INTEGRATION OF NOMINAL PREDICATES INTO A PARSER: AN EXPERIMENT WITH THE CONSTRUCTIONS WITH THE SUPPORT VERB *DAR* 'GIVE' IN BRAZILIAN PORTUGUESE

Amanda RASSI*
Jorge BAPTISTA**
Oto VALE***
Nuno MAMEDE****

■ ABSTRACT: This article describes the methodology for the integration of nominal predicates, that is, support verb constructions (*SVC*), in the XIP parser, which is used by STRING, a Portuguese processing chain. More specifically, 580 *SVC* with the support verb (*Vsup*) *dar* 'give' and a predicative noun (*Npred*), whose syntactic-semantic properties have been described, formalized and then integrated into the Portuguese grammar of XIP, by means of rules, in order to extract the syntactic dependency (noted SUPPORT) between the *Npred* and the *Vsup*. The need to automatically treat *SVC* derives from the fact that they are different from full-verb constructions, have complex syntactic structures, have specific syntactic-semantic properties, and allow several systematic, albeit lexically determined, syntactic transformations. The concept of *SVC*, as well as the lexical-syntactic approach here adopted, follows the theoretical and methodological principles of the Lexicon-Grammar theory. As a result of integrating these data into the XIP parser, the system achieved 85% precision, 87% recall, 80% accuracy and 86% F-measure on an evaluation *corpus*, specifically built for this purpose.

■ KEYWORDS: Support verb. Predicative noun. Support verb construction. Causative verb-operator. XIP parser.

Introduction

Support verb constructions (*SVC*) are nominal predicates formed by a support verb (*Vsup*) and a predicative noun (*Npred*). In this sense, to identify an *SVC*, it is necessary to identify both verbs that can function as *Vsup* and the predicative nouns

---

* Universidade Federal de São Carlos (UFSCar), Centro de Educação e Ciências Humanas, São Carlos – São Paulo – Brasil. Departamento de Letras. amandarassi85@gmail.com. ORCID: 0000-0001-5314-1868

** Universidade do Algarve (UAlg), Gambelas – Faro – Portugal. jbaptis@ualg.pt. ORCID: 0000-0003-4603-4364

*** Universidade Federal de São Carlos (UFSCar), Centro de Educação e Ciências Humanas, São Carlos – São Paulo – Brasil. Departamento de Letras. otovale@ufscar.br. ORCID: 0000-0002-0091-8079

**** Instituto de Engenharia de Sistemas e Computadores (INESC), Investigação e Desenvolvimento. Lisboa – Portugal. nuno.mamede@l2f.inesc-id.pt. ORCID: 0000-0001-6033-158X

that are constructed with them. In this work, we adopt the notion of support verb from the transformational grammar of operators of Harris (1991) and from the Lexicon-Grammar approach (GROSS, 1975, 1981).

In addition to the concept of *SVC*, there are also different properties that can be used to identify these constructions (RANCHHOD, 1990; BAPTISTA, 2005). The main test, which represents a necessary and sufficient property of *SVC*, is the close relationship between *Npred* and (typically[1]) the subject of the *SVC* (*e.g. Pelé deu um chute na bola* 'Pelé kicked in the ball', interdicting the construction **Pelé deu o chute do Neymar na bola* 'Pelé gave Neymar's kick in the ball'). This relation has the same semantic nature as the relation between the verb and its subject, in a verbal predicate (*e.g. Pelé chutou a bola* 'Pelé kicked the ball').

In addition to this test, there are others that can be indicative of an *SVC*, such as (a) replacing the construction with *Vsup* with a corresponding full verb (such as *dar um abraço* 'give a hug' = *abraçar* 'to hug/embrace', or such as *dar um beijo* 'give a kiss = *beijar* 'to kiss'; (b) the restrictions on the determiners (as in *Ana deu uma passeada no parque* 'Ana took (lit. gave) a walk in the park'; interdicting the construction **Ana deu minha passeada no parque* 'Ana took (lit. gave) my walk in the park'); (c) the descent of the adverb, which allows an adverb modifying a verbal construction to "descend" to adnominal modifier position as the corresponding adjective in the equivalent nominal construction (*e.g. Rui chutou fortemente a bola* 'Rui kicked strongly the ball' = *Rui deu um chute forte na bola* 'Rui kicked the ball hard (lit. gave a strong kick in the ball)'); among other tests.

For a more general view of *SVC*, see, among others, Gross (1981, 1994, 1998), Giry-Schneider (1978, 1987), Meunier (1981), Vivès (1983), and Ranchhod (2005). The literature on automatic processing of *SVC* offers at least two distinct approaches to this phenomenon: (i) one of them considers *SVC* as a block whose constituents are relatively fixed, as a subtype of *multi-word expressions*, such as compound words and many idiomatic expressions (CALZOLARI et al., 2002; SAG et al., 2002; DIAB; HUTADA, 2009); (ii) and another perspective that considers *SVC* as a complex syntactic structure, which follows the same rules of the general grammar of the language, but has specific properties and admits systematic syntactic transformations. This work adopts this second approach, which recognizes and describes the networks of syntactic relations existing among the constituents of an *SVC*.

The *SVC*, because they are complex phenomena, present a series of challenges for their automatic processing, such as, for example, the fact that *SVC* are not always the result of nominalizations; the *Vsup* of the *SVC* is not always explicit in the sentence, as base SVC form may have undergone several types of reduction; nominal constructions do not necessarily maintain the same number of arguments, but only a subset of argument domain of their equivalent verbal constructions (while keeping the same distributional constraints); etc. As a result, syntactic parsers in general do not address this phenomenon.

---

[1] In standard SVC, this relation holds between the agentive argument in the subject slot and the *Npred*, as in *A Ana deu um beijo no Rui* 'Ana gave a kiss to Rui', while in converse SVC, like *O Rui recebeu um beijo da Ana* 'Rui received a kiss from Ana', the agentive argument is placed in a prepositional complement slot.

The parsers (automatic syntactic analyzers) available in Portuguese, such as PALAVRAS (BICK, 2000) and the LX-*parser* (SILVA et al., 2010), apparently do not yet have information on nominal predicates formed by *Vsup* and *Npred*.

Though there are different types of nominal predicates, in this work we will deal specifically with nominal constructions whose predicative nucleus is a noun (called *predicative noun*, *Npred*) and this noun is auxiliated by a verb (called a *support verb*, *Vsup*). In this sense, we developed a systematic linguistic analysis of *SVC*, we adopted a formalization of the data based on the proposal of the Lexicon-Grammar (GROSS, 1975, 1981), we integrated the data in an automatic processing chain of Portuguese, STRING (MAMEDE et al., 2012), and we evaluated the result of the system based on the manual annotation of a *corpus*.

The analysis, description and classification of the data were done in three different works: 1,815 nominal predicates formed by the support verb *fazer* 'do/make' (BARROS, 2014); 2,273 nominal predicates with the support verb *ter* 'have' (SANTOS, 2015); and 1,489 nominal predicates with the support verb *dar* 'give' (RASSI, 2015). All these data have been systematically analyzed, described and formalized in the Lexicon-Grammar (LG) matrices of the Portuguese nominal constructions.

In the LG methodology, the description of the linguistic phenomena is often presented in the form of binary matrices: the lines contain the lexical entries (in this case, the *Npred*) and the columns represent the syntactic-semantic properties of each entry. For example, each predicative noun imposes distributional constraints on the type of arguments it selects, the preposition that introduces the essential complement(s) and the determiner of the predicative noun. The matrix also encodes the standard and the converse support verbs (see below)*,* as well as their aspectual and/or stylistic variants; it codifies the thematic or semantic roles of the arguments; the possibility of accepting or not the *Conversion*, the *Passive*, and the *Symmetry* transformations, among other properties.

Although descriptions of the nominal predicates with the support verbs *fazer* 'do/ make', *ter* 'have' and *dar* 'give', are already available in tabular format, this work presents only the results of the integration of the nominal constructs with the support verb *dar* 'give' in STRING.

STRING is a Portuguese processing chain with a modular structure that performs the main basic tasks of Natural Language Processing (NLP), such as tokenization, textual segmentation, labeling part-of-speech tags (*POS-tagging*), morphosyntactic disambiguation, *chunking*, deep syntactic analysis, such as the extraction of dependencies (subject, complement, *etc.*), among other tasks. For several of these tasks, but mainly for parsing, STRING uses the *Xerox Incremental Parser* (XIP), which is a statistical and rule-based parser (MOKHTAR *et al*.,2002).

The data of the constructions with the support verb *dar* 'give' were integrated into the processing chain as one of the components of the Portuguese grammar, implemented in XIP. This was done in the form of lexical-syntactic dependency extraction rules in order to automatically extract the dependency we call SUPPORT between *Vsup* and *Npred* and between the *Npred* and its arguments.

In a previous work (RASSI et al., 2015), we described a general proposal for extracting events and dependencies associated to constructions with *Vsup* in STRING. In that work, we indicated the strategy adopted for the implementation of support verb constructions in that system. Remember that *SVC* can form standard constructions (*Ana deu um beijo no Rui* 'Ana gave a kiss on Rui' - SUPPORT[vsup-standard]), with an active semantic orientation, or converse constructions (*O Rui recebeu um beijo de Ana* 'Rui received a kiss from Ana' - SUPPORT[vsup-converse]), with a passive semantic orientation.

In this paper, we will especially describe the results of the automatic processing of *SVC* with *Vsup dar* 'give' in STRING, and we compare the system's output with the manual annotation of a sample of the *corpus* PLN.Br Full (BRUCKSCHEN et al., 2008). The total sample has 2,646 sentences randomly extracted from PLN.Br Full, with verb-noun pairs candidates for *Vsup* and *Npred* status. In this work, however, we refer only to 580 phrases of this total sample, which correspond to the sentences involving the verb *dar* and its variants.

State of the art

Much of the work that describes automatic tasks related to *SVC* deals with the identification or the extraction of these constructions from *corpora*, whether based on lexical patterns (through regular expressions) or based on manual annotated *corpus* and machine learning techniques.

