1[lemma:abeirar],\#2) || MOD(\#1[lemma:aproximar],\#2) || MOD(\#1[lemma:acercar],\#2)
    & PREPD(\#2,?[lemma:de]) & EVENT[other](\#1) & ~EVENT(\#1,\#2) )
  EVENT[locative-destination=+](\#1,\#2).


// Exemplo: O Pedro a\c{c}aimou o cao -> EVENT_object-generic(a\c{c}aimar,cao)
//
if (CDIR(\#1[SR-N1-object-cl],\#2[UMB-Animal]) & EVENT[other](\#1) & ~EVENT(\#1,\#2) )
  EVENT[object-generic=+](\#1,\#2).

//=========================================================
```

```
// Semantic Roles for LOCATIVE
//=======================================================

// If, in a PP, the preposition is a locative Prep, and if the head of the PP
// is a locative noun, then the semantic role of locative is attributed to the PP.
// The specific locative role (SOURCE, DESTINATATION, PLACE or PATH) is derived
// from the features of the preposition.
//
// Example: O Joao leu o livro em Lisboa   ->   EVENT_locative-place(leu,Lisboa)
//
// Example: O Pedro trouxe de Lisboa para o Porto atraves da Estremadura o livro e leu-o no Algarve
//                                  ->   EVENT_LOCATIVE-SOURCE(trouxe,Lisboa)
//                                  ->   EVENT_LOCATIVE-DESTINATION(trouxe,Porto)
//                                  ->   EVENT_LOCATIVE-PATH(trouxe,Estremadura)
//                                  ->   EVENT_LOCATIVE-PLACE(leu,Algarve)
//

//RT 2013-07-23 added umbrellas [geo] and [UMB-location] and [sem-cc] and [proxadv] and [sem-con:+]
if ( (MOD(\#1,\#2[geo]) || MOD(\#1,\#2[geo-toponym]) || MOD(\#1,\#2[UMB-location]) ||
    MOD(\#1,\#2[sem-cc]) || MOD(\#1,\#2[proxadv])  || MOD(\#1,\#2[sem-con:+]))
    & PREPD(\#2,?[preplocsource]) & EVENT[other](\#1) & ~EVENT(\#1,\#2) )
  EVENT[locative-source=+](\#1,\#2).

//RT 2013-07-23 added umbrellas [geo] and [UMB-location] and [sem-cc] and [proxadv] and [sem-con:+]
if ( (MOD(\#1,\#2[geo]) || MOD(\#1,\#2[geo-toponym]) || MOD(\#1,\#2[UMB-location]) ||
    MOD(\#1,\#2[sem-cc]) || MOD(\#1,\#2[proxadv])  || MOD(\#1,\#2[sem-con:+]))
    & PREPD(\#2,?[preplocdest]) & EVENT[other](\#1) & ~EVENT(\#1,\#2) )
  EVENT[locative-destination=+](\#1,\#2).

//if ( MOD(\#1,\#2) & PREPD(\#2[human:~],?[preplocdest]) & EVENT[other](\#1) & ~EVENT(\#1,\#2) )
//  EVENT[locative-destination=+](\#1,\#2).

//RT 2013-07-23 added umbrellas [geo] and [UMB-location] and [sem-cc] and [proxadv] and [sem-con:+]
if ( (MOD(\#1,\#2[geo]) || MOD(\#1,\#2[geo-toponym]) || MOD(\#1,\#2[UMB-location]) ||
    MOD(\#1,\#2[sem-cc]) || MOD(\#1,\#2[proxadv])  || MOD(\#1,\#2[sem-con:+])) & PREPD(\#2,?[preplocpath])
    & EVENT[other](\#1) & ~EVENT(\#1,\#2) )
  EVENT[locative-path=+](\#1,\#2).

//RT 2013-07-23 added umbrellas [geo] and [UMB-location] and [sem-cc] and [proxadv] and [sem-con:+]
if ( (MOD(\#1,\#2[geo]) || MOD(\#1,\#2[geo-toponym]) || MOD(\#1,\#2[UMB-location]) ||
    MOD(\#1,\#2[sem-cc]) || MOD(\#1,\#2[proxadv])  || MOD(\#1,\#2[sem-con:+])) & PREPD(\#2,?[preplocplace]
    & EVENT[other](\#1) & ~EVENT(\#1,\#2) )
  EVENT[locative-place=+](\#1,\#2).

