The fuzzy boundaries of operator verb and support verb constructions with *dar* "give" and *ter* "have" in Brazilian Portuguese

Amanda Rassi[1,2,3], **Cristina Santos-Turati**[1,2,3], **Jorge Baptista**[2,3], **Nuno Mamede**[3], **Oto Vale**[1]

{aprassi,mcturati,jbaptis}@ualg.pt, nuno.mamede@inesc-id.pt, otovale@ufscar.br

[1]UFSCar, Rodovia Washington Luís, km 235 - SP-310 São Carlos-SP, Brazil
[2]UAlg/CECL, Campus de Gambelas, 8005-139 Faro, Portugal
[3]INESC-ID Lisboa/L2F, Rua Alves Redol, n.º 9, 1000-029 Lisboa, Portugal

Abstract

This paper describes the fuzzy boundaries between support verb constructions (SVC) with *ter* "have" and *dar* "give" and causative operator verb (VopC) constructions involving these same verbs, in Brazilian Portuguese (BP), which form a complex set of relations: (i) both verbs are the support verb of the same noun (SVC); (ii) *dar* is the standard (active-like) SVC while *ter* is a converse (passive-like) SVC; and (iii) *dar* is a VopC, operating on a *ter* SVC. In this paper we have systematically studied these complex relations involving SVC and VopC for BP, which constitute a challenge to Natural Language Processing (NLP) systems, and have been often ignored in related work. The paper proposes a lexically-based strategy to implement SVC in a fully-fledged, rule-based parsing system, yielding an adequate semantic structure of the events (predicates) denoted by predicative nouns in SVC.

1 Introduction: basic concepts and a little history

The notion of support verb has been in use for a long time, under many different theoretical perspectives and various terminologies. In this paper, we adopt the Zellig S. Harris's (1991) transformational operator grammar framework. As early as 1964, Harris (1964, p.216-7) proposed the concept and named this particular type of construction as "U operator" nominalizations, linking sentences such as *He studies eclipes = He makes studies of eclipses*. It was, however, M. Gross (1981) who first provided the definition of support verb we will rely upon here. The support verb *make* (in the example above) can be seen as a sort of an auxiliary of the predicative noun *studies*, in charge of carrying the grammatical values of tense and person-number agreement that the noun is morphologically unable to express. In many cases, support verbs are practically devoid of meaning. For lack of space, we cannot detail further the properties of SVC, and only the briefest outline is provided here; a good overview can be found in (Gross, 1996; Gross, 1998; Lamiroy, 1998).

One of the most important theoretical contribution of the notion of support verb came from the fact that it provides a natural framework to adequately include in the kernel sentences of the language the large number of 'abstract' nouns, which do not have neither a verbal nor an adjectival counterpart; that is, they are *isolated* or *autonomous* nouns, lacking any nominalizations (in a synchronic perspective, at least). This phenomenon is particularly evident in Romance languages (French, Italian, Portuguese, Romanian and Spanish): FR: *Jean a fait grève* "Jean did strike"; IT: *Giovanni ha fatto sciopero* "*id.*"; PT: *O João fez greve* "*id.*"; RU: *Ioan a făcut grevă* "*id.*"; SP: *Joan hizo huelga* "*id.*"; cp. EN: **John did strike*, *John was on strike*).

Finally, nominal constructions are unlike any other predicative part-of-speech by the fact that predicative nouns can present more than one construction with different support verbs, while still expressing the same semantic predicate. Hence, for example, *greve* "strike" can have a SVC with both *fazer* "to make" (as above) and *estar em* "to be in": *O João está em greve* "João is on strike" (Ranchhod, 1990). Each SVC has its own specific properties, *e.g.* only SVC with *fazer* can undergo passive, while the general predicate remains the same.

In this paper, we also consider the concept of *operator verb* (*VopC*), introduced in the same paper (Gross, 1981, p. 23-39); two relatively clear situations were distinguished:

- a *causative operator verb* (*VopC*), which adds a new element to an elementary sentence; this element has an easily identifiable meaning: CAUSE; distributionally, this element suffers very loose constraints (and we define this as a *distributionally non constraint* position (*Nnr*)); if the base sentence under the operator is a support verb construction 1, the *VopC* may "absorb" the support verb and it may also introduce some formal changes in that sentence 1;

(1) *Isso* dá # *Max* tem (*fome + medo + sede*). [1] "This gives # Max has (hungry + fear + thirst)."

(2) *Isso* dá (*fome + medo + sede*) *em Max.* "This gives Max (hungry + fear + thirst)."

In (2), the support verb *ter* is absorbed under the operator *dar* and its subject becomes a dative, indirect complement, though the semantic roles of subject of *dar* (CAUSE) and of the subject of the predicative noun (EXPERIENCER), after this restructuring, remain the same.