Grefenstette and Teufel (1995) present a method of identifying support verbs from an unlabeled *corpus*, by comparing the arguments related to verbal forms and the nominalized potential forms, that is, the argumental network is transferred from the verbal constructions to the nominal constructions potential candidates. The authors seek to find the most likely support verbs for each predicative noun but considering only the *Npred* that are nominalizations of verb forms. It is known that many *Npred* are nominalizations of verbs, such as in the pairs {*abraço*, *abraçar*} 'a hug, to hug', {apresentação, apresentar} 'presentation, to present', {*chute*, *chutar*} 'a kick, to kick', *etc*., but there are also *Npred*, which are called *autonomous predicative nouns*, that are not derived from verbs such as *greve* 'strike', *sermão* 'sermon', *cólica* 'colic', *etc.* Thus, the method presented by the authors does not capture these autonomous *Npred*. In that work, Grefenstette and Teufel (1995) extracted from an English *corpus* 6,704 sentences with candidates for support verbs and candidates for nominalizations, producing a list of potential support verb constructions that occur with the nominalized forms. In addition to disregarding the *Npred* autonomous, another problem of this approach consisted in considering that the nominal construction maintains the same number of arguments in their argument domain as the equivalent verbal construction, which is not always the case.

For Spanish, Páez (2014) extracted from a *corpus* 81,274 phrases with candidates of *SVC* from which the most representative are the support verbs *tener* 'have', *hacer* 'do' and *dar* 'give'. The author also automatically extracted combinations of any noun and a set of 12 verbs, all common variants of *Vsup tener*, *hacer* and *dar*. She then ordered the main combinations of verb and noun by frequency and calculated the likelihood of co-occurrence of such a verb with that noun, using 3 association measures (log-likelihood, Student's T-score and Maximum Likelihood Estimator). At the end of the task, the author listed the 15 *SVC* most recurrent in Spanish according to the association measures used and concluded that approximately 69% of the *SVC* in this list were correctly identified.

In the literature, we find other works, similar to that of Paez (2014), which start from a previous list of verbs that can work as *Vsup* or a list of nouns that can function as *Npred*. The proposal of Duran et al. (2011) differs from these approaches by starting from the syntactic patterns of combinations of grammatical categories to find the *SVC*. They used patterns such as V N PRP (*abrir mão de* 'give up smthg.'), V PRP N (*pôr de lado*, 'leave aside'), V DET N PRP (*virar as costas para* 'turn one's back to'), V DET ADV (*dar o fora* 'run away'), V ADV (*ir atrás de* 'go after smthg.'), V PRP ADV (*dar para trás* 'reject/waver'), V ADJ (*dar duro* 'work hard')[2].

Using this method, Duran et al. (2011) were able to identify 773 *complex predicates*, which were then annotated manually. According to the authors, these complex predicates include (but are not limited to) *SVC*, which the authors call *light verb* constructions[3]. We consider, however, that the use of regular expressions with combinations of grammatical categories will not be the most appropriate approach for the unique identification of *SVC*, since *SVC*, as a rule, are formed by V (DET) N, a pattern that is syntactically identical to the structures of ordinary verbal predicates, composed of a full verb (V), followed by a direct object (N), eventually with a determiner (DET).

On the other hand, there are also works that aim to process (not only identify) these constructions, for example, Barreiro et al. (2014), which evaluated two automatic translation systems, OpenLogos (based on rules) and Google Translate (based on neural networks), in the task of translating constructions with verb support in five languages: French, German, Italian, Portuguese and Spanish. To perform the experiments and the evaluation, the authors produced a set of 100 phrases that they analyzed as *SVC*[4] and annotated them manually. As a result of the evaluation of the two systems, the authors

---

[2]   Notation: ADJ = adjective, ADV = adverb, DET = determiner, N = noun, PRP = preposition e V = verb.

[3]   SVC are often referred to in the literature as light verb constructions (SCHER, 2004; DURAN et al., 2011; TU; ROTH, 2011; BUTT; GEUDER, 2001; ISTVÁN; VINCZE; FARKAS, 2013). The two terms, *support verb* and *light verb*, are commonly interpreted as synonyms, though there are conceptual differences between this terminology. In this work, we adopt the term *support verb* (Portuguese: *verbo-suporte*) since we consider that the main function of the Vsup is to "support" (or carry) the inflectional features of person-number and tense (temporal features but also including modality and aspect).

[4]   In fact, and to be precise, not all the sentences selected by the authors correspond to nominal constructions with support verb, and they also include adjectival and prepositional constructions, and even sentences with operator-verbs (GROSS, 1981).

concluded that Google Translate translates *SVC* better than OpenLogos, attributing this result to its rich lexical knowledge.

In the present work, with the intention of contributing to the tasks of processing the *SVC* and aiming to fill the gap for its automatic identification, we present the methodology and the results of the integration of *SVC* with the *Vsup* dar 'give' in the STRING system, using the *parser* XIP. The results of the performance of this system were evaluated against a reference *corpus*, manually and independently annotated, which will be presented in the next section.

Construction of the reference *corpus* for SVC

In this section, we will briefly explain the procedures adopted for the constitution of the reference *corpus*, its annotation and the selection of a sub-sample to be processed in STRING. The entire process of construction and annotation of this *corpus* has already been dealt with in detail in previous work (RASSI et al., 2015).

The matrices of the Lexicon-Grammar (predicative nouns and the verbs *fazer* 'do/make', *ter* 'have' and *dar* 'give') were intersected with Unitex[5] reference graphs in order to systematically search in the *corpus* PLN.Br for all possible combinations of each of these support verbs with each predicative noun, considering only the combinations predicted in the matrices. Through this methodology, 121,198 sentences were identified in the *corpus* presenting the candidate pairs {*Vsup, Npred*}{*Vsup, Npred*}, that is, sentences in which simultaneously occur a potential *Vsup* and a predicative noun.

We selected a sample of these 121,198 sentences, keeping it proportional to the number of occurrences of each pair {*Vsup, Npred*}. The sample consists of 2,646 sentences and corresponds to 2.18% of the total sentences. This selection retrieved at least one instance of all pairs {*Vsup, Npred*} that have at least 21 occurrences. Table 1 summarizes the main information about the *corpus* and the sample selected.

Table 1 – Sample data compared to the data *corpus*.

	Corpus	**Sample**	**%**
{*Vsup, Npred*} pairs (# sentences)	121,198	2,646	2.18%
# different {*Vsup, Npred*} pairs	4,668	1,130	24.2%

Source: Author's elaboration.

The sample has 1,130 different pairs of {*Vsup, Npred*}, which corresponds to 24.2% of the *corpus*, which is composed of 4,668 different pairs.

The annotation of the 2,646 sentences with candidate *SVC* was done manually by 5 Portuguese native speaker annotators who were also specialists in *SVC*. For this

---

[5] Unitex is a software that allows for the processing of large textual *corpora*, and it is available at: http://www-igm. univ-mlv.fr/~unitex/

task, an already existing *corpora* annotation tool (SUÍSSAS, 2014) was adapted. The annotation consisted in labelling, for each sentence, a (conventional) code that corresponds to the type of syntactic construction indicated by the pair {*Vsup, Npred*} that appears in parentheses at the beginning of each phrase. The possible labels were:

SVC-STANDARD - for standard support verb constructions
> Ex.: (***dar, tapa***) *Ana **deu** um **tapa** em Rui.*
> 'Ana **gave** a **slap** on Rui'

SVC-CONVERSE - for constructs with converse support verb
> Ex.: (***levar, tapa***) *Rui **levou** um **tapa** da Ana.*
> 'Rui **got** a **slap** from Ana'

VOPC - for constructions with causative operator verb
> Ex.: (***dar, medo***) *O vento sombrio **deu medo** na Ana.*
> 'The dark wind **gave fear** in Ana'
> Ex.: (***fazer, medo***) *O vento sombrio **fez** com que Ana tivesse **medo***
> 'The dark wind **made** Ana to have **fear**'

OTHER - for any other type of construction
> Ex: (***fazer, academia***) *O Rui **fez** (=construiu) uma **academia**.*
> 'Rui **did** (= built) a **gym**' [full verb]
> Ex.: (***dar, tiro***) *O governo **deu** um **tiro** no próprio pé.*
> 'The government has **given** a **shot** to itself in the foot' [fixed expression]
> Ex: (**ter, controle**) *Rui **tem** Max sob seu **controle**.*
> 'Rui **has** Max under his **control**' [linking operator-verb]

At the end of the process, the annotations were tabulated into 5 columns and the *ReCal 0.1 Alpha for 3+ Coders* tool[6] was used to calculate the agreement among the annotators. The mean agreement among the 5 scorers was 80.8%. The tool also calculates the Kappa coefficient (COHEN, 1960), which is a statistical measure that evaluates the agreement between pairs of evaluators, also called an *inter-annotator agreement* or *inter-rater agreement*. The Cohen's Kappa average attributed to the annotation was 0.604.

From the manual annotation of the 2,646 phrases, those that had agreement equal or superior to 60%, that is, those in which 3 or more annotators assigned the same label were selected. Table 2 shows the number of sentences per degree of agreement between the annotators in the general sample.

---

6   Disponível em: http://dfreelon.org/recal/recal3.php#result1

Table 2 – Distribution of sentences per degree of agreement
between annotators in general sample.

Concordance	Number of phrases	% of phrases
2 annotators	44	1.7%
3 annotators	326	12.3%
4 annotators	627	23.7%
5 annotators	1,574	59.5%
* sentences with error	75	2.8%
TOTAL	**2,646**	**100%**

Source: Author's elaboration.

As it turns out, the sentences labelled in the same way by all the annotators correspond to almost 60%. Adding the sentences annotated with the same label by the majority of the annotators (3 or more), the agreement corresponds to 95.5% of the total, which represents a good result.

We randomly selected, from the total sample set, all the sentences that had a co-occurrence, on one hand, of a predicative noun, and, on the other hand, of the verb *dar* 'give' or one of its standard variants (*aplicar* 'apply', *conceder* 'grant', *fazer* 'do/ make') or its converse support verbs (*ter* 'have', *receber* 'receive', *levar* 'take' and *tomar* 'get/take'); simultaneously, those sentences had been annotated with the same label by at least 3 annotators. In this sense, 580 sentences (22% of the general sample) were selected to compose a sub-sample and then analyzed in STRING.