//============================
// TOP node modifiers
//============================

//NJM+JB,2013/04/01
// TOP node modifiers are associated with all the events!!!

if ( MOD(?[cat:0],\#2[location]) & PREPD(\#2,?[preplocplace]) & EVENT[other](\#1) & ~EVENT(\#1,\#2) )
```

```
  EVENT[locative-place=+](\#1,\#2).

// Note: In the case of locative NPs in subject or direct object position,
// the semantic role of locative can only be ascertained by refering to the lexical
// information about SRs encoded in the verb the NP depends on (mainly viper classes 38Lx).

//=========================================================
// Semantic Roles for TIME
//=========================================================

// Example: O Pedro fez isso em Agosto. -> EVENT_TIME-CALENDAR(fez,em Agosto)
//
if ( MOD(\#1,\#3) & HEAD(\#3,\#2) & NE[t-date](\#2) & EVENT[other](\#1) & ~EVENT(\#1,\#2) )
  EVENT[time-calendar=+](\#1,\#2).

// Example: O Pedro fez isso durante dois dias. -> EVENT_TIME-DURATION(fez,durante dois dias)
//
if ( MOD(\#1,\#3) & HEAD(\#3,\#2) & NE[t-duration](\#2) & EVENT[other](\#1) & ~EVENT(\#1,\#2) )
  EVENT[time-duration=+](\#1,\#2).

// Example: O Pedro fez isso todos os dias. -> EVENT_TIME-FREQUENCY(fez,todos os dias)
//
if ( MOD(\#1,\#3) & HEAD(\#3,\#2) & NE[t-frequency](\#2) & EVENT[other](\#1) & ~EVENT(\#1,\#2) )
  EVENT[time-frequency=+](\#1,\#2).

//============================
// TOP node modifiers
//============================

//NJM+JB,2013/04/01
// TOP node modifiers are associated with all the events!!!

// Example: Em agosto, o Pedro fez isso. -> EVENT_TIME-CALENDAR(fez,em Agosto)
//
if ( MOD(?[cat:0],\#3) & HEAD(\#3,\#2) & NE[t-date](\#2) & EVENT[other](\#1) & ~EVENT(\#1,\#2) )
  EVENT[time-calendar=+](\#1,\#2).

// Example: Durante dois dias, o Pedro fez isso. -> EVENT_TIME-DURATION(fez,durante dois dias)
//
if ( MOD(?[cat:0],\#3) & HEAD(\#3,\#2) & NE[t-duration](\#2) & EVENT[other](\#1) & ~EVENT(\#1,\#2) )
  EVENT[time-duration=+](\#1,\#2).

// Example: Todos os dias, o Pedro fez isso. -> EVENT_TIME-FREQUENCY(fez,todos os dias)
//
if ( MOD(?[cat:0],\#3) & HEAD(\#3,\#2) & NE[t-frequency](\#2) & EVENT[other](\#1) & ~EVENT(\#1,\#2) )
  EVENT[time-frequency=+](\#1,\#2).

//=========================================================
// Semantic Roles for MANNER
//=========================================================

// If a MODifier has as head a manner operator-noun ('modo', 'maneira', 'forma' and 'jeito' [PB]),
```

```
// then a new participant of the event is created and the feature 'manner' is added to that participant.
// In this case the participant is the adj that modifies the operator-noun.
//

// Exemplo: O Pedro cheira de um modo estranho --> EVENT_MANNER(cheira,estranho)
//
if (MOD(\#1,\#3[nop]) & MOD(\#3,\#2[adj]) & EVENT[other](\#1) & ~EVENT(\#1,\#3) & ~EVENT(\#1,\#2) )
  EVENT[mannersr=+](\#1,\#2).

// Exemplo:  Nao ha exemplo, por isso foi comentado!!!!
//
//if (MOD(\#1,\#3[nop]) & ~MOD(\#3,\#2[adj]) & EVENT[other](\#1) & ~EVENT(\#1,\#3) & ~EVENT(\#1,\#2) )
//  EVENT[mannersr=+](\#1,\#3).