- a *linking operator-verb* (*VopL*), which hardly modifies the meaning of the underlying sentence; it also adds an argument to the base sentence 1, but this is not a new one since it is bounded linked to a noun complement of the base sentence 1 (Ranchhod, 1990).

(3) *Max* tem # *Ana* está sob *o controle do Max.* "Max has # Ana is under Max's control."

(4) = *Max$_i$* tem *Ana sob o* (*seu$_i$ + *meu + *teu*) *controle.* "Max$_i$ has Ana under (his$_i$ + *my + *your) control."

This paper reports an ongoing research to systematically classify the predicative nouns built with the support verbs *dar* and *ter* in Brazilian Portuguese (Rassi and Vale, 2013; Santos-Turati, 2012). Similar work has already been developed for the European variety (Vaza, 1988; Ranchhod, 1990; Baptista, 1997; Baptista, 2005). For many languages, including Portuguese, the studies on support verb constructions and causative constructions use a lexical approach, aiming at building dictionaries or lists of predicative nouns or at identifying those constructions (semi)automatically, *e.g.* for Portuguese (Hendrickx et al., 2010; Duran et al., 2011), for English (Grefenstette and Teufel, 1995), for German (Hanks et al., 2006; Storrer, 2007) and many other languages. As far as we could ascertain, no implementation of these SVC constructions has been made yet in NLP systems, particularly in parsers. Most systems considering these constructions just treat them as multiword expressions, ignoring their internal syntactic structure.

In this paper, we will show the complex set of relations involved in these SVC, where these verbs can function not only as support but also as operator verbs, thus rendering their description remarkably complex, particularly in view of Natural Language Processing. We aim at capturing the syntactic dependencies involved in these expressions, not as multiword, fixed strings, but as analyzable syntactic structures.

The paper is structured as follows: Next, Section 2 presents the current state of the collection and classification if these SVC in Brazilian Portuguese; Section 3 illustrates the syntactic-semantic relations between different constructions of *ter* and *dar*; Section 4 proposes a strategy for implementing the data in a rule-based parsing system; and, finally, Section 5 presents some concluding remarks and perspectives on future work.

2 Support verb constructions with *ter* "have" and *dar* "give"

The predicative nouns in this paper select the support verbs *dar* "give" and *ter* "have", and were retrieved from previous lists of predicative nouns in European Portuguese (Vaza, 1988; Baptista, 1997) and from the PLN.BR Full corpus (Bruckschein et al., 2008). This corpus contains 103,080 texts, with 29 million tokens, consisting of news pieces from *Folha de São Paulo*, a Brazilian newspaper (from 1994 to 2005). All these constructions were validated in real data, and in some cases also ressourcing to the web.

---

[1] In the examples, elements between brackets and separated by the '+' sign can all appear in that given syntactic slot. The symbol '#' delimits clauses, while the '*' mark signals the sentence as unacceptable. Correferent elements are linked by correference indexes $_i$. For clarity, all support verbs will be shown without italics in the examples. An approximate translation of Portuguese examples is provided, but its acceptability is irrelevant for the paper.

2.1 Nominal predicates with support verb *ter* "have"

We adopted several criteria that allowed us to constitute lexical-syntactic, relatively homogeneous, classes. These criteria were inspired in those taken from previous classifications, developed in the Lexicon-grammar framework of Maurice Gross (1975; 1988; 1996), for both Portuguese and other languages. The main classification criteria can be summarized as follows: (i) the number of arguments, considering constructions with a subject and one or two essential complements as arguments; (ii) the possibility of a noun admitting a sentential construction (in subject or complement position); (iii) the distributional nature of the arguments: if they are obligatorily human or allow for non-human nouns; (iv) the property of *symmetry* [2] between the arguments.

Following these criteria, we have so far classified around 1,000 nominal constructions from a list with 3,000 candidates of predicative nouns censed in the corpus (Santos-Turati, 2012). The already classified nominal predicates that select the support verb *ter* "have" in Brazilian Portuguese were divided into 9 classes (Table 1) [3].