Table 3 presents the distribution of the sentences in this sub-sample, by level of agreement between the annotators.

Table 3 – Distribution of sentences' level of agreement
between annotators in the sub-sample.

Concordance	Number of phrases	% of phrases
3 annotators	95	16.4%
4 annotators	137	23.6%
5 annotators	337	58.1%
*error	11	1, 9%
TOTAL	**580**	**100%**

Source: Author's elaboration.

In Table 3, we did not indicate sentences with concordance between 2 annotators because these sentences were not selected for the sub-sample. As it can be seen, the distribution of sentences by degree of agreement in the sub-sample (Table 3) is practically proportional to their distribution in the global sample (Table 2).

One can also analyze the distribution of sentences by category (or label) assigned by most annotators. Table 4 shows the distribution of sentences from the sub-sample, by category and level of agreement between the annotators.

Table 4 – Number of sentences per degree of agreement and by category.

	3 scorers	4 scorers	5 scorers	TOTAL	%
STANDARD	42	78	227	347	59.82%
CONVERSE	35	38	77	150	25.86%
OTHER	15	21	33	69	11.89%
VOPC	3	0	0	3	0.51%
* sentences with error	---	---	---	11	1.89%
TOTAL	95	137	337	580	100%

Source: Author's elaboration.

As can be seen, the category of SUPPORT[vsup-standard] is the most consensual among annotators, corresponding to almost 60%, followed by the category SUPPORT[vsup-converse], with about 25%. The remaining cases add up to 14.29%.

Data integration in STRING

Following the strategy outlined in previous work (RASSI et al., 2015), a set of programs was built that automatically converts the information contained in the Lexicon-Grammar matrices (from constructions with the support verb *dar* 'give') into the dependency extraction rules that XIP parser uses to determine the syntactic relations between *Vsup* and *Npred*. Two features distinguish the dependency relation regarding Conversion: SUPPORT[vsup-standard] and SUPPORT[vsup-converse]. These rules also cover the cases with the causative verb-operator (dependency VOP-CAUSE), however, in constructions with *give*, this category practically does not occur (only 3 cases), so we will not mention it here any further. Thus, for example, based on the information of the entry *kiss*, the system generates the following rule (Fig. 1):

Figure 1 – First dependency extraction rule producing the SUPPORT[vsup-standard] dependency for *Npred beijo* 'kiss'.

```
if (VDOMAIN(#1,#2[lemma:"dar"])&
   (MOD[post,relat](#3[lemma:"beijo"],#2) ||
    CDIR(#2[transf-passiva:~],#3[lemma:"beijo"]) ||
    SUBJ(#2[transf-passiva],#3[lemma:"beijo"]) ||
    (ANTECEDENT[relat](#3[lemma:"beijo"],#4[pronrel])&SUBJ(#2[transf-passiva],#4)))&
    ~SUPPORT[vsup-standard](#3,#2) )
              SUPPORT[vsup-standard=+](#3,#2)
```

Source: Author's elaboration.

Dependency rules consist essentially of two parts: first, the conditions (`if`) that must be verified for a dependency to be extracted are drawn; then the dependency to be extracted is stated. In this case (Fig. 1), the rule first instantiates a variable (# 2) whose lemma is that of the support verb *dar* 'give' ([lemma: "dar"]); then, a set of conditions in alternative (‖) are stated:

- the first condition (line 2 of Fig. 1) corresponds to the situation in which the *SVC* undergoes a relativization, which is captured by the dependency `MOD` between the support verb (in this case, its past participle) and the predicative noun *beijo* 'kiss' that is the antecedent of the relative pronoun (e.g. *o beijo que Rui deu na Ana* 'the kiss that Rui gave to Ana');
- the second condition (line 3) is triggered if the name *beijo* 'kiss' is in a direct complement relation (dependency `CDIR`) with the verb *dar* 'give', and this verb has not been marked with the feature [`trans-passive`], which could have been attributed previously if a passive structure had been identified (e.g. *Rui deu um beijo na Ana* 'Rui gave a kiss to Ana');
- the third condition (line 4) is the opposite to the previous one: it identifies a subject relation (`SUBJ`) between the predicative noun and the support verb when this one is in a passive construction (e.g. *Um beijo foi dado por Rui na Ana* 'A kiss was given by Rui to Ana');
- the following condition (line 5) applies when the subject of passive construction has as its antecedent a relative pronoun (#4[`pronrel`]), which is then subject of the support verb employed in passive (e.g. *O beijo que foi dado por Rui na Ana...* 'A kiss was given by Rui to Ana');
- Finally (line 6), the rule checks if the dependency `SUPPORT[vsup-standard]` has not yet been extracted between *beijo* 'kiss' and *dar* 'give', in order to avoid repetition of the procedure.

If any of these conditions are verified, the rule is triggered and the dependency `SUPPORT[vsup-standard]` is extracted.

In addition to this rule, STRING also generates rules for the case of reduced passive constructions (as in *O beijo dado por Rui na Ana* 'the kiss given by Rui to Ana'), in which *Vsup* has been deleted (Fig.2):

Figure 2 – Second dependency extraction rule producing the `SUPPORT[vsup-standard]` dependency for the *Npred beijo* 'kiss'.

```
|pastpart#2[lemma:"dar",pass-ser=+]|
if (MOD(#3[lemma:"beijo"],#2) & ~SUPPORT[vsup-standard](#3,#2) )
SUPPORT[vsup-standard=+](#3,#2)
```

Source: Author's elaboration.

This rule differs from the previous one by the context statement, marked between vertical bars (|...|), in the first line. It adds to the past participle of *dar* 'give' the feature of the passive construction with auxiliary verb *ser* 'be' [`pass-ser=+`]. This participle must then be modifying the predicative noun. Rule order is relevant: this rule is triggered only if the previous rule has not fired yet.

The two previous rules serve to extract the dependency `SUPPORT[vsup-standard]` from *SVC*. In cases where *SVC* admits Conversion, a piece of information that is encoded in the Lexicon-Grammar matrices, STRING also generates the corresponding dependency extraction rules `SUPPORT[vsup-converse]` (Fig. 3) to enable the system to capture sentences such as *Ana recebeu um beijo de Rui* 'Ana received a kiss from Rui'.

Figure 3 – First and second extraction dependency rules, producing the `SUPPORT[vsup-converse]` dependency to *Npred beijo* 'kiss'.

```
if (VDOMAIN(#1,#2[lemma:"receber"])&(MOD[post,relat](#3[lemma:"beijo"],#2)||
 CDIR(#2[transf-passiva:~],#3[lemma:"beijo"]) ||
 SUBJ(#2[transf-passiva],#3[lemma:"beijo"]) ||
 (ANTECEDENT[relat](#3[lemma:"beijo"],#4[pronrel])&SUBJ(#2[transf-passiva],#4))) &
 ~SUPPORT[vsup-converse](#3,#2) )
SUPPORT[vsup-converse=+](#3,#2)


|pastpart#2[lemma:"receber",pass-ser=+]|
if (MOD(#3[lemma:"beijo"],#2) & ~SUPPORT[vsup-converse](#3,#2) )
SUPPORT[vsup-converse=+](#3,#2)
```

Source: Author's elaboration.

The difference between these two rules (Fig. 3) and those above (Fig. 1 and 2) is basically the lemma of the verb, which is *receber* 'receive', and consequently the dependency becomes `SUPPORT[vsup-converse]` instead of `SUPPORT[vsup-standard]`. For more information on the operation of the XIP parser dependency rules, see Mamede et al. (2012).

Evaluation

The sub-sample of the phrases (with *dar* 'give' and its variants) randomly selected from the set of manually annotated *SVC* was processed by the STRING system and its output was analyzed in comparison with the reference *corpus*. The results are presented in Table 5:

Table 5 – First evaluation of the system performance.[7]

TP	FP	FN	TN	Precision	Recall	Accuracy	F-Measure
350	91	114	25	79%	75%	65%	77%

Source: Author's elaboration.

The usual metrics were used to evaluate the output of the system, namely *Precision*, which measures the fraction of correctly found instances (true-positives) over the total of instances found (true-positives plus false-positives): (TP/(TP+FP)); *Recall*, measuring the fraction of relevant instances that were found: (TP/(TP+FN)); *Accuracy,* which computes both the correct found instances and the correctly missed cases: ((TP+TN)/(TP+TN+FP+FN)); and *F-measure*, which is the harmonic mean between precision and recall: (2PR/(P+R)).

Of the 580 sentences analyzed, STRING correctly extracted the dependence of 350 sentences (TP), incorrectly extracted the dependence of 91 sentences (FP) and did not extract any dependence of another 139 sentences, of which 114 should have been extracted (FN) and 25 sentences that should not (TN).

In addition to these results, the system captured 47 other dependencies that had not been noted in the reference, since they involve pairs of words that were not the target of the sentence extracted from the *corpus*. For example, in the sentence: *O varejo, em contrapartida, pode **dar** descontos no **valor** cobrado à indústria por determinado espaço na loja.* 'Retail, by contrast, can **give** discounts in **value** charged to the industry for certain space in the store', the target pair to be labelled was {*dar*, *valor*} 'give, value', but in this case there is no relation between the verb and the noun, so the annotators did not label it. On the other hand, in this same example, the verb *dar* 'give' is the support verb of the predicative noun *descontos* 'discounts', a dependency that the STRING system captured well. Since the SUPPORT[vsup-standard] dependency between *dar* and *desconto* was not in the reference (this was not the targeted pair), in a second moment of the evaluation, the reference *corpus* was completed, adding the missing dependencies that should have been taken into account.

The following section is an analysis of the main problems identified in the output of the system. After this analysis, we corrected some data in the Lexicon-Grammar and processed the *corpus* again. The results of this second evaluation will be presented later.

---

[7]   TP = true-positives: instances correctly found/labelled by the system; FP = false-positives, incorrectly found/labelled instances; TN = true-negatives: instances correctly missed/unlabeled; FN = false-negatives: instances incorrectly missed/unlabeled.