// Exemplo: O Pedro cheira bem                    --> EVENT_MANNER(cheira,bem)
// Example: O Pedro fez isso inteligentemente     --> EVENT_MANNER(fez,lentamente)
//
if ( (MOD(\#1,\#2[advmanner]) || MOD(\#1,\#2[advmansubj]) ) &
     EVENT[other](\#1) & ~EVENT(\#1,\#2)
   )
  EVENT[mannersr=+](\#1,\#2).

if ( MOD(\#1,\#2[advmanner]) &
     EVENT[other](\#1) & ~EVENT(\#1,\#2)
   )
  EVENT[mannersr=+](\#1,\#2).

//*****************************************************************************
//*****************************************************************************
//***                                                                       ****
//***                            FOURTH PART                                ****
//***                           (exceptions)                                ****
//***                                                                       ****
//*****************************************************************************
//*****************************************************************************

//abarrotar de object-gen hum (o verbo permite hum e nao hum na posi\c{c}ao de N1)

//if (subj(\#1[NOhum],\#2[UMB-Animal]) & EVENT[other](\#1) & ~EVENT(\#1,\#2) )
//  EVENT[agent-generic=+](\#1,\#2).

//*****************************************************************************
//*****************************************************************************
//***                                                                       ****
//***                            FIFTH PART                                 ****
//***               (For verbs that do not belong to Viper)                 ****
//***                                                                       ****
//*****************************************************************************
//*****************************************************************************

//=====================================================
// Semantic Roles Heuristics:  AGENT-GEN
```

```
//=========================================================

// if the subject has the distributional feature of human, the semantic role
// of \texttt{AGENT} is attributed
//
// Example: O Joao leu o livro   ->   EVENT_AGENT(leu,Joao)
//
if ( SUBJ(\#1,\#2[human]) & EVENT[other](\#1) & ~EVENT(\#1,\#2) )
   EVENT[agent-generic=+](\#1,\#2).

//=========================================================
// Semantic Roles Heuristics:  INSTRUMENT
//=========================================================

// if a prepositional phrase introduced by the preposition ``com" (with) has a name
// with the concrete countable or tool semantic feature (Sem-cc, Sem-tool) as its
// head, the semantic role of \texttt{INSTRUMENT} is attributed to the complement
//
// Example: O Joao bateu no prego com o martelo   ->   EVENT_INSTRUMENT(bater,martelo)
//
//if ( PREPD(\#2[SEM-tool],\#3[lemma:com]) & EVENT[other](\#1) & MOD(\#1,\#2) & ~EVENT(\#1,\#2) )
//   EVENT[instrument=+](\#1,\#2).

//RT 2013-07-23 sem features and UMB added
if (MOD(\#1,\#2[sem-tool:+]) & PREPD(\#2,?[lemma:com]) & EVENT[other](\#1) & ~EVENT(\#1,\#2) )
  EVENT[instrument=+](\#1,\#2).

// Exemplo: O Joao falou ao telefone -> EVENT_instrument(falar,telefone)
//
if (MOD(\#1,\#2[sem-tool:+]) & PREPD(\#2,?[lemma:a]) & EVENT[other](\#1) & ~EVENT(\#1,\#2)
    & ~EVENT[instrument](\#1,?) )
  EVENT[instrument=+](\#1,\#2).

// captures ``a facada"   NaO ESTa A FUNCIONAR
if (MOD(\#1,\#2[instrumental]) & EVENT[other](\#1) & ~EVENT(\#1,\#2) & ~EVENT[instrument](\#1,?) )
  EVENT[instrument=+](\#1,\#2).

//=========================================================
// Semantic Roles Heuristics:  MANNER
//=========================================================

// if a prepositional phrase introduced by the preposition ``com" (with)
// has a name with the mass abstract or countable abstract features as its head,
// the complement is marked with the semantic role of MANNER
//
// Example: O Pedro andava com leveza   ->   EVENT_MANNER(andar,leveza)
//
if ( PREPD(\#2,\#3[lemma:com]) &
     EVENT[other](\#1) &
     (MOD(\#1,\#2[SEM-am]) | MOD(\#1,\#2[SEM-ac])) &
     ~EVENT(\#1,\#2)
   )
```

```
   EVENT[mannersr=+](\#1,\#2).