Class	Structure	Example/Gloss	Count
TH1	*Nhum$_0$ ter Npred*	*Ana* tem *uma beleza impressionante* "Ana has an amazing beauty"	465
TNH1	*N-hum$_0$ ter Npred*	*A tinta* tem *um tom escuro* "The paint has a dark tone"	138
TR1	*N±hum$_0$ ter Npred*	*(Ana + a música)* tem *um ritmo contagiante* "(Ana + the music) has a contagious rhythm"	139
TH2	*Nhum$_0$ ter Npred Prep Nhum$_1$*	*Ana* tem *respeito por Max* "Ana has respect for Max"	111
TNH2	*N-hum$_0$ ter Npred Prep Nhum$_1$*	*O bombom* tem *gosto de avelã* "The bonbon has taste like hazelnut"	6
TR2	*N±hum$_0$ ter Npred Prep N-hum$_1$*	*(O carro + a cidade)* tem *um alto consumo de água* "(The car + the city) has a high consumption of water"	22
TS2	*Nhum$_0$ ter Npred Prep Nhum$_1$* (Simetry)	*O patrão* tem *um acordo com o empregado* "The boss has an agreement with the employee"	38
TQF1	*QueF$_0$ ter Npred Prep N$_1$*	*Esse fato* tem *uma grande importância para Ana* "This fact has a great importance for Ana"	6
TQF2	*N$_0$ ter Npred Prep QueF$_1$*	*Ana* tem *medo de dirigir na estrada* "Ana has fear to drive on the road"	80
TOTAL			1,005

Table 1: SVC with support verb *ter* (Santos-Turati, 2012)

2.2 Nominal predicates with support verb *dar* "give"

The same criteria were also adopted for SVC with verb *dar* "give" (Rassi and Vale, 2013), though two differences were considered: (i) the constructions with a body-part noun (*Npc*) as argument were distinguished as a special class for their particular properties; and (ii), no symmetric constructions were found. We classified 900 support verb constructions with verb *dar* "give" in Brazilian Portuguese into 11 classes (Table 2).

3 Relations between *ter* "have" and *dar* "give"

First of all, it is necessary to distinguish three different kinds of relations established between verb *dar* and verb *ter* constructions. The first type of relation considers the verbs *dar* "give" and *ter* "have" as synonymous and classified as standard support verb constructions. The verb *dar* can replace the verb *ter* without any changes in the meaning of the sentence or in the selection restrictions of the arguments:

---

[2]The notion of symmetry in verbal constructions was initially presented by Borillo (1971) for French verbs - *Paul rencontre son frère* "Paul meets his brother" / *Paul et son frère se rencontrent* "Paul and his brother meet". In the case of the Portuguese nominal constructions, symmetry was presented in Ranchhod (1990) and Baptista (2005), who described the nominal predicates with the support verbs *estar com* and *ser de*, respectively.

[3]In Table 1 and Table 2, the left column shows the conventional codes for designating each class; and the second column represents its syntactic structure, indicated as follows: *Nhum* and *N-hum* for human and non-human noun respectively; *N±hum* for both human or non-human noun; *Npc* for body-part noun; the indexes '$_0$' and '$_1$' indicate the subject and the complement position, respectively; *Npred* stand for the predicative noun; *Prep* for preposition; *QueF* for completive.

Class	Structure	Example/Gloss	Count
DH1	$Nhum_0$ *dar Npred*	*Ana* deu *uma pirueta* "Ana gave a pirouette"	133
DNH1	$N\text{-}hum_0$ *dar Npred*	*O balão* deu *um estouro* "The baloon gave a burst"	20
DR1	$N\pm hum_0$ *dar Npred*	*(Max + O clima)* deu *uma refrescada* "(Max +The weather) gave a refreshed"	51
DH2	$Nhum_0$ *dar Npred Prep $Nhum_1$*	*Max* deu *um castigo para a Ana* "Max gave a punishment to Ana"	217
DNH2	*Nhum dar Npred Prep $N\text{-}hum_1$*	*Ana* deu *uma cozida nos legumes* "Ana gave a cooked in the vegetables"	137
DPC2	$Nhum_0$ *dar Npred Prep Npc_1*	*Max* deu *um tapa na cara da Ana* "Max gave a slap in Ana's face"	114
DQF2	$Nhum_0$ *dar Npred Prep $QueF_1$*	*Max* deu *um jeito de consertar o carro* "Max gave a way to fix the car"	52
DHR2	$Nhum_0$ *dar Npred Prep $N\pm hum_1$*	*Ana* deu *destaque ao (Max + problema)* "Ana gave emphasis to (Max + the problem)"	60
DRH2	$N\pm hum_0$ *dar Npred Prep $Nhum_1$*	*(Ana + O telhado)* deu *proteção ao Max* "(Ana + The roof) gave protection to Max"	32
DR2	$N\pm hum_0$ *dar Npred Prep $N\text{-}hum_1$*	*(Ana+A lei)* deu *embasamento à teoria* "(Ana+The law) gave basis to the theory"	25
D3	N_0 *dar Npred Prep N_1 Prep N_2*	*Ana* deu *um apelido de macaco ao Max* "Ana gave the nickname monkey to Max"	59
TOTAL			900

Table 2: SVC with support verb *dar* (Rassi and Vale, 2013)

(5) *Ana* (deu + teve) *um + um*(*a*) (*birra + chique + pirepaque + tremelique + troço*).
  "Ana (gave + had) (a + an) (tantrum + hissy fit + outburst + shiver + thing)."