Error Analysis

a) false-positive

False positives (FP) correspond to the cases where (i) the system extracted the wrong dependency, as in the cases of ambiguity between `SUPPORT[vsup-standard]` and `support[vsup-converse];` or (ii) the system extracted the `SUPPORT` dependency between a pair of words that do not hold such relation, mainly due to problems on previous syntactic processing or an incorrect morpho-syntactic disambiguation. The two cases will be analyzed in detail:

(i) ambiguity between *SVC*-standard and *SVC*-converse

Typically, the support verb *dar* 'give' is selected to form standard constructions (active orientation). On the other hand, the support verb *receber* 'receive' is often selected by the same predicative nouns to form typical converse constructions (passive orientation). There are, however, other verbs - such as *ter* 'have', for example, that can enter both the standard and the converse constructions.

(1) **[built example]**[8]: *O Jô Soares (**deu** + **teve**) sua **participação** no Programa da Hebe.*
'Jô Soares (**gave** + **had**) his **participation** in Hebe Program'

(2) **[built example]**: *O Programa da Hebe (**recebeu** + **teve**) a **participação** do Jô Soares.*
'The Hebe Show (**received** + **had**) the **partaking** of Jô Soares'

The first sentence is typically a standard construction, while the second phrase is typically a converse construction. Both can be formed by the verb *ter* 'have' and the same predicative noun *participação* 'participation'. Therefore, the verb *ter* 'have' was encoded in the Lexicon-Grammar matrices both as a variant of the standard *Vsup dar* 'give' and also as a variant of converse *Vsup receber* 'receive'. It was decided to systematically adopt the dependency `SUPPORT[vsup-standard]`, instead of `SUPPORT[vsup-converse]`, in cases of ambiguity of classification, in which the verb can function both as a standard and converse *Vsup*, when supporting the same *Npred*. This is done in XIP by a "cleanup" rule, which removes duplicate dependencies at the end of the processing.

---

[8] Most examples presented in this paper have been retrieved from the sub-sample of the *corpus* used for the evaluation, in other words, they are real, naturally occurring texts. In some specific situations, some examples were devised by the authors to demonstrate a precise point in the argument, or to highlight certain phenomena, following the Lexicon-Grammar guidelines for example building (GROSS, 1981). All built examples are indicated as such. Some examples result from the simplification of real instances taken from the *corpus* and are also marked as such. Real examples are preceded by the target pair (*Vsup*, *Npred*).

This decision led to some misclassification, such as in the following cases, which were labelled by STRING as `SUPPORT[vsup-standard]`, and were (correctly) marked by most or all of the annotators as `SUPPORT[vsup-converse]`:

(3) *(ter, participação) A mesa-redonda, com início às 14h, **terá** a **participação** do historiador José Murilo de Carvalho, da UFRJ (Universidade Federal do Rio de Janeiro), e dos cientistas políticos Renato Lessa e César Guimarães, ambos do IUPERJ.*

'(**have**, **participation**) The roundtable, starting at 14h, will **have** the **participation** of historian José Murilo de Carvalho, from UFRJ (Federal University of Rio de Janeiro), and political scientists Renato Lessa and César Guimarães, both from IUPERJ.'
`SUPPORT[vsup-standard](participação,terá)`

(4) *(ter, prazo) Martins disse ter decidido indiciar Teixeira indiferentemente do resultado da perícia técnica no caminhão, que **tem prazo** de 30 dias a partir do acidente para ser concluída. ''Não dá para acreditar que alguém possa dirigir um caminhão desse tipo e não perceber que a caçamba está levantada'', disse.*

'(**have**, **deadline**) Martins said he decided to indict Teixeira regardless of the result of technical expertise to the truck, which **has** a **deadline** of 30 days [counting] from the accident to be complete. "One can not believe that someone may drive a dump truck like that and not realize that the bucket is up," he said.'
`SUPPORT[vsup-standard](prazo,tem)`

(5) *(ter, voto) O PMDB conta com 5 integrantes, mas **terá** um **voto** a menos se Juvêncio estiver na presidência.*

'(**have**, **vote**) The PMDB has 5 members, but will **have** less one **vote** unless Juvêncio is in the presidency.'
`SUPPORT[vsup-standard](voto,terá)`

(6) *(ter, prejuízo) Pará deve **ter prejuízo** com jogo do Brasil.*

'(**have**, **injury**) Pará might **have injury** with Brazilian game.'
`SUPPORT[vsup-standard](prejuízo,ter)`

On the other hand, there are also cases that STRING has labelled as `SUPPORT[vsup-converse]`, whereas the majority of the human annotators considered them as instances of `SUPPORT[vsup-standard]`. This is the case of constructions with *Vsup ter* 'have' and the following predicative nouns: *acordo* 'agreement', *alvará* 'commercial license', *apelido* 'surname', *apresentação* 'presentation', *cargo* 'position/job', *conhecimento* 'knowledge', *destino* 'destination/destiny', *dica* 'hint', *explicação* 'explanation', *financiamento* 'financing', *importância* 'importance', *informação* 'information', *início* 'beginning', *liberdade* 'freedom', *limitação* 'limitation', *motivo* 'reason, *nome* 'name', *nota* 'grade', *orientação* 'orientation', *ponto* 'point', *prioridade* 'priority', *privilégio* 'privilege', *redução* 'reduction', *renda* 'income', *sinal* 'sign', *título* 'title' and *treinamento* 'training'.

These predicative nouns, associated with the *Vsup have* 'ter', can not only form standard constructions like (7), but they can also be the result of a conversion from another distinct construction with the standard support verb *dar* 'give' (8):

(7) **[built example]**: *Ana **tem** um vasto **conhecimento** sobre geografia*.
'Ana **has** a vast **knowledge** of geography'

(8) **[built example]**: *A Ana **deu conhecimento** neste documento*
'Ana checked (lit. **gave knowledge**) this document'
[Conversion] = *Este documento **teve** o **conhecimento** da Ana*.
'This document was checked by (lit. **had knowledge of**) Ana'

As it can be seen, (7) and (8) are different standard constructions of two distinct predicative nouns: the first one refers to a human quality, someone's intellectual ability; and the other construction has a technical sense, and it refers to the act of signing or checking a document. Because they are different constructions, the only one that is listed in the Lexicon-Grammar matrix used in this work is the construction (8), with standard *Vsup dar* 'give', whose conversion is done with *ter* 'ter'. The construction illustrated in (7) is also a base sentence, but it does not select the verb *dar* 'give', so it should be described in another matrix, which takes into account the base nominal constructions with *Vsup ter* 'ter'.

When both constructs are available in the matrix, we have cases of ambiguity, which causes two rules to be triggered. In such cases, the "cleaning" rule referred to above is applied.

There are also other sentences that were also marked by STRING as SUPPORT (vsup-standard), and that were tagged by most or all of the annotators as constructions with a causative operator-verb (VOP-CAUSE):

(10) *(dar, sorte) Colocar roupa branca e pular sete ondas **dão sorte** porque são rituais para atrair coisas boas e, se você acredita, funcionam.*
'(**give**, **luck**) Putting on white clothes and jumping seven waves **give luck** (=attract good luck) because they are rituals to attract good things and, if you believe, they work.'

(11) *(dar, prejuízo)* *Fraude on-line **dá prejuízo** de R$ 100 mi.*
'(**give**, **loss**) Fraud on-line **gives** a **loss** of R$ 100 mi.'

(12) *(dar, voto) Motivo: administra o orçamento para construção de casas populares, que é polpudo e **dá votos**.*
'(**give**, **vote**) Reason: [he] manages the budget for the construction of commoners' houses, which is substantial and **gives votes** (=brings in votes).'

Constructions like *dar sorte* 'lit. give luck', *dar prejuízo* 'lit. give loss' and *dar voto* 'lit. give vote' are, of course, acceptable as standard constructions in other situations, hence they have been classified as such in the Lexicon-Grammar matrix:

(13) **[built example]**:   *Ana **deu sorte** na loteria.*
                           'Ana **got lucky** in the lottery'

(14) **[built example]**:   *A empresa da Ana **deu prejuízo** durante todo o ano.*
                           'Ana's company **gave loss** all year around'

(15) **[built example]**:   *Ana **deu** seu **voto** para o candidato da oposição.*
                           'Ana **gave** her **vote** to the opposition's candidate'

In these cases, the identification of the semantic roles of the arguments could help in the disambiguation of the dependency rules that should apply to extract the SUPPORT dependency (i.e., standard or converse *Vsup*). In (10), if the subject sub-clause (e.g. *Colocar roupa branca e pular sete ondas* 'putting on white clothes and skipping seven waves') was correctly labeled with the thematic role of **cause** for *dar sorte* 'lit. give luck', this would be an indication that the construction is a causative, not an *SVC*. The example (11) is ambiguous, although it has been noted by most human annotators as a causative construction. It is ambiguous because it allows for two different interpretations: (i) online fraud has suffered some loss; or (ii) online fraud has caused some loss to someone. Without further (contextual/situational) information, this ambiguity cannot be solved. In (12), the subject of *dar votos* 'lit. give votes' is *orçamento* 'budget', which means that this noun is the **cause** behind someone having votes.

Although the semantic roles had been encoded in the Lexicon-Grammar matrix, this information was not used in the SUPPORT dependency extraction process because STRING's automatic semantic role labelling module does not yet yield sufficiently good results.

(ii) Problems of syntactic processing or morphosyntactic disambiguation

Some sentences were incorrectly tagged by STRING as SUPPORT[vsup-standard], due to syntax processing issues. The system was expected to recognize a dependency relation between a verb and a noun, and it recognized instead a relation between another verb and/or another noun. For example:

(16) *(dar, aula) Ela dirige atualmente a Companhia Os Bobos da Corte, criada há dois anos, e **dá aulas** de voz e **interpretação** na Escola de Teatro da Universidade Federal da Bahia.*

'(**give**, **class**) She currently directs the *Company of the Fools of the Court*, created two years ago, and [she] **gives classes** (=teaches) voice and **interpretation** at the Theater School of the Federal University of Bahia.'

SUPPORT[vsup-standard](aulas,dá)
SUPPORT[vsup-standard](interpretação,dá)

(17) *(ter, vantagem) Essa solução **teria** a **vantagem** de rapidez e **rentabilidade**, trazendo ao Tesouro receita maior e evitando disputas jurídicas inerentes ao processo de cisão de ativos.*

'(**have**, **advantage**) This solution would **have** the **advantage** of speed and **profitability**, bringing higher revenue to the Treasury and avoiding legal disputes inherent in the asset scission process.'