//=========================================================
// Semantic Roles Heuristics:  CO-AGENT
//=========================================================

// e preciso que esta regra tenha primeiro determinado que os argumentos simetricos
// nao se encontram coordenados ja noutra posi\c{c}ao sintactica; depende do tipo de constru\c{c}ao
// simetrica: 35S e 36S2 trata-se de coordena\c{c}ao de sujeitos e nos restantes 36S2
// de coordena\c{c}ao de objectos (CDIR)

// if a prepositional phrase introduced by the preposition ``com" (with), has a
// name with the human feature as its head; and if the verb on which the
// prepositional phrase depends is a symmetric verb (Vsim, 35S, 42S, 36S1, 36S2);
// and if no CO-AGENT has been marked yet; the semantic role of CO-AGENT
// is attributed to the complement
//
// Example: O Pedro casou com a Joana   ->   EVENT_CO-AGENT(casar,Pedro)
//
if ( PREPD(\#2,\#3[lemma:com]) &
     MOD(\#1,\#2)
     ( EVENT[other](\#1[vsim]) | EVENT[other](\#1[35S]) | EVENT[other](\#1[42S]) |
       EVENT[other](\#1[36S1]) | EVENT[other](\#1[36S2]) ) &
     ( MOD(\#1,\#2[human]) | MOD(\#1,\#2[SEM-ac]) ) &
     ~EVENT(\#1,\#2)
   )
   EVENT[co-agent=+](\#1,\#2).

//Nota: faltam os nomes e os adjetivos simetricos que tambem podem ter CO-AGENT, etc.

//=========================================================
// Semantic Roles Heuristics:  OBJECT-CL
//=========================================================