The second type of relation concerns the transformation named *Conversion* by G. Gross (1982; 1989), in which the predicative noun is maintained and their arguments change their relative position, without, however, changing their semantic roles. In these constructions, the sentence with AGENT subject is called the *standard* construction, while its equivalent sentence with the reversed argument order is called the *converse* construction. Usually, the support verbs of the standard and the converse construction are different, as it is also the preposition introducing the converse complement:

(6) *Ana deu algum apoio ao Max.* "Ana gave some support to Max."
  [Conv.] = *Max teve algum apoio da Ana.* "Max had some support from Ana."

The third kind of relation linking the sentences with the verb *ter* and the verb *dar* is the causative operator construction (already mentioned in §1):

(7) *Isso* deu # *Ana* tem *coragem.* "This gave # Ana has courage."
  = *Isso* deu *coragem à Ana.* "This gave courage to Ana."

These three types of relations are presented in the table below, with an example and the respective number of constructions in each type. From the intersection between the list of predicative nouns constructed with verb *ter* "have" and those with verb *dar* "give", we found 693 predicative nouns, distributed as shown in Table 3.

dar "give"	*ter* "have"	Example/Gloss	Count
SVCstandard	SVCstandard	*Ana* deu *um chique* "Ana gave a hissy fit" *Ana* teve *um chique* "Ana had a hissy fit"	35
SVCstandard	SVCconverse	*O policial* deu *uma multa ao Max* "The officer gave Max a fine" *Max* teve *uma multa* "Max had a fine"	72
VopCausative (VopC)	SVCstandard	*A flor* deu *alergia a Ana* "The flower gave allergy to Ana" *Ana* tem *alergia à flor* "Ana has an allergy"	586

Table 3: Comparative table with syntactic relations

3.1 Verbs *dar* and *ter* in standard SVC

Around 4.8% of the predicative nouns (35 constructions) accept both support verbs *dar* and *ter* in standard constructions, such as:

(8) *A empresa* (dá + tem) *atenção ao cliente.* "The company (gives + has) attention to the client."

(9) *O remédio* (dá + tem) *um efeito positivo no organismo.* "The medicine (gives + has) a positive effect on human body."

(10) *O resultado* (deu + teve) *um impacto significativo para o time.* "The result (gave + had) a significant impact to the team."

In Brazilian Portuguese, around 35 predicative nouns, such as *febre* "fever" and *dengue* "dengue", besides having both *dar* and *ter* as their support verb also allow *dar* as a causative operator on them (examples taken from the web):

[VopC]: *[Sua lição de casa:] água parada* dá *dengue.* "[...] still water gives (= causes) dengue."
[CVS_dar]: *Inclusive, a vizinha também* deu *dengue.* "Inclusive, the neighbour gave (= had) dengue."
[CVS_ter]: *O meu esposo já* teve *dengue.* "My husband already had dengue."

A few nouns (around 10), such as *amor* "love", *confiança* "trust" and *respeito* "respect", besides admitting the two support verbs in their basic construction, also admit *ter* in a converse construction:

(11) *O filho* dá *respeito à mãe.* "The son gives respect to the mother."
= *O filho* tem *respeito pela mãe.* "The son has respect for the mother."
[conv.] = *A mãe* tem *o respeito do filho.* "The mother has respect from her son."

3.2 Verb *dar* as standard SVC and *ter* as converse SVC

Around 10.4% of the predicative nouns (72 constructions) admit the verb *dar* in the standard construction and the verb *ter* in a converse construction, but not *ter* as a standard support. In Brazilian Portuguese, predicative nouns constructed with the support verb *dar* in a standard construction accept other converse verbs beyond the verb *ter* "have", such as *receber* "receive", *ganhar* "gain", *levar* "get" and *tomar* "take"[4].

(12) *Ana* deu *proteção ao Max.* "Ana gave protection to Max."
= *Max* (teve + recebeu) *a proteção da Ana.* "Max (had + received) the protection from Ana."

(13) *Ana* deu *uma ajuda ao Max.* "Ana gave a help to Max."
= *Max* (teve + ganhou) *uma ajuda da Ana.* "Max (had + gained) a help from Ana."

(14) *Ana* deu *uma resposta no Max.* "Ana gave an answer to Max."
= *Max* (teve + levou) *uma resposta da Ana.* "Max (had + got) an answer from Ana."