SUPPORT[vsup-standard](vantagem,teria)
SUPPORT[vsup-standard](rentabilidade,teria)

The target pairs, whose dependencies should have been extracted are {*dar*, *aulas*} 'give, lesson' and {*dar*, *vantagem*} 'give, advantage', respectively in examples (16) and (17). In addition to correctly extracting these two dependencies, the system also recognized the dependency between *dar* 'give' and *interpretação* 'interpretation' in (16) and between *dar* 'give' and *rentabilidade* 'profitability' in (17). In these cases, there is an issue with the coordination of noun phrases and the proper extraction of direct complement (CDIR) dependence. The system analyzed (16) as a coordination between *aulas* 'classes' and *interpretação* 'interpretation', and not between *voz* 'voice' and *interpretação* 'interpretation', considering incorrectly that there was a coordination between two direct complements of the verb *dar* 'give': 'she gives voice lessons' and 'she gives interpretation'. In the same way, the processing chain analyzed (17) as a coordination between *vantagem* 'advantage' and *rentabilidade* 'profitability', and not between *rapidez* 'speed' and *rentabilidade* 'profitability', considering the coordination between two direct complements of the verb *ter* 'have': 'this solution would have the advantage of speed' and 'this solution would have profitability' .

In other cases, the problem results from an incorrect assignment of the morphosyntactic labels of grammatical categories (part-of-speech tags, PoS) or their inadequate disambiguation. In such cases, STRING incorrectly extracted the dependency SUPPORT[vsup-standard] from sentences like this:

(18) *(dar, saída) À noite, feito criança no mato, ele **dá** uma saidinha e **volta** com duas rãs e um sapo.*

(**take**, **walk**) 'At night, like a child in the bush, he takes (lit. **gives**) a little walk and **returns** with two frogs and a toad.'

SUPPORT[vsup-standard](volta,dá)

In this sentence, the diminutive form of *saidinha* (derived from *saída* 'walk') was not recognized by the system, reason why, although it was assigned the PoS noun, it was not possible to extract the dependency SUPPORT, which requires the identification of the lemma. On the other hand, the PoS disambiguation of *volta* 'return' was not properly made and the word was labeled as a noun, when the verb PoS tag should have been chosen. Now, since there is an additive conjunction *e* 'and' between *saidinha* 'walk' and *volta* 'return', and the latter was marked as a name, the system analyzed this sequence as the coordination of two nouns. In the second moment, the direct complement dependency (CDIR) between *dar* 'give' and *saidinha* 'walk' is percolated to the noun coordinated with the latter, which triggers the rule that extracts the SUPPORT dependency between *dar* 'give' and *volta* 'walk'.

Other processing problems have also been considered in cases where the SVC is partially identical to a fixed or frozen, idiomatic, expression (idioms) and the STRING processing chain extracts two dependencies for the same constituents. STRING has a frozen expressions' analysis module (BAPTISTA *et al.*, 2014), whose lexicon contains a few thousand idioms and it uses FIXED dependency extraction rules to capture the verb and the fixed elements of those idiomatic expressions. Consider next example:

(19) **(dar, volta)** *Até lá, não custa nada ter esperança de que pelo menos um grande clube carioca está **dando a volta por cima** e reconquistando seu lugar de honra na elite do futebol nacional.*

'(**give**, **turn**) Until then, it does not hurt to hope that at least one great club from Rio de Janeiro is turning around (lit. **giving the turn on top**) and regaining its place of honor in the national soccer elite.'
SUPPORT[vsup-standard](volta,dando)
FIXED (dando,volta,cima)

In this sentence, the expression *dar a volta por cima* 'turn around' was analyzed by STRING in two different ways, and two different dependencies were extracted for the same constituents: (i) as SUPPORT[vsup-standard], similar to the predicate *dar um passeio* 'take a walk'; and (ii) as a fixed construction (FIXED), meaning 'turn around'. It should be noted that most of the annotators have assigned the OTHER tag to this phrase, which may correspond to a fixed expression.

To fix these problems, a general "cleanup" rule was created, which gives preference to extracting the FIXED dependency and excludes the SUPPORT dependency. At the end of processing, only the second dependency (as a fixed or idiomatic expression) should remain, which is often the correct analysis.

Other fixed expressions were analyzed incorrectly by both the STRING system and most annotators, e.g.:

(20) *(dar, tiro) O PT está **dando um tiro no** próprio **pé** ao tentar abortar a CPI do caso Waldomiro Diniz.*

 '(**give**, **shot**) The PT is firing on (lit. **giving** a **shot**) its own **foot** trying to abort the CPI of the Waldomiro Diniz case.'
SUPPORT[vsup-standard](tiro,dando)

(21) *(dar, passo) Lee-Huang **deu** um **passo à frente** em relação à pesquisa de Gallo, diz David Lewi, infectologista da Unifesp.*

 '(**take**, **step**) Lee-Huang stepped up on (lit. **took** a **step** ahead of) Gallo's research, says David Lewi, an infectologist at Unifesp.'
SUPPORT[vsup-standard](passo,deu)

In (20), the elements of the SVC (*dar* 'give' and *tiro* 'shot') form a subset of the frozen expression *dar tiro no pé* 'fire (lit. give a shot) on own foot'. In (21), the frozen expression *dar um passo à frente* 'take a step ahead of' is ambiguous with the SVC *dar um passo* 'take a step'. From the semantic point of view, both can literally mean 'move one leg forward', or, metaphorically, 'move on, overcome some challenge'. The same problem occurs with other expressions, such as *dar o primeiro passo* 'take the first step', *dar um passo decisivo* 'take a decisive step', *dar passos firmes* 'take firm steps', etc., in which the *SVC* construction of the noun *passo* 'step' is probably at the origin of these idiomatic constructions.

Both the manual annotation and the automatic classification of these cases should be reviewed in order to maintain the consistency of the linguistic description. The frozen expressions of Brazilian Portuguese were described by Vale (2001) and many of them have already been classified in European Portuguese (BAPTISTA et al., 2004) and inserted in the lexicon of STRING (BAPTISTA; MAMEDE; MARKOV, 2014).

Another false-positive case concerns the classification of constructions with linking operator-verb as if they were constructions with support verb. The two following examples are cases of constructions with linking operator-verb:

(22) *(ter, nome) Em 94, vários delegados denunciados por Luz **tiveram** os **nomes** encontrados nos livros de contabilidade do jogo do bicho.*

 '(**have**, **name**) In 94, several delegates denounced by Luz **had** the[ir] **names** found in the accounting books of the illegal lottery.'
SUPPORT[vsup-standard](nomes,tiveram)

(23) *(ter, nome) O participante que **tiver** o **nome** confirmado deverá se dirigir à Bovespa, rua XV de novembro, 275, centro, São Paulo, no horário marcado, munido de identidade.*

 '(**have**, **name**) The participant that **has** [his/her] **name** confirmed should address the Bovespa, November 15 Street, 275, center, São Paulo, at the time scheduled, with [his/her] identity [card].'
SUPPORT[vsup-standard](nome,tiver)

In both cases, the *Npred nome* 'name' was identified as a direct complement (`CDIR`) of the verb *ter* 'have', and so the two dependency extraction rules `SUPPORT[vsup-standard=+]` and `SUPPORT[vsup-converse=+]` were triggered. Because of the rule that selects the standard dependency in ambiguous cases, the system extracts only the dependency `SUPPORT[vsup-standard=+]`. These cases, however, correspond to constructions with a linking operator-verb, as was noted manually by most annotators.

The linking operator-verb is a concept introduced by Gross (1981) and later studied by Ranchhod (1990), for European Portuguese, to refer to verbs that operate on base constructions, adding to them an argument which is already present in the base sentence in the complement position.

(24) **[built example]**: *O Rui tem # A Ana está sob o controle do Rui.*
'Rui has # Ana is under the control of Rui.'

(24a) = *O Rui$_i$ tem a Ana sob o (seu$_i$ + *meu + *teu) controle.*
'Rui has Ana under his control.'

Some cases had been classified by both humans and STRING, in the first evaluation, as `SUPPORT[vsup-standard]`. After a systematic review of the annotations of the reference *corpus*, however, we considered that these are other cases of linking operator-verb. In general, the problem is related to nouns such as *notícia* 'news', *orientação* 'guidance', *informação* 'information', *explicação* 'explanation', *opinião* 'opinion', *solução* 'solution', *resposta* 'response/answer', *exemplo* 'example', *definição* 'definition', *dica* 'hint/tip', *pista* 'clue', *sugestão* 'suggestion', *argumento* 'argument', *parecer* 'opinion', *etc.*, associated with *Vsup ter* 'have', which admits two different interpretations, one with passive sense, shown in (25), and another with an active sense, shown in (26), e.g.:

(25) **[built example]**: *Zé **teve** uma **notícia** ruim <quando Ana lhe contou sobre sua doença>.*
'Zé **had** some bad **news** <when Ana told him about her illness>.'

(26) **[built example]**: *Zé **tem** uma **notícia** ruim <para dar à Ana>.*
'Ze **has** bad **news** <to give Ana>.'

Verbal tenses, strictly speaking the 'punctual' aspect of the perfect tense in (25), and the 'durative' aspect of the present (or the imperfect) in (26) allow us to distinguish these two uses. The example (25) is less controversial, clearly being considered a converse *SVC*, since it has as its counterpart the standard construction:

(25a) ≡ *A Ana **deu** uma **notícia** ruim ao Zé.*
'Ana **gave** some bad **news** to Zé.'

The statute of this converse construction did not raise any doubts among the annotators. On the other hand, the same pair (*ter* 'have', *notícia* 'news'), in (26), seems to have a special status, since it resembles an active orientation construction (e.g. *Ana deu uma notícia ruim ao Zé* 'Ana gave a bad news to Zé'), but the action is not fully achieved (imperfective aspect).

The basic predicate in (26) is *dar uma notícia* 'give a [piece of] news', since it can be reconstituted in the infinitive sub-clause introduced by *para* 'to' (e.g. *Zé tem uma notícia ruim para dar à Ana* 'Ze has some bad news to give to Ana'). The verb *ter* 'have', in this sense, serves only to connect the subject argument (*Zé*) to the predicate *dar uma notícia* 'give some news'. This argument (*Zé*) is not new, as it already existed in the base sentence. In these conditions, the verb *ter* 'have', in (26), also has the status of a linking operator-verb.

