Support Verb Constructions across the Ocean Sea

Jorge Baptista[1,2], Nuno Mamede[2,3], Sónia Reis[1,2]
[1]University of Algarve, [2]INESC-ID, Lisbon, [3]Instituto Superior Técnico, University of Lisbon
[1]Campus de Gambelas, 8005-139 Faro, [2]R. Alves Redol 9, 1000-029 Lisboa,
[3]Av. Rovisco Pais 1, 1049-001 Lisboa
jbaptis@ualg.pt, nuno.mamede@tecnico.ulisboa.pt, smreis@ualg.pt

Abstract

This paper analyses the *support* (or *light*) *verb constructions* (SVC) in a publicly available, manually annotated corpus of multiword expressions (MWE) in Brazilian Portuguese. The paper highlights several issues in the linguistic definitions therein adopted for these types of MWE, and reports the results from applying STRING, a rule-based parsing system, originally developed for European Portuguese, to this corpus from Brazilian Portuguese. The goal is two-fold: to improve the linguistic definition of SVC in the annotation task, as well as to gauge the major difficulties found when transposing linguistic resources between these two varieties of the same language.

Keywords: support-verb constructions, light verbs, predicate noun

1. Introduction

Support-verb (or *light verb*) constructions (SVC) (Gross, 1981; Gross, 1996; Gross, 1998). are a fundamental component of the lexicon and grammar of any language (Constant et al., 2017), conveying a large variety of semantic predicates, in as much the same way as full (or distributional) verbs, predicative adjectives and other predicative elements do. In broad traits, a SVC can be defined as a multiword expression (MWE) that consists of an *elementary* (or *base*) sentence (Gross, 1981) where the predicative nucleus is formed by a *predicate noun* (Npred), which conveys the lexical meaning of the expression, and a *support-verb* (Vsup), an auxiliary element that serves basically to "conjugate" the predicative noun (Gross, 1989, p.38), mostly conveying grammatical values – person-number and tense, but also aspect and modality, and eventually, some stylistic values (Gross, 1998), that the morphology of the predicate noun cannot by itself express. (In this paper, we do not distinguish the terms *light* or *support* verbs, following, among others, (Fotopoulou et al., 2021), who consider that the way authors use them is not consistently correlated with differences between the properties of the constructions.)

Clear examples of SVC, in Portuguese, are: *O Pedro tem fome* lit:'Pedro has hunger' 'Pedro is hungry' (Santos, 2015), *O Pedro deu um abraço ao João* 'Pedro gave a hug to João' (Baptista, 1997b; Calcia, 2022), *O Pedro fez/está em greve* 'Pedro is on strike' (Chacoto, 2005; Dias de Barros, 2014), *É do interesse do Pedro que o João faça isso* 'It is in Pedro's interest that João do this' (Baptista, 2005b).

A key aspect of SVC is that the most relevant of their syntactic-semantic properties result from each verb-noun combination and, though some regularities can be found across large subsets of the SVC *lexicon-grammar* (in the sense of (Gross, 1996)), those properties cannot (and, in our view, should not) be generalized over the lexicon, neither of predicate nouns, nor of the verbs that can function as support-verbs. Definitory formal properties have been discovered, particularly since the (Giry-Schneider, 1978), that allow for a clear distinction between SVC and other, formally identical, constructions with full (or distributional) verbs (see (Ranchhod, 1990; Baptista, 2005b) for an overview). For example, since the predicative noun expresses a semantic predicate, it selects at least one other element for its subject argument (Gross, 1981). In the examples above, this is the relation holding between the predicate nouns and the subject of the SVC, which precludes the possibility of inserting a complement *de N* 'of N' (or a possessive pronoun) modifying the predicate noun that is not correferent to the subject: **O Pedro tem a fome do Rui/a tua fome* lit:'Pedro has Rui's/your hunger', **O Pedro deu um abraço do Rui ao João* lit:'Pedro gave João Rui's hug', **O Pedro está em/fez a greve do Rui* 'Pedro is on Rui's strike', **É do teu interesse do Pedro que o João faça isso* 'It is in Pedro's your interest that João do this'. (Some of these sentences can only be interpreted in the comparative sense of 'the same Npred that', or 'in place/instead of' hence, they are not elementary or base sentences.)

This paper analyses *support-verb constructions* represented in a publicly available, manually annotated corpus of verbal idioms in Brazilian Portuguese (PT-BR). The paper highlights several issues in the linguistic definitions therein adopted for the annotation of this type of MWE. It then reports the results from applying STRING (Mamede et al., 2012; Baptista and Mamede, 2020) [1] to this corpus from Brazilian Portuguese. STRING is a statistical and rule-based nat-

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

Proceedings of the 18th Workshop on Multiword Expressions (MWE 2022) n@LREC2022, pages 26–36

ural language processing pipeline, specifically developed for European Portuguese (PT-PT).