(15) *O policial* deu *uma multa ao Max.* "The officer gave a fine to Max."
= *Max* (teve + tomou) *uma multa do policial.* "Max (had + took) a fine from the officer."

3.3 Verb *dar* as *VopC* and *ter* as SVC

Around 84.8% (586 predicative nouns) of the elementary constructions with the support verb *ter* "have" also allow the causative operator verb *dar* "give"; some of these nouns constitute relatively homogenous semantic sets, *e.g.* the predicative nouns that express <feeling>, <sensation>, <emotion> or those that indicate <disease> (this semantic classification is just approximative):

---

[4]For European Portuguese equivalent converse constructions, see Baptista (1997); for a comparison between the two language variants, see Rassi *et al.* (2014).

(16) *Ana* tem *alegria*. "Ana has happiness."
   (*Zé* + *A vinda do Zé* + *O fato de Zé ter voltado* + *Isso*) deu *alegria à Ana*.
   "(Zé + Zé's coming + The fact of Zé has came + That) gave happiness to Ana."

(17) *Ana* tem *cólica*. "Ana has colic."
   (*O chocolate* + *O fato de ter comido chocolate* + *Isso*) deu *cólica na Ana*.
   "(The chocolat + The fact of she has eaten chocolat + That) gave a colic in Ana."

These predicative nouns allow a particular (impersonal?) construction with *dar*, in which the argument in subject position is not explicit, so the CAUSE element is also absent, and the sentence has the same overall meaning of the SVC with verb *ter* standard, but with an inchoative aspect; notice that the verb *dar* must be in the 3rd person singular, and it does not agree with the predicative noun:

(18) (Deu + *Deram) (*uma*) (*alegria* + *cólica*) *na Ana*. "Gives/gave (a) (hapiness + colic) in Ana."
   = *Ana* teve (*uma*) (*alegria* + *cólica*). "Ana had (a) (hapiness + colic)."

(19) (Deu + ?*Deram) umas (*palpitações* + *cólicas*) *na Ana*. "Gives/gave some (palpitations + colics) in Ana."
   = *Ana* teve *umas* (*palpitações* + *cólicas*). "Ana had some (palpitations + colics)."

3.4 Formalization into the Lexicon-Grammar

Because of the complex relations and the different syntactic status that the verbs *dar* and *ter* may show, these constructions are essentially determined by the lexicon, *i.e.*, they depend on the specific predicative noun. It is only natural that a lexically-based approach be taken in order to describe this properties, particularly in view of the implementation of such type of expressions in NLP systems. The Lexicon-Grammar framework constitutes such a methodological setting, as it presupposes the extensive and systematical survey and formal representation of the lexicon properties.

In the Lexicon-Grammar, a systematic description of linguistic phenomena is usually presented in the form of binary matrices: the lines contain the lexical entries while the columns represent syntactic-semantic properties of each entry. For example, for each predicative noun, distributional constraints on the arguments are represented; the elementary support verb and the main variants of this verb are encoded; the possibility of accepting conversion and the converse support verbs are explicitly provided; and all these syntactic-semantic informations are specified for each predicative noun. Besides its intrinsic linguistic interest, the main purpose for this formalization requirements is the application of the data in NLP. In the next section, we present a preliminary proposal for the implementation problems of these type of SVC in a rule-based parsing system of Portuguese.

4 Towards the implementation of SVC in a NLP system

Besides its linguistic interest, one of the goals of the formal representation of the lexical properties of predicative nouns and SVC into a Lexicon-Grammar such as described above (§3.4) is to allow for the implementation of these data in NLP systems. In this section an outline of the strategy adopted for its implementation specifically into a rule-based system, namely STRING (Mamede et al., 2012) [5], is presented. This is still an on-going work, so in the next lines we briefly sketch the system's architecture (§4.1.) and then (§4.2.) we present the strategy that we intend to implement for the adequate parsing of SVC with *ter* and *dar*, having in mind the complex structures and relations mentioned in §3.

4.1 STRING architecture

STRING is an NLP chain with a modular structure that executes all the basic processing tasks, namely: tokenization and text segmentation, part-of-speeh tagging, morphosyntactic disambiguation, shallow parsing (chuking) and deep parsing (dependency extraction). The parsing stage is performed by the rule-based parser XIP (Xerox Incremental Parser) (Mokhtar et al., 2002). XIP identifies the elementary

---

[5]http://string.l2f.inesc-id.pt/

constituents of a sentence, such as noun phrases (`NP`) or prepositional phrases (`PP`), and then these are structured by binary dependencies between them, corresponding to the syntactic relations, such as subject (`SUBJ`), direct complement (`CDIR`) or modifier (`MOD`). STRING also extracts Named Entities, performs time expressions identification and normalization, Anaphora Resolution and some Word-Sense disambiguation (WSD).