The same phenomenon can be observed in several other *Npred* associated with the verb *ter* 'have'. Sentences (27) to (32), taken from the reference *corpus*, exemplify the problem.

(27) *(ter, notícia) Segundo Zeca, o Estado vizinho de Mato Grosso tem "quase uma dezena de usinas instaladas na bacia do Paraguai sem que se **tenha** tido **notícia** de um único acidente ambiental".*
   '(**have**, **news**) According to Zeca, the neighboring state of Mato Grosso has "almost a dozen factories installed in the basin of Paraguay and one **has** no **news** of a single environmental accident."'
SUPPORT[vsup-standard](notícia,tenha)

(28) *(ter, informação) A delegada diz que é importante que os passageiros que sejam furtados ou roubados registrem a ocorrência na delegacia do aeroporto, para que a polícia **tenha** mais **informações** sobre o modo como os bandidos agem.*
   '(**have**, **information**) The delegate says that it is important that passengers who have been robbed make a record of the occurrence at the airport police station, so that the police **have** more **information** on how the criminals act.'
SUPPORT[vsup-standard](informações,tenha)

(29) *(ter, solução) A disputa entre juízes e a direção da liga, que aparentemente **teria** uma **solução** rápida, deve durar algumas rodadas.*
   '(**have**, **solution**) The dispute between judges and the league's management, which apparently would **have** a quick **solution**, should last a few rounds'
SUPPORT[vsup-standard](solução,teria)

(30) *(ter, notícia) O "The Wall Street Journal" **tem** boas **notícias** para todos vocês, ratos de sofá.*
   (**have**, **news**) '"The Wall Street Journal" **has** good **news** for all of you, couch rats'
SUPPORT[vsup-standard](notícias,tem)

(31) *(ter, informação) A página* **tem informações** *sobre o clube, fotos e os nomes dos membros.*
     '(**have**, **information**) The page **has informations** about the club, photos and the names of members'
     SUPPORT[vsup-standard](informações,tem)

(32) *(ter, solução) Quem ousaria dizer que* **tem a solução** *para o caso?*
     '(**have**, **solution**) Who would dare say that he **has** the **solution** to the case?'
     SUPPORT[vsup-standard](solução,tem)

For all these cases, STRING extracted the SUPPORT[vsup-standard] dependency. The first three examples (27), (28) and (29), however, are converse *SVC* and the last three (30), (31) and (32) should not have been extracted, since they correspond to constructs with linking operator-verb. As the reference *corpus* itself was also incorrect in these cases, it was subsequently revised and corrected.

b) False-Negative

As presented in Table 5, the SUPPORT dependency was not extracted from 114 sentences. The main reasons for the desired dependency not being identified is the distance between the support verb and the predicative noun or, else, the failure of the processing chain in some previous stage.

(33) *(dar, mostra) Para fortalecer essa espécie de revolução democrática que, iniciada com a decisão de desalojar um usurpador do poder Executivo,* **dá mostras** *agora de que deseja ir fundo na moralização e na republicanização do Poder Legislativo.*
     '(**give**, **sign**) In order to strengthen this kind of democratic revolution, which, having began with the decision to evict an usurper from the executive power, now **gives signs** that it will go deep into the moralization and republicanization of the Legislative Branch.'

The nominal predicate *dar mostras* 'give signs' is found in a relative-restrictive subclause, but the relative pronoun is separated from the supporting verb by a parenthetical, appositive sub-clause (a so-called participle-reduced sub-clause). Given the complexity of the sentence and the interaction of the rules of grammar, the parser analyzes *mostras* 'signs' as the subject (SUBJ) and not as the direct complement (CDIR) of *dar* 'give'. For this reason, the SUPPORT dependency has not been extracted. It should be noted that, in a simpler sentence, the parser's analysis is already adequate:

(33a) **[Simplified Example]** *A revolução democrática **dá mostras** de que deseja ir fundo na moralização.*

'The democratic revolution **gives signs** that it will go deep into the moralization.'
`SUPPORT[vsup-standard](mostras,dá)`

In other cases, the system did not properly extract the `SUPPORT` dependency, due to failure in processing some type of anaphora, a step that occurs in a previous stage of the chain, before the `SUPPORT` dependency extraction step, *e.g.*:

(34) *(dar, declaração) A **declaração** não tem valor legal, já que não foi **dada** em um depoimento formal.*

'**(give**, **statement**) The **statement** has no legal value, since [it] was not **given** in a formal way.'

In this example, the *Npred declaração* 'statement' is the subject of a passive construction that is subordinated to the main clause. However, the subject of that subclause has been elided (*já que [**essa declaração**] não foi **dada*** 'since [it=**this statement**] was not **given'**), since it already occurs as the subject of the main clause. Although STRING contains a module for the resolution of this type of anaphora (PEREIRA; ZAC, 2010), in this case the system failed to capture adequately the elliptical subject and, therefore, did not extract the `SUPPORT` dependency of the *SVC*. However, the system adequately captures the passive nominal construction in a sentence whose subject is explicit:

(34a) **[Simplified example]** *A **declaração** não foi **dada** em um depoimento formal.*

'The **statement** was not **given** in a formal way'
`SUPPORT[vsup-standard](declaração,dada)`

The 114 nominal predicates were tested individually in the processing chain using simple sentences as examples (simplified examples). For all of them, it was possible to obtain the adequate analysis, which means that the problem is not in the linguistic data formalized in the Lexicon-Grammar, but results from the complex process of the parser analysis. In some cases where the coordinate conjunction is not explicit and has been replaced by a comma, the system has not been able to extract the dependency:

(35) (**dar, amasso**) *Ah, Lorena, você só **dá** uns **beijinhos** nele, uns **amassos** e pronto.*

'(**give**, **make out**) Oh, Lorraine, you just **give** a few **kisses** on him, some **make out**, and that's it.'
`SUPPORT[vsup-standard](beijinhos,dá)`

In this sentence, only the dependency between *dar* 'give' and *beijinhos* 'kisses' was extracted. The dependency between the target pair {*dar, amasso*} 'give, make out'

was not identified by STRING. It turns out that, in this example, the word *pronto* was analyzed as an adjective and it is not in a formal context that allows to form a nominal phrase that would then be coordinated with the nouns *beijinhos* 'kisses' and *amassos* 'make out'. This is the reason why the rules of coordination did not fire; therefore, the coordination between *beijinhos* and *amassos* has not been extracted either and the SUPPORT could not be made to percolate from the first to the second noun.

As it is currently implemented in STRING, coordination is treated as a strictly local phenomenon, linking nominal and/or prepositional phrases, including cases of enumeration of 3 or more elements, in which there are ellipses of intermediate conjunctions (*e.g. laranjas, bananas e maçãs* 'oranges, bananas, and apples'). If we make explicit the coordinative conjunction between the two noun phrases of the sentence (35), *uns beijinhos* 'some kisses' and *uns amassos* 'some make-out', then the system correctly extracts the two SUPPORT dependencies:

(35a) **[Simplified example]** *Ah, Lorena, você só* **dá** *uns* **beijinhos** *e uns* **amassos** *nele e pronto.*

  'Ah, Lorraine, you just **give** a few **kisses** <u>and</u> some **make out** on him and that's it.'
SUPPORT[vsup-standard](amassos,dá)
SUPPORT[vsup-standard](bejinhos,dá)

It should be noted that in other cases of coordination between predicative nouns with the same support verb, STRING correctly extracted the SUPPORT dependency, as in:

(36) *(ter, aprovação) Três cadernos "Guerra na América" (12/9, 13/9 e 14/9), contados à parte,* **tiveram** *a maior* **leitura** <u>e</u> *a maior* **aprovação** *da semana (média de 95% do ótimo/bom).*

  '(**have**, **approval**) Three notebooks "War in America" (12/9, 13/9 and 14/9), counted separately, **had** the highest **reading** <u>and</u> highest **approval** ratings of the week (average of 95% of the very good/good [ratings]).'
SUPPORT[vsup-converse](aprovação,tiveram)
SUPPORT[vsup-converse](leitura,tiveram)

(37) *(receber, confirmação) Até hoje e apesar do prazo fixado para este efeito em 20 de junho de 2005, este Ofício não* **recebeu** <u>nem</u> **resposta** <u>nem</u> *a* **confirmação** *de procedimentos feitos pelas autoridades brasileiras competentes necessários à retirada da documentação suíça.*

  '(**receive**, **confirmation**) To date, and despite the deadline set for this purpose on June 20, 2005, this Office has **received** <u>no</u> response <u>nor</u> **confirmation** of procedures performed by the competent Brazilian authorities [that were] necessary to the withdrawal of the Swiss documentation.'
SUPPORT[vsup-converse](confirmação,recebeu)
SUPPORT[vsup-converse](resposta,recebeu)

In the two examples, the direct complement dependencies (CDIR) that were extracted between {*ter, leitura*} 'have, reading' and {*receber, resposta*} 'receive, answer/response' were percolated to the other nominal groups with which these nouns are coordinated, namely {*ter, aprovação*} 'have, approval' and {*receber, confirmação*} 'receive, confirmation', respectively.

c) True-negatives

Of the 139 cases in which STRING did not extract SUPPORT dependency, there are 25 sentences for which the system, in fact, should not have extracted the dependence, since there is no syntactic relation between *Vsup* and *Npred*. The human annotators also did not consider that there was a construction with support verb in these sentences. These are, therefore, true-negative cases.

(38) *(receber, título) Por exemplo, no documento dizia que foi **recebido** a **título** de horas extras CR$ 200 mil.*
   '(**receive, overtime**) For example, in the document it was said that CR $ 200 thousand was **received** as **overtime**.'

(39) *(levar, ponto) Roteiro de um dia **leva** aos **pontos** altos de San Francisco.*
   '(**lead**, **highlight**) A one-day tour **leads** to the **highlights** of San Francisco.'

(40) *(ter, sorte) Sua **sorte** foi **ter** sido socorrido com rapidez.*
   '(**have**, **luck**) His **luck** was that he **has** been rescued quickly.'