The goal of the paper therefore is two-fold: (i) to discuss several issues in the linguistic definition of SVC adopted in the annotation of the corpus, helping to contribute to the clarification of several key concepts; and (ii) to gauge the major difficulties found when applying linguistic resources originally built for PT-PT to a text written in PT-BR, shedding some light on the degree of linguistic similarity between the SVC of these two varieties of the same language.

The paper is structured as follows: Next, Section 2 presents related work on SVC, with a special focus on Portuguese, both European (PT-PT) and Brazilian (PT-BR); Section 3 describes the PARSEME corpus, used in this paper; Section 4, briefly presents the processing of the corpus in STRING and the experiments performed; Section 5 presents and discusses the results obtained; and, finally, Section 6 draws the main conclusions and refers to future work.

2. Related work

Though the idea of nouns as predicative elements in language is quite old in grammar and in language studies, a modern thread can be sourced on (Harris, 1955) and subsequent work (Harris, 1964; Harris, 1976; Harris, 1982; Harris, 1991), while the terms *support-verb* (*Vsup*) and *predicative noun* (*Npred*), and the corresponding concepts here adopted, have been coined by (Gross, 1981), and later extended in (Gross, 1998). Extensive/systematic descriptions of SVC have been produced, within the Lexicon-Grammar framework (Gross, 1996), both for romance and non-romance languages, mostly in the early 80s and in the 90s (and for Brazilian Portuguese mostly since the early 2010s); see (Fotopoulou et al., 2021) for a brief, tough non-exhaustive overview.

For European Portuguese (PT-PT), the language variety that is the focus of this paper, landmarks in this descriptive campaign started in the late 80s, with (Vaza, 1988; Ranchhod, 1990; Baptista, 1997b), and continue until the mid-2000s, (Baptista, 2005b; Chacoto, 2005). Specific constructions, such as Converse (i.e. passive-like) SVC, as originally defined by (Gross, 1989) received attention in multiple works (Vaza, 1988; Baptista, 1997a; Baptista, 1997b); the description of specific transformations, such as Fusion (Gross, 1981) and particular classes of predicate nouns, like instrument nouns (Baptista, 2004) and communication predicates (Reis et al., 2021); or in the context of the more general phenomenon of Symmetry (Baptista, 2005a), i.e. intrinsically reciprocal constructions, as originally defined by (Borillo, 1971).

For Brazilian Portuguese (PT-BR), the language variety of the corpus used in this paper, mention should be made to the SVC with support-verb *fazer* 'do/make' (Dias de Barros, 2014), *dar* 'give' (Rassi, 2015) and *ter* 'have' (Santos, 2015); and for specific aspects of SVC, like the Converse constructions involving the support-verb *dar* 'give' (Calcia, 2016; Calcia and Vale, 2019); the aspectual variants of support-verbs, (Picoli et al., 2021), and non-agentive constructions with *fazer* 'do/make' (Dias de Barros et al., 2013).

Few works have been dedicated to the systematic comparison of the lexicon and grammar of the PT-PT and PT-BR variants, exception made to (Rassi et al., 2016), who compared a subset of converse SVC. An annotated corpus of SVC with support-verb *dar* 'give' has also been produced (Rassi et al., 2015b).

Extensive literature exists on SVC across multiple languages, on their place within the description of multiword expressions (Constant et al., 2017), their relation with fixed, verbal idioms and the challenges they pose to Natural Language Processing (NLP) (Sag et al., 2002). A comprehensive set of references and the current trends in MWE processing can be found in (Ramisch et al., 2020, p.222–223) and in (Cook et al., 2021), among others. Processing SVC Portuguese has been the topic of, among others, (Baptista et al., 2015; Rassi et al., 2014; Rassi et al., 2015a), with recent developments in (Mota et al., 2018; Baptista and Mamede, 2020; Barreiro et al., 2022).

In recent years, the study of multiword expressions (MWE) received a significant boost by the collaborative efforts of a multilingual community gathered around the PARSEME project (Savary et al., 2015)[2], developed under the European Union COST framework. The PARSEME project is aimed at "characterizing MWEs in lexicons, grammars and corpora and enabling systems to process them" (Ramisch et al., 2020, p.107). Under this project, several initiatives were taken, including a Shared Task on automatic identification of MWE. For the Shared Task 1.2 (edition) (Ramisch et al., 2020), a (Brazilian) Portuguese corpus, has been manually annotated for verbal MWE. A major contribution of the project, within this Shared Task, was the construction of "unified guidelines for all the participating languages, in order to avoid heterogeneous, hence incomparable, datasets"[3]. These guidelines take the form of *decision trees*, with specific branches for each one of the two main verbal MWE categories addressed by the project: support-verb constructions and verbal idioms. SVC (or light verb constructions *LVC*, in the authors' terminology), are considered "universal, that is, valid for all languages participating in the task" (Ramisch et al., 2020, p. 224), though there are reasons to believe that they may pertain to many types of languages. Within PARSEME, SVC are organized in 2 subsets (Portuguese examples from the PARSEME training corpus; the succinct defini-

[2] https://typo.uni-konstanz.de/parseme/index.php [last access: June 13, 2022]. All URL in this paper were last checked on this date.