if (MOD(\#1,\#2[UMB-Anatomical]) & PREPD(\#2,\#3[lemma:em]) & EVENT[other](\#1) & ~EVENT(\#1,\#2) )
  EVENT[object-cl=+](\#1,\#2)
```

Appendix F

Semantic Role Set Validation Corpus

The next pages contain the corpus of 68 sentences that were randomly selected from the corpus and manually annotated with the semantic roles of the first full verb (in bold font), in order to assess the inter-annotator agreement when using the set of SR defined in this project. The golden standard has been determined by two linguists, working together and carefully discussing each choice. Before each sentence, in square brackets, there is the random ID number attributed to the sentence, its length (in characters) and the ViPEr verb class.

[4 128 32C] Finalmente, e provando que não brinca em serviço, o jornalista do "El País" vai **buscar** "A Capital" para dar a última estocada.
Agent-gen(buscar,jornalista)
Object-gen(buscar,"A Capital")

[8.1 79 9] No antetítulo,**refere-se** que "Tribunal de Lisboa recusa providência cautelar".
Locative-place(refere,antetítulo)

[8.2 131 13] Estes responsáveis foram **acusados** de ocultar informação ao mercado sobre as suas intenções de compra de 75% da Sevillana e Fecsa.
Patient(acusados,responsáveis)

[13 136 35LD] Igual evolução tiveram os lucros do grupo, que **baixaram** de 59,2 mil milhões de ienes para 8,5 mil milhões de ienes, uma quebra de 86%.
Object-gen(baixaram,que(os lucros))
Object-q(baixaram,59,2 mil milhões)
Object-q(baixaram,8,5 mil milhões)

[23 125 35LD] Os cidadãos norte-americanos não estão **autorizados** a entrar no Líbano, devido a uma interdição federal em vigor desde 1985.
Patient(autorizados,cidadãos)

[26 118 9] A CNMV **declara** agora que tem fortes indícios de que o BBV "violou o segredo de salvaguarda de informação reservada".
Agent-speaker(declara,CMVM)
Time-calendar(declara,agora)

[43 137 9] Ferreira do Amaral, um eurocético convicto, vai mais longe, ao **defender** que Portugal não deveria entrar já no primeiro pelotão do euro.
Agent-speaker(defender,Ferreira do Amaral)

[45 65 6] Bayer **quer** poupar cem milhões com a redução de 4 mil efectivos.
Experiencer-vol(quer,Bayer)

[54 82 9] **Refere-se** a uma hipótese veiculada pela revista Scrip com base numa fonte da CE.
Topic(refere,hipótese)

[63 109 33NM] O volume de negócios **cresceu** 10% no exercício passado, registando um valor de 6.659 mil milhões de dólares.
Object-f(cresceu,volume de negócios)
Object-q(cresceu,10%)
Time-calendar(cresceu,ano passado)

[71 137 33NM] No total, a venda dos títulos da Elf e do BNP **representa** mais de 2,5 mil milhões de francos e uma mais-valia de 550 milhões de francos.
Object-gen(representa,venda)
Object-q(representa,milhões)

[74 75 32C] A conta fica assim nas mãos da empresa que **trabalha** a publicidade do BCP.
Agent-gen(trabalha,que(empresa))
Object-f(trabalha,publicidade)

[75 139 9] A própria satisfação face ao líder palestiniano, Arafat, está em queda: 54% **dizem-se** satisfeitos da acção de Arafat, contra 68% em Abril.
Agent-speaker (dizem,54%)

[76 131 1t] "São os pontos críticos apontados no relatório EHLASS e, por isso, essas áreas **merecem** uma atenção regulamentadora", acrescentou.
Object-f(merecem,áreas)
Object-f(merecem,atenção)

[98 141 35LD] Dirac, um dos fundadores da mecânica quântica, vai **buscar** o fundamento das suas "descobertas" à harmonia estética das equações matemáticas.
Agent-gen(buscar,fundadores)

[100 76 1t] Os títulos da operadora **reagiram** em alta à notícia do Wall Street Journal.
Object-gen(reagiram,títulos)
Object-f(reagiram,notícia)

[109 120 6] O MCM África, o primeiro canal musical sobre África lançado em Maio de 1996 pelo grupo MCM, **festeja** o seu aniversário.
Agent-gen(festeja,MCM África) Occurrence(festeja,aniversário)

[110 73 9] b) **Solicitar** ao Infarmed a devolução de todos os documentos copiados; .
Addressee(solicitar,Infarmed)
Object-f(solicitar,devolução)

[114 120 32C] Os Laboratórios Euro-Labor ficam à espera que o Diário Económico **publique** esta carta ao abrigo do direito de resposta.
Agent-gen(publique,Diário Económico)
Object-f(publique,carta)

[119 106 6] "Os direitos dos cidadãos **devem** ser respeitados, tal como está estipulado na nossa Constituição", disse.