(41) *(ter, aula) Quando estiver pronto será a sede da administração do campus e também **terá** salas de **aula**.*
   '(**have**, **class**) When it will be ready, it will be the headquarters of the campus administration and it will also **have class** rooms.'

As we said above, the extraction of sentences from the PLN.Br *corpus* was based only on the fact that the pair *Vsup Npred* was present in the sentence and it did not consider any syntactic relations. For this reason, such sentences selected for the sample are counterexamples that contribute to measuring the quality of the parsing process.

Particular Cases of True-Positives

The previous section presented the main problems found in the automatic extraction of the SUPPORT dependency in sentences taken from the *corpus*. In addition to these, it should be highlighted that in some cases the system detected a SUPPORT relation,

although they were not the target for which the sentence had been extracted from the *corpus*.

In the examples below, both the target pair and the pair extracted by STRING are shown in bold. The target pair (automatically extracted from the *corpus*, using Unitex) is inserted at the beginning of the example, in parentheses; the pair {*Vsup*, *Npred*}, not targeted but properly parsed by STRING, is underlined in the body of the example.

(42) *(dar, show) Os ingressos custam R$ 30,00 e <u>darão</u> <u>direito</u> a diversos shows de dança e música e a um jantar típico com especialidades da China, Japão, Coréia, Tailândia, Indonésia e Índia.*
   '(**give**, **show**) Tickets cost R$ 30,00 and will **entitle** (lit: <u>**give right**</u>) to various dance and music **shows** and a typical dinner with specialties from China, Japan, Korea, Thailand, Indonesia and India.'
`SUPPORT[vsup-standard]`(direito,darão)

(43) *(dar, nome) Segundo Vassoureiro, esse costume, que ainda persiste, <u>deu</u> <u>origem</u> ao nome "papangu".*
   '(**give**, **name**) According to Vassoureiro, this custom, which still persists, <u>**gave rise**</u> to the **name** "papangu"'
`SUPPORT[vsup-standard]`(origem,deu)

(44) *(ter, nó) Pollack - Os europeus têm as mesmas <u>informações</u> que nós temos.*
   'Pollack - Europeans **have** the same <u>**information**</u> as we do (lit. that we **have**).'
`SUPPORT[vsup-converse]`(informações,têm)
`SUPPORT[vsup-converse]`(informações,temos)

(45) *(receber, comissão) Sua candidatura ao COI <u>recebeu</u> o <u>aval</u> da comissão executiva da entidade, que é formada por 11 pessoas, entre elas o presidente, Juan Antonio Samaranch.*
   '(**receive**, **comission**) His candidature to the IOC has <u>**received**</u> the <u>**endorsement**</u> of the executive **committee** of the organization, which is formed by 11 people, among them the president, Juan Antonio Samaranch.'
`SUPPORT[vsup-converse]`(aval,recebeu)

Since STRING performs both a shallow parsing (chunking) and deep syntax processing (with extraction of dependencies between constituents), the system recognizes constituents that actually have some relation and ignore those that do not.

It is also worth noting that, in (44), the chain correctly extracted two dependencies: one, in which *informações* 'informations' is a direct complement of *ter* 'have', and another, in which *informações* 'informations' is the antecedent of the relative pronoun, which is a direct complement of *ter* 'have'.

By integrating the Lexicon-Grammar data into STRING and its automatic syntactic analysis, it was also possible to identify other pairs of *Vsup* and *Npred* that had not previously been considered, such as:

(46) **(ter, condição)** *Como o MEC não **tem condições** de fiscalizar todos os 5.506 municípios brasileiros, pretende **contar com** a **ajuda** dos Estados.*
     '**(have, condition)** As the MEC is not able (lit: does not **have** the **conditions**) to inspect all 5,506 Brazilian municipalities, it intends to **count on the help** of the States.'
`SUPPORT[vsup-converse]`(ajuda,contar com)

In addition to the target pair {*ter, condição*} 'have, condition' that allowed this phrase to be extracted automatically from the *corpus* using Unitex, STRING also identified the pair {*contar com, ajuda*} 'count on, help', for which it correctly extracted the `SUPPORT[vsup-converse]` dependency.

Second evaluation of the system's performance

As mentioned above, after the error analysis, the linguistic data of the Lexicon-Grammar was corrected in the matrix and the sentences of the sub-sample were processed again. Table 6 presents and compares the results of the first and second evaluations.

Table 6 – First and second evaluations of STRING performance.

	TP	FP	FN	TN	Precision	Recall	Accuracy	F-measure
First evaluation	350	91	114	25	79%	75%	65%	77%
Second evaluation	325	56	84	115	85%	87%	80%	86%

Source: Author's elaboration.

As it can be noticed, the system's performance improved in the second evaluation run: Precision increased by 6%, Recall augmented 12%, Accuracy rose 15%, and, consequently, the F-measure also rose by 9%.

It is worth highlighting the significant increase in the number of true-negative from the first (TN = 25) to the second (TN = 115) evaluation run. The first evaluation considered as the golden standard the annotation of the majority or the unanimity of the annotators, without checking whether that annotation was consistent or not. By systematically reviewing annotations, we identified all cases in which the verb *ter* 'have' functioned as a linking operator verb and we correct that data in the reference *corpus*. Therefore, this has significantly increased the number of sentences in which STRING aptly did not extract the `SUPPORT` dependencies.

The improvement in the system performance is also due to the correction of Lexicon-Grammar data, as the dependencies that were being extracted as `SUPPORT[vsup-converse]` became `SUPPORT[vsup-standard]`.

In addition to correcting the reference *corpus* and the linguistic data of Lexicon-Grammar, we inserted in the dictionary used by STRING (v.g. **PB.dic**), the degree inflection/derivation information for the nouns ending in *-ada*/*-ida*. The dependencies of nouns such as *saidinha* 'little exit', *fugidinha* 'little escape' e *arrumadinha* 'little tidy up' were not being extracted because the system did not recognize these nouns in *saída* 'exit', *fugida* 'escape' e *arrumada* 'tidy up', respectively. By associating the adequate inflection/derivation paradigm to these nouns, STRING now correctly extracts the dependencies involving them.

In addition, a "cleaning" rule was created for the cases of frozen (idiomatic) expressions. STRING extracted at the same time the FIXED and SUPPORT dependencies for the constructions whose constituents either were frozen expressions or form an *SVC*. The rule created for the processing chain gives preference to the extraction of the FIXED dependency and excludes the SUPPORT dependency, in cases of duplicated dependencies between the same constituents.

Conclusions and future work

As a result of this work, we produced a golden standard of constructions with the support verb *dar* 'give' and its variants for Portuguese. This golden standard consists of a *corpus* annotated automatically with the `SUPPORT[vsup-standard]` and `SUPPORT[vsup-converse]` dependencies by STRING, and then manually revised by a team of linguists.

The results of the task indicate gains in the different parameterization of the rules. It should be noticed, however, that the experiments were done for a small set of *SVC*, involving only one elementary *Vsup* and its variants.

In the near future, we also intend to integrate into STRING the Lexicon-Grammar matrices referring to the nominal constructions with the Vsup *fazer* 'do/make' and *ter* 'have', and to evaluate, in a more comprehensive way, the performance of the system, using in full the 2,646 manually annotated sentences.

One of the specificities of Brazilian Portuguese in relation to European Portuguese is the great productivity of predicative nouns that can be created by derivation with the suffix *-ada*/*-ida*. Virtually all verbs of action and many verbs denoting processes can give rise to predicative nouns with the suffix *-ada*/*-ida* (SCHER, 2004), which select for the most part the support verb dar 'give', such as *abanar* = *dar uma abanada* 'shake = give a shake-ada', *enxugar* = *dar uma enxugada* 'dry = give a dry-ada', *enrugar* = *dar uma enrugada* 'wrinkle = give a wrinkle-ada', *crescer* = *dar uma crescida* 'grow = give a grow-ida', *sumir* = *dar uma sumida* 'disapear = give a disapear-ida (=disapearance)', etc. In the same way, some nouns designating objects, instruments and body-part nouns

can also be given the suffix -*ada* to form predicative nouns (BAPTISTA, 2004), such as *bater com uma cadeira = dar uma cadeirada* 'hit (someone) with a chair = give a chair-ada', *bater com o cinzeiro = dar uma cinzeirada* 'hit (someone) with a ashtray = give a ashtray-ada', *bater com o ombro = dar uma ombrada* 'hit (someone) with the shoulder = give a shoulder-ada', etc.). This phenomenon is quite productive in the nominal constructions with *Vsup dar* 'give'. In this sense, it is intended, in future works, to expand the list of predicative nouns and to integrate them into the dictionaries of STRING so that more constructions can be identified in real texts.

However, these *N-ada* nouns are often ambiguous with the past participle of the corresponding verbs, which raises several processing problems, especially for the difficulty in part-of-speech (POS) disambiguation.

Acknowledgements

- *RESUMO: Este artigo descreve a metodologia para a integração de predicados nominais, do tipo construções com verbo-suporte (CVS), no analisador sintático automático XIP, que é utilizado pela cadeia de processamento do Português STRING. Trata-se, mais especificamente, de 580 CVS com o verbo dar e um nome predicativo, cujas propriedades sintático-semânticas foram descritas, formalizadas e, em seguida, integradas à gramática do XIP, por meio de regras, a fim de extrair a dependência SUPPORT entre o nome predicativo (Npred) e o verbo-suporte (Vsup). A necessidade de tratar automaticamente as CVS decorre do fato de que elas são diferentes de construções com verbo pleno, possuem estruturas sintáticas complexas, possuem propriedades sintático-semânticas específicas e admitem transformações sintáticas sistemáticas, ainda que lexicalmente determinadas. O conceito de CVS, bem como a abordagem léxico-sintática adotada, segue os princípios teóricos e metodológicos do Léxico-Gramática. Como resultado da integração desses dados ao parser XIP, o sistema atingiu precisão de 85%, abrangência de 87%, acurácia de 80% e medida-F de 86%.*

- *PALAVRAS-CHAVE: Verbo-suporte. Nome predicativo. Construção com verbo-suporte. Verbo-operador causativo. Parser XIP.*

REFERENCES

BAPTISTA, J. Instrument nouns and fusion. Predicative nouns designating violent actions. In: LECLÈRE, C.; LAPORTE, E.; PIOT, M.; SILBERZTEIN, M. (Eds.). Lexique, Syntaxe et Lexique-Grammaire (Syntax, Lexis & Lexicon-Grammar), Homenagem a Maurice Gross, **Linguisticae Investigationes** Supplementa Amsterdam/Philadelphia: John Benjamins Publishing Comp, 2004. p.31-40.