At the final stages of parsing, the system extracts the text events (or predicates) and their participants (arguments). The system currently extracts the `EVENT` structure for all full verbs and predicative nouns. In the case of verbs, it associates the events to their participants and circumstances, identifying their corresponding semantic roles (Talhadas, 2014), based on the sentence parse and the information available on the Portuguese full verbs Lexicon-Grammar (Baptista, 2012) [6]. Hence, for a sentence such as (20), the system parser extracts the event structure by way of the following dependencies:

(20) *Max costuma ler o jornal no café às sextas-feiras.* "Max uses to read the newspaper at the caffée on Fridays."

```
EVENT_AGENT(ler,Max)
EVENT_OBJECT(ler,jornal)
EVENT_LOC-PLACE(ler,café)
EVENT_TIME-FREQUENCY(ler,a_as sextas-feiras)
```

4.2 Strategy

In the case of a predicative noun in a SVC, one would want the predicative noun also to be captured as an `EVENT`, but *not* the support verb, since its role is basically that of an auxiliary of the noun. However, since the support verb conveys several important grammatical information, particularly the standard/converse orientation of the predicate[7], a `SUPPORT` dependency is first extracted, so in sentences such as in (21) one would get the dependency shown below:

(21) *Max* deu *um beijo na Ana.* "Max gave a kiss in Ana."

```
SUPPORT_STANDARD(beijo,deu)
```

To do so, one needs to provide the system with the information that *dar* is the (basic) standard support verb of the predicative noun *beijo* "kiss". It is also necessary to know that in this construction, the predicative noun is the direct complement (`CDIR`) of the support verb and that the dative complement can be introduced, in Brazilian Portuguese, by preposition *em* "in/on". The following rules illustrate (in a simplified way[8]) the functioning of the rule-based system:

```
if (CDIR(#1[lemma:dar],#2[lemma:beijo]) & ~SUPPORT(#2,#?))
    SUPPORT[vsup-standard](#1,#2)

if (SUPPORT(#1,?))
    EVENT[OTHER=+](#1).

if (SUPPORT[vsup-standard](#1[lemma:beijo],#2) &
    EVENT[other](#1) & SUBJ(#2,#3))
    EVENT[agent-generic=+](#1,#3).

if (SUPPORT[vsup-standard](#1[lemma:beijo],#2) & EVENT[other](#1) &
    ^MOD(#2,#3) & PREPD(#3,?[lemma:em]) )
    COMPL(#1,#3),
    EVENT[patient=+](#1,#3).
```

---

[6] This semantic role information is still not available for the predicative nouns, but it is currently being encoded.

[7] The support verb can convey aspectual, modal and even stylistic values, which are encoded in the lexicon and remain available in the system's output, even if not necessarily visible in the `EVENT` representation.

[8] The rule system should also take into account the distributional constraints on the argument slots, but, for simplicity, we dismissed it in this paper.

```
if (SUPPORT[vsup-standard](#1[lemma:beijo],#2) & EVENT[other](#1) &
   (^MOD[dat](#2,#3) || ^CLITIC(#2,#3[dat]) ) )
   CINDIR[dat=~](#1,#3),
   EVENT[patient=+](#1,#3).
```

The rules read as follows: First, a SUPPORT dependency with the feature _VSUP-STANDARD is extracted when the noun *beijo* "kiss" is the direct complement of the verb *dar* "give" (and no other support verb was extracted yet for that noun); based on this dependency, an EVENT (unary) dependency is extracted for the predicative noun; then, the subject of the standard support verb is assigned the `agent` semantic role (`agent-generic` in STRING's terminology); next, the prepositional phrase modifying (MOD) the support verb and introduced by preposition *em* "in" is converted into a complement (COMPL) of the predicative noun and assigned a semantic role of `patient`; a similar procedure is used for the dative complement, when reduced to a dative pronominal form, but in this case, instead of COMPL the CINDIR (indirect complement) dependency is used. All these rules are automatically produced for each predicative noun, from the information in the Lexicon-Grammar. The corresponding EVENT structure is represented below:

```
SUPPORT_VSUP-STANDARD(beijo,deu)
EVENT_AGENT(beijo,Max)
EVENT_PATIENT(beijo,Ana)
```

For the converse construction 4.2, while the EVENT structure remains the same, the SUPPORT dependency is:

(22) *Ana* ganhou *um beijo do Max.* "Ana got a kiss from Max."

```
SUPPORT_VSUP-CONVERSE(beijo,ganhou)
```

The converse construction entails the "swapping" of the arguments' syntactic function, while keeping their respective semantic roles. The detection of the converse construction triggers a set of rules that also swap the semantic roles associated to the predicative noun's syntactic slots. In the case where the same verb is both the standard and the converse support of a predicative noun, they are both extracted, at first, and then the presence of prepositional complements or the determiner of the noun can be used for disambiguation. This will be part of future work as, for the moment, whenever this happens, the converse construction is discarded. The assigning of semantic roles to the predicative noun's arguments is then made only once, and by general rules, both in the standard and in the converse constructions.