Object-f(respeitados,direitos)

[127 65 9] Em Abril, 40% **aprovava** acções violentas contra o Estado hebreu.
Experiencer-gen(aprovava,40%)
Object-f(aprovava,acções)

[128 78 33NM] Os lucros do grupo **cresceram** 26,1%, passando para 33,6 mil milhões de ienes.
Object-f(cresceram,lucros)
Object-q(creseceram,26,1%)

[134 125 32H] **Empregando** 142 mil empregados em todo o mundo, a Bayer pretende aumentar os efectivos em 2000, mas apenas fora da Alemanha.
Patient(empregando,empregados) Location-place(empregando,mundo)

[138 129 9] Na sexta-feira passada, o grupo havia já **revelado** a venda de 0,7% do capital da Elf Aquitaine por 1,226 mil milhões de francos.
Time-calendar(revelado,sexta-feira)
Agent-speaker(revelado,grupo)
Object-f(revelado,venda)

[139 95 9] Se se **confirmar** a infracção, a multa a aplicar representará o dobro, ou o triplo desse valor.
Object-f(confirmar,infração)

[142 140 9] Em contrapartida, o grupo alemão **compromete-se** a renunciar aos despedimentos e investir mais de mil milhões de marcos na Alemanha por ano.
Agent-speaker(compromete,grupo)

[145 77 36LD] Neste caso **insere-se** a valorização de 0,29% das acções da Portugal Telecom.
Object-f(insere,valorização)

[156 117 6] Mais 0,2% que os 3% do défice face ao PIB **exigidos** por Maastricht equivale a cerca de 4 biliões de liras, explicou.
Experiencer-vol(exigidos,Maastricht)

[158 66 6 O tribunal não **reconheceu** razão à Astra em nenhum desses pontos.
Experiencer-gen(reconheceu,tribunal)
Object-f(reconheceu,razão)

[161 83 32C] De acordo com o quotidiano, o negócio está **avaliado** em 50 mil milhões de dólares.
Object-gen(avaliado,negócio)
Object-q(avaliado,milhões)

[162 128 33MV] Passou-se depois a uma votação global sobre o projecto de estatutos, tendo os presentes **votado** favoravelmente por unanimidade.
Agent-gen(votado,presentes)
Manner(votado,favoravelmente)
Manner(votado,unanimidade)

[163 95 36TA] E mesmo o nosso poeta Gedeão **faz** corresponder ao binómio de Newton a beleza da Vénus de Milo.
Agent-gen(faz,poeta)

[172 102 1t] Esses dois trabalhos **destacam** ainda situações de insegurança em parques infantis e mesmo em escolas.
Object-f(destacam,situações)

[174 111 9] c) **Notificar** o Infarmed que deve desapreciar os documentos copiados para próximos processos de reavaliação; .
Addressee(notificar,Infarmed)

[176 65 32R] Mas, **traça** um quadro cinzento do futuro da economia portuguesa.
Object-gen(traça,quadro)

[179 94 33NM] Segundo o jornal “Les Echos", a mais-valia desta operação **atingiu** os 400 milhões de francos.
Object-f(atingiu,mais-valia)
Object-q(atingiu,milhões)

[187 139 32H] A AT&T **vai** fundir-se com a SBC Communications, uma operadora regional de telecomunicações norte-americana, segundo o Wall Street Journal.
Object-gen(fundir,AT&T)
Object-gen(fundir,SBC)

[191 108 9] O líder comunista **salientou**, todavia, que a discussão do Governo com os parceiros sociais vai prosseguir".
Agent-speaker(salientou,líder)

[197 146 33NM] Os valores não cotados em Bolsa não **podem** exceder 25% do valor líquido global do fundo e nenhum título pode representar mais de 10% da carteira.
Object-f(exceder,valores)

[199 124 32C Esta operação, **realizada** através de um intermediário estrangeiro, proporcionou ao grupo um lucro de 35 milhões de pesetas.
Manner(realizada,através)

[212 118 33NM No segmento não residencial o SRE **situou-se** nos 8% negativos, ainda assim abaixo do registado no trimestre anterior.
Object-gen(situou,SRE)

[216 117 6] Mas, apesar de tanto pessimismo só 24% dos palestinianos **apoiam** atentados anti-israelitas, contra 65% que se opõem.
Experiencer-gen(apoiam,palestiniano)
Object-f(apoiam,atentados)

[223 124 32A] O construtor americano General Motors vai **criar**, no próximo mês, uma joint-venture com um grupo automóvel de Xangai, SAIC.
Agent-creator(criar,construtor)
Time-calendar(criar,próximo mês)
Object-gen(criar,joint-venture) Comitative(criar,grupo)

[228 105 35LD] Quer dizer: o facto **vai** para antetítulo (menos importante do que o título); a hipótese vai para título.
Object-gen(vai,facto) Location-destination(vai,título)

[229 110 31R] A Bolsa de Derivados do Porto **fechou** com um volume de negócios de 54,3 milhões de contos, após 309 negócios.