BAPTISTA, J. **Sintaxe dos predicados nominais com *SER DE***. Lisboa: Fundação Calouste Gulbenkian/Fundação para a Ciência e Tecnologia, 2005.

BAPTISTA, J.; CORREIA, A.; FERNANDES, G. Frozen Sentences of Portuguese: Formal Descriptions for NLP. In: **Proceedings of the Workshop on Multiword Expressions: Integrating Processing**, EACL. Barcelona, Spain, July, 2004. p.72-79.

BAPTISTA, J.; MAMEDE, N.; MARKOV, I. Integrating a lexicon-grammar of verbal idioms in a Portuguese NLP system. In: **WG2:** Parsing techniques for MWEs, PARSEME meeting, 10-11 March, Athens, 2014.

BARREIRO, A.; MONTI, J.; ORLIAC, B.; PREUß, S.; ARRIETA, K.; LING, W.; BATISTA, F.; TRANCOSO, I. Linguistic Evaluation of Support Verb Constructions by OpenLogos and Google Translate. In: CALZOLARI, N.; CHOUKRI, K.; DECLERCK, T.; LOFTSSON, H.; MAEGAARD, B.; MARIANI, J.; MORENO, A.; ODIJK, J.; PIPERIDIS, S. (Eds.). **Proceedings of LREC'14**. Ninth International Conference on Language Resources and Evaluation. European Language Resources Association (ELRA), May, Reykjavik, Iceland. 2014. p.35-40.

BARROS, C. D. de. **Descrição e classificação dos predicados nominais com o verbo- suporte FAZER em Português do Brasil**. 2014. 270 f. Tese (Doutorado em Linguística) – Centro de Educação e Ciências Humanas, Universidade Federal de São Carlos, São Carlos, 2014.

BICK, E. **The Parsing System Palavras, Automatic Grammatical Analysis of Portuguese in a Constraint Grammar Framework**, Aarhus University Press, 2000.

BRUCKSCHEN, M.; MUNIZ, F.; SOUZA, J. G. C.; FUCHS, J. T.; INFANTE, K.; MUNIZ, M.; GONÇALVES, P. N.; VIEIRA, R.; ALUISIO, S. Anotação linguística em XML do *corpus* PLN-BR. **Série de relatórios do NILC**, NILC - ICMC - USP, 2008.

BUTT, M.; GEUDER, W. On the (semi)lexical status of light verbs. In: CORVER, N.; RIEMSDIJK, H. (Eds.). **Semi-lexical categories**. Berlin, Germany: Mouton de Gruyter, 2001. p.323-370.

CALZOLARI, N.; FILLMORE, C. J.; GRISHMAN, R.; IDE, N.; LENCI, A.; MACLEOD, C.; ZAMPOLLI, A. Towards best practice for Multiword Expressions in Computational Lexicons. In: **Third International Conference on Language**

Resources and Evaluation, LREC. Las Palmas, Canary Islands - Spain, May, 2002. p.1934-1940.

COHEN, J. A coefficient of agreement for nominal scales. **Educational and Psychological Measurement,** 20 (1), p.37-46, 1960.

DIAB, M.; BHUTADA, P. Verb Noun Construction MWE Token Supervised Classification. In: **Proceedings of the Workshop on Multiword Expressions:** identification, interpretation, disambiguation and applications, MWE'09. Association for Computational Linguistics, Stroudsburg, PA, USA, 2009. p.17-22.

DURAN, M.; RAMISCH, C.; ALUISIO, S.; VILLAVICENCIO, A. Identifying and analyzing Brazilian Portuguese complex predicates. In: **Proceedings of MWE'11**. Workshop from Parsing and Generation to the Real World. Association for Computational Linguistics, 2011. p.74-82.

GIRY-SCHNEIDER, J. **Les nominalisations en français:** l'opérateur *faire* dans le lexique. Genève: Librairie Droz, 1978.

GIRY-SCHNEIDER, J. **Les prédicats nominaux en français:** les phrases simples à verbes support. Genève: Librairie Droz, 1987.

GREFENSTETTE, G.; TEUFEL, S. *Corpus*-based method for automatic identification of support verbs for nominalizations. In: **Proceedings of EACL'95**. 7th Conference of the European Chapter of the Association for Computational Linguistics, March, Sttutgart, Germany, 1995.

GROSS, M. **Méthodes en syntaxe**. Paris: Hermann, 1975.

GROSS, M. Les bases empiriques de la notion de prédicat sémantique. **Langages**, 63 (15), p.7-52, 1981.

GROSS, M. The lexicon grammar of a language: application to french. In: ASHER, R. E. **Encyclopedia of Language and Linguistics**. London: Pergamon, 1994. p.2195-2205.

GROSS, M. La fonction sémantique des verbes supports. **Travaux de linguistique**, 37, p.25-46, 1998.

HARRIS, Z. **A Theory of Language and Information:** a mathematical approach. New York: Oxford University Press, 1991.

ISTVÁN, N.; VINCZE, V.; FARKAS, R. Full-coverage Identification of English Light Verb Constructions. In: **Proceedings of IJCNLP**, 2013.

LAMIROY, B. Le Lexique-grammaire: Essai de synthèse. In: LAMIROY, B. (Ed.). **Travaux de Linguistique**, 37, p.7-23, 1998.

MAMEDE, N.; BAPTISTA, J.; DINIZ, C.; CABARRÃO, V. String: an hybrid statistical and rule-based natural language processing chain for Portuguese. **International**

Conference on Computational Processing of Portuguese (Propor 2012), Demo Session. Coimbra, Portugal, April, 2012.

MEUNIER, A. **Nominalisations d'adjectifs par verbes supports**. 1981. 215 f. Tese (Thèse de Troisième cycle) – Laboratoire Automatique Documentaire et Linguistique, Université Paris 7, 1981.

MOKHTAR, S. A.; CHANOD, J. P.; ROUX, C. Robustness beyond shalowness: Incremental dependency parsing. **Natural Language Engineering**, p.121-144, 2002.

PÁEZ, S. M. C. Extraction et représentation des constructions à verbe support en Espagnol. In: **Proceedings of ACL**. 21ème Traitement Automatique des Langues Naturelles, Marseille, 2014. p.419-424.

PEREIRA, S.; ZAC, P. B. An Annotated *Corpus* for Zero Anaphora Resolution in Portuguese. In: **Proceedings of the Student Research Workshop. Association for Computational Linguistics**, Borovets, Bulgaria, 2010. p.53-59.

RANCHHOD, E. M. **Sintaxe dos predicados nominais com *ESTAR***. INIC – Instituto Nacional de Investigação Científica, Lisboa, 1990.

RANCHHOD, E. M. Groupes nominaux négatifs issus de la réduction de verbes supports: Exemples du portugais, de l'anglais et du français. **Lingvisticae Investigationes**, 27 (2), p.283-294, 2005.

RASSI, A. P. **Descrição, classificação e processamento automático das construções com o verbo DAR em Português Brasileiro**. 2015. 327 f. Tese (Doutorado em Linguística) – Centro de Educação e Ciências Humanas, Universidade Federal de São Carlos, São Carlos, 2015.

RASSI, A. P.; BAPTISTA, J.; MAMEDE, N.; VALE, O. A. Integrating support verb constructions into a parser. In: **Atas do Symposium in Information and Human Language Technology** (STIL'2015), 04-06 November 2015, Natal, Rio Grande do Norte, Brazil, 2015.

RASSI, A. P.; BAPTISTA, J.; VALE, O. A. Um *corpus* anotado de construções com verbo-suporte em Português. **Revista Gragoatá**, v.20, n.38, p.207-230, 2015.

SAG, I. A.; BALDWIN, T.; BOND, F.; COPESTAKE, A. A.; FLICKINGER, D. Multiword Expressions: A Pain in the Neck for NLP. In: GELBUKH, A. (Ed.) **Proceedings of the Third International Conference**, CICLing – Computational Linguistics and Intelligent Text Processing. Mexico City, Mexico, February17-23, 2002. p.1-15.

SANTOS, M. C. A. dos. **Descrição e classificação dos predicados nominais com o verbo-suporte TER em Português do Brasil**. 2015. 215 f. Tese (Doutorado em Linguística) – Centro de Educação e Ciências Humanas, Universidade Federal de São Carlos, São Carlos, 2015.

SCHER, A. P. **As construções com o verbo leve *dar* e nominalizações em *-ADA* no Português do Brasil**. 2004. 234 f. Tese (Doutorado em Linguística) – Instituto de Estudos da Linguagem, Universidade Estadual de Campinas, Campinas, 2004.

SILVA, J.; BRANCO, A.; CASTRO, S.; REIS, R. Out-of-the-Box Robust Parsing of Portuguese. In: **Proceedings of the 9th International Conference on the Computational Processing of Portuguese (PROPOR'10)**, 2010. p.75-85.

SUÍSSAS, G. **Verb Sense Disambiguation**. Dissertation Project. Universidade de Lisboa – Instituto Superior Técnico/INESC-ID Lisboa – Spoken Language Laboratory, 2014.

TU, Y.; ROTH, D. Learning English Light Verb Constructions: Contextual or Statistics. **Proceedings of ACL'11**. Workshop on Multiword Expressions: from Parsing and Generation to the Real World, 2011.

VALE, O. A. **Expressões Cristalizadas do Português do Brasil:** uma proposta de tipologia. 2001. 250 f. Tese (Doutorado em Letras) – Faculdade de Ciências e Letras, Universidade Estadual Paulista, Araraquara, 2001.

VIVÈS, R. **Avoir, prendre, perdre:** Constructions à verbe support et extensions aspectuelles. 1983. 338 f. Tese (Thèse de Troisième cycle) – Laboratoire d'Automatique Documentaire et Linguistique, Université Paris 8, Paris, 1983.