The situation is somewhat similar in the case of a causative-operator verb 4.2:

(23) *Essa notícia* deu *estresse no Max.* "This news gave stress in Max."

In this case, since the Lexicon-Grammar has encoded that the verb *dar* can be an operator on *ter*, and since the predicative noun *estresse* "stress" does not allow for *dar* to be its support, a general rule can apply, extracting the CAUSE relation expressed by the *VopC*, in a similar way as for the SUPPORT dependency. The EVENT structure is thus construed as shown below:

```
VOPC(estresse,deu)
EVENT(estresse,other)
EVENT_EXPERIENCER(estresse,Max)
EVENT_CAUSE(estresse,notícia)
```

However, when the same verb can be both a support and an operator verb, in the absence of tell-tale prepositional complements or other syntactic evidence, the detection of the adequate structure can not be done at this stage. We found only 35 predicative nouns which can be associated to the verb *dar* "give" with both categories, *i.e.* as a support and a *VopC*. It is also possible that both dependencies SUPPORT and CAUSE be extracted in order to disambiguate them at a later stage.

5 Conclusions and future work

In the near future, we intend to use the data encoded in the Lexicon-Grammar of these predicative nouns and build a SVC identification module for STRING. For the moment, the identification of all the syntactic phenomena, constituting as many different parsing cases as possible, is underway, in order to fully automatize the processing of converting the Lexicon-Grammar tables into the STRING, with XIP-compliant rules, in a similar way as it has already been done for the verbs (Baptista, 2012; Travanca, 2013; Talhadas, 2014). After implementing all the data in STRING we also intend to evaluate the system in order to check the extraction of the dependencies involving the support verbs and predicative nouns.

An important task ahead is the systematic comparison of the structures and properties here described against those of European Portuguese. First of all, the set of nouns available in each variant is not exactly the same, even if the concepts are shared; for example, *carona* in BP corresponds to the EP *boleia* "ride"; in other cases, the choice of the nominalization suffixes differ: in BP one uses the term *parada cardíaca*, while its equivalent in EP is *paragem cardíaca* "cardiac arrest". False-friends are also common: in BP, *chamada* "rebuke" is unrelated to EP *chamada* "phone call" (but, in this sense, it is also used in BP); the set of support verbs for each noun are different: as a synonym of *rebuke* we find the pair *dar-levar* (only in BP), while as equivalent to *phone call* the basic support verbs are *fazer-receber* (the same in BP and EP). Naturally, much in both variants is quite similar, though some patterns begin to emerge: the different choice of prepositions for the complement, mostly the alternation between *em* "in" in BP and *a* "to" in EP (both as dative complements); the choice of support verbs, with some being used for these predicative noun exclusively in BP (*ganhar* "gain" and *tomar* "take") or in EP (*pregar* "throw" and *apanhar* "take").

References

Jorge Baptista. 1997. *Sermão, tareia e facada*. Uma classificação das construções conversas *dar-levar*. In *Seminários de Linguística*, volume 1, pages 5–37, Faro. Universidade do Algarve.

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

Jorge Baptista. 2012. A Lexicon-grammar of European Portuguese Verbs. In Jan Radimsky, editor, *Proceedings of the 31st International Conference on Lexis and Grammar*, volume 31, pages 10–16, Czech Republic, September. Università degli Studi di Salerno and University of South Bohemia in Nové Hrady.

Andrée Borillo. 1971. Remarques sur les verbes symétriques. *Langue Française*, (11):17–31.

Mírian Bruckschein, Fernando Muniz, José Guilherme Camargo Souza, Juliana Thiesen Fuchs, Kleber Infante, Marcelo Muniz, Patrícia Nunes Gonçalvez, Renata Vieira, and Sandra Maria Aluisio. 2008. Anotação linguística em XML do corpus PLN-BR. Série de relatórios do NILC, NILC- ICMC - USP.

Magali Sanches Duran, Carlos Ramisch, Sandra Maria Aluísio, and Aline Villavicencio. 2011. Identifying and analyzing Brazilian Portuguese complex predicates. In *Proceedings of the Workshop on Multiword Expressions: from Parsing and Generation to the Real World*, pages 74–82, Portland, USA.