Object-gen(fechou,Bolsa de Derivados do Porto)

[229 128 9] O presidente da Electrafina, Albert Frére **acrescentou** que a empresa investiu mil milhões de dólares (170,6 milhões de contos).
Agent-speaker(acrescentou,Albert)

[235 83 11] O banco holandês ABN Amro Bank foi **admitido** à cotação no New York Stock Exchange.
Object-gen(admitido,banco)
Object-f(admitido,cotação)

[238 125 1t] O regresso da CNN à capital libanesa vai permitir **cobrir** mais alargadamente e em profundidade o mundo árabe, disse Johnson.
Manner(cobrir,alargadamente)
Manner(cobrir,em profundidade)
Topic(cobrir,mundo árabe)
Agent-speaker(disse,Johnson)

[249 63 8] Mudanças pelas quais nos continuaremos a **bater**", acrescentou.
Object-f(bater,as quais(mudanças))

[260 130 9] O Diário Económico **diz**, citando o responsável da Astra, que o tribunal "deu como provados os principais pontos da queixa Astra".
Agent-speaker(diz,Diário Económico)

[270 143 6] Para 1997/98, a Toshiba **espera** realizar um resultado antes de impostos de 140 mil milhões de ienes, para um lucro de 75 mil milhões de ienes.
Experiencer-vol(espera,Toshiba)

[271 77 9] Comissão Nacional de Justiça e Paz **discute** nova ética nas relações sociais.
Agent-speaker(discute,Comissão Nacional de Justiça) Topic(discute,ética)

[276 115 6] O estudo **considera** ainda que a Portucel tem um dos mais baixos custos de produção da indústria da pasta europeia.
Object-gen(considera,estudo)

[279.1 82 32H] Este caso foi **julgado** no dia 21.4.1997, no 15 Juízo Cível da Comarca de Lisboa.
Object-gen(julgado,caso)
Time-calendar(julgado,dia) Location-place(julgado,Juízo Cível da Comarca de Lisboa)

[279.2 125 1t] O forte rateio que vai **rodear** a operação desmotivou os aforradores, que nas últimas sessões haviam procedido a mais-valias.
Cause(rodear,rateio)
Object-f(rodear,operação)

[284 85 9] A AT&T **recusou-se**, segundo a France Press, a comentar a existência das negociações.
Agent-speaker(recusou,AT&T)

[300 147 6] Aos espanhóis **parece-lhes** preocupante que a imprensa lusa se preocupe tanto com as relações comerciais entre os dois países da península ibérica.
Experiencer-gen(parece,espanhóis)

[302 98 9] Pelo menos no que toca aos artigos onde o seu país é, de forma directa ou indirecta, **mencionado**.
Object-f(mencionado,país)

[308 140 9] A administração do novo grupo, nascida da fusão entre as duas maiores seguradoras francesas, não **confirmou** a venda da participação no BNP.
Agent-speaker(confirmou,administração)

Object-f(confirmou,venda)

[322 125 6] A PT **liderou** mesmo a sessão, ao movimentar títulos no montante de 3,7 milhões de contos, à última cotação de 6.710 escudos.
Agent-gen(liderou.PT)
Object-f(liderou,sessão)

[331.1 119 32C] A RTP **organiza** um espectáculo para dia 3 de Junho, terça-feira, para assinalar cinco anos de emissões internacionais.
Agent-gen(organiza,RTP)
Ocurrence(organiza,espetáculo)

[331.2 137 9] Foi assim que Netanyahu, o primeiro-ministro de Israel **comentou** o encontro de três horas, em Sharm el-Sheikh, com o Presidente egípcio.
Agent-speaker(comentou,primeiro-ministro)
Topic(comentou,encontro)
Comitative(Presidente,comentou)

[335 70 9] Assunto: Bruxelas **admite** procedimento contra Portugal no caso Astra.
Agent-speaker(admite,Bruxelas)
Object-f(admite,procedimento)
Patient(admite,Portugal)

[340 142 32H] Os amotinados **mantinham** em seu poder, à hora de fecho desta edição, oito reféns, na sua maioria mulheres e crianças, familiares de reclusos.
Agent-gen(mantinham,amotinados)
Time-calendar(mantinham,hora de fecho)
Patient(mantinham,reféns)

[348 82 33NM] A Central de Cervejas **comprou** 9,53% do capital da Companhia Andaluza de Cervezas.
Agent-gen(comprou,Central de Cervejas)
Object-q(comprou,capital)

[350 94 9] Infelizmente, não há nenhum sinal de tal mudança na política americana", **disse** o Presidente.
Agent-speaker(disse,Presidente)

[357 137 32H] A Endesa foi igualmente **sancionada** com uma multa de 300 milhões de pesetas (357 mil contos) por ocultar informação sobre esta operação.
Patient(sancionada,Endesa)

[395 109 1t] Esta operação **insere-se** na estratégia definida pelo banco de aproximação aos investidores norte-americanos.
Object-f(insere,operação)
Object-f(insere,estratégia)