Gregory Grefenstette and Simone Teufel. 1995. Corpus-based Method for Automatic Identification of Support Verbs for Nominalizations. *CoRR*, cmp-lg/9503010.

Maurice Gross. 1975. *Méthodes en syntaxe*. Hermann, Paris.

Maurice Gross. 1981. Les bases empiriques de la notion de prédicat sémantique. *Langages*, 63(3):7–52.

Gaston Gross. 1982. Un cas des constructions inverses: *donner* et *recevoir*. *Lingvisticae Investigationes*, 2:1–44.

Maurice Gross. 1988. Methods and tactics in the construction of a Lexicon-grammar. In The linguistic Society of Korea, editor, *Linguistics in the Morning Calm 2. Selected papers from SICOL-1986*, pages 177–197, Seoul. Hanshin Publishing Company.

Gaston Gross. 1989. *Les constructions converses du français*. Droz, Genebra.

Maurice Gross. 1996. Lexicon grammar. In K. Brown and J. Miller, editors, *Concise Encyclopedia of Syntactic Theories*, pages 244–259. Pergamon, Cambridge.

Maurice Gross. 1998. La fonction sémantique des verbes supports. In Béatrice Lamiroy, editor, *Travaux de Linguistique*, number 37, pages 25–46.

Patrick Hanks, Anne Urbschat, and Elke Gehweiler. 2006. German light verb constructions in corpora and dictionaries. *International Journal of Lexicography*, 19(4):439–457.

Zellig Harris. 1964. The Elementary Transformations. *Transformations and Discourse Analysis Papers*, (54):211–235.

Zellig Harris. 1991. *A Theory of Language and Information: A Mathematical Approach*. Clarendon Press, New York.

Iris Hendrickx, Amália Mendes, Sílvia Pereira, Anabela Gonçalves, and Inês Duarte. 2010. Complex predicates annotation in a corpus of Portuguese. in: Proceedings of the 4th ACL. In *Proceedings of the 4th ACL Linguistic Annotation Workshop*, pages 100–108, Uppsala, Sweden.

Béatrice Lamiroy. 1998. Le Lexique-grammaire: Essai de synthèse. In Béatrice Lamiroy, editor, *Travaux de Linguistique*, volume 37, pages 7–23.

Nuno Mamede, Jorge Baptista, Vera Cabarrão, and Cláudio Diniz. 2012. STRING: An hybrid statistical and rule-based natural language processing chain for Portuguese. In *International Conference on Computational Processing of Portuguese (Propor 2012)*, volume Demo Session, Coimbra, Portugal, April.

Salah Ait Mokhtar, Jean-Pierre Chanod, and Claude Roux. 2002. Robustness beyond shalowness: Incremental dependency parsing. *Natural Language Engineering*, pages 121–144.

Elisabete Ranchhod. 1990. *Sintaxe dos predicados nominais com Estar*. INIC - Instituto Nacional de Investigação Científica, Lisboa.

Amanda Rassi and Oto Vale. 2013. Predicative Nouns Suffixation associated to the verb *dar* (give) in Brazilian Portuguese. In Jorge Baptista and Mario Monteleone, editors, *Proceedings of the 32nd International Conference on Lexis and Grammar*, volume 32, pages 151–158, Faro, September. UAlg.

Amanda Rassi, Nathalia Perussi, Jorge Baptista, and Oto Vale. 2014. Estudo contrastivo sobre as construções conversas em PB e PE. In Cristina Fargetti, Odair Silva, Clotilde Murakawa, and Anise Ferreira, editors, *Anais do I CINELI - Congresso Internacional Estudos do Léxico e suas Interfaces*, volume 1, page (no prelo), Araraquara-SP, Maio. Universidade Estadual Paulista - UNESP.

Maria Cristina Andrade Santos-Turati. 2012. Descrição da estrutura argumental dos predicados nominais com o verbo-suporte *ter*. In *Seminário do GEL*, number 60, pages 20–21, São Paulo, Brasil. Grupo de Estudos Linguísticos do Estado de São Paulo - GEL.

Angelika Storrer. 2007. Corpus-based investigations on german support verb constructions. In Christiane Fellbaum, editor, *Collocations and Idioms: Linguistic, Lexicographic, and Computational Aspects*, pages 164–188. Continuum Press, London.

Rui Talhadas. 2014. Automatic Semantic Role Labeling for European Portuguese. Master's thesis, Universidade do Algarve, Faro.

Tiago Travanca. 2013. Verb Sense Disambiguation. Master's thesis, Instituto Superior Técnico, Universidade de Lisboa, June.

Aldina Vaza. 1988. Estruturas com nomes predicativos e o verbo-suporte *dar*. Master's thesis, Faculdade de Letras - Universidade de Lisboa.