[3] The full guidelines for Shared Task Edition 1.1 can be found at: https://parsemefr.lis-lab.fr/parseme-st-guidelines/1.1/

tions below were also taken from (Ramisch et al., 2020, p. 224)): (a) **LVC.full**, "in which the verb is semantically totally bleached", e.g. *fazer uma palestra* 'make a speech'; (b) **LVC.cause**, "in which the verb adds a causative meaning to the noun", e.g. *dar origem a* (lit. 'give origin to', 'gives rise to'). The project's participant teams produced *corpora* manually annotated for MWE, and for several languages (+18), including (Brazilian) Portuguese (PT-BR). These *corpora* have been updated and extended throughout several editions of the Shared Task, and in this paper Portuguese data from the latest Shared Task 1.2. edition (2020) will be used.

3. SVC in the PARSEME corpus

The PARSEME Portuguese corpus[4] is divided into training (80%), development (10%) and testing (10%) partitions. In this paper, only the testing partition was considered (though the entire corpus has been processed by STRING).

According to the information distributed with the corpus, it consists of 27,904 sentences, 638,002 tokens, where 3,145 SVC (or *light verb* constructions, noted 'LVC.full', in the authors' terminology), and 94 LVC.cause (=causative constructions) have been manually annotated. The testing partition consists of 2,770 sentences, +62.6 thousand tokens, and, according to the documentation, it includes 337 LVC.full and 7 LVC.cause.

According (Ramisch et al., 2020, p.226), "the Portuguese corpus contains sentences from the informal Brazilian newspaper *Diário Gaúcho* and from the training set of the [Universal Dependencies] treebank" (UD_Portuguese-GSD v2.1). We could not find information on the method used to sample the sentences included in the corpus.

A sample of 1,000 sentences (4.54% of the corpus) is reported (idem, p.227) to have been double-annotated, and the following agreement metrics were reported: F_{span} (0.713) is the F-measure between annotators, K_{span} (0.684) is the agreement on the annotation span and K_{cat} (0.837) is the agreement on the VMWE category. These results seem to indicate the sufficiency of the content of the guidelines and the training of the annotators, yielding reasonable consistency of the annotation process, given the complexity of the task.

Lexical variety

Digging deeper into the corpus content, one can note that the distribution of some SVC constructions seems somewhat skewed. A large number come from text on sport (football), and not all support-verb/predicate noun combinations seem natural: 134 instances of *gol* 'goal' (*fazer* 'do' and *marcar* 'score', ?*dar* 'give'); 30 *falta* (*fazer*, *marcar*, *sofrer* 'suffer', ?*cometer* 'comit'); 13 *passe* (*fazer* 'make, do', *receber* 'receive'); 11 *pênalti* 'penalty' (*marcar*, *sofrer*); etc. Because of the lack of information on the sampling strategy, no further comment can be made on the fact that 6.8% of the SVC instances concern vocabulary from this specific domain.

As shown in the examples above, the lexical variety of SVC in the corpus looks sometimes skewed by the occurrence of several instances of the same predicate nouns with the same support-verb, without any obvious relevant change, neither in the meaning nor in the syntactic structure of the SVC, which would add value to their inclusion in the corpus. For example, a certain number of nouns designate measurable quantities, usually accompanied by a quantification phrase, e.g. *área* 'area' (x8), *população* 'population' (x2), where one can find some that are technical terms drawn from Astronomy: *ascenção* (*reta*) 'right ascension', *declinação* 'declination' (x2), *excentricidade* 'excentricity' (x6), *inclinação* 'inclination', e.g.

> *Possui uma excentricidade de 0.03574140 e uma inclinação de 11.03095º.* 'It has an eccentricity of 0.03574140 and a tilt of 11.03095º.'.

The purpose of including these astronomical terms in the corpus is not entirely clear. Still, it is interesting that other senses of these predicate nouns, such as *excentricidade* 'excentricity' *inclinação* 'inclination', as illustrated below, are absent from the data, e.g. *O Pedro é de uma certa excentricidade* 'Pedro is of a certain eccentricity' ('Pedro is eccentric') (Baptista, 2005b). On the other hand, many of these nouns designating measurable quantities can undergo a restructuring transformation (or alternation) (Baptista and Ranchhod, 1998), leaving them superficially as a complement of the measuring unit:

> *O mundo tem 510 bilhões de km2 de área total* 'The world has 510 billion km2 of total area' [5]
> = *O mundo tem uma área total de 510 bilhões de km2* 'The world has a total area of 510 billion km2'

however, none of these restructured forms seems to have been captured by the corpus.

Naturally, the fact that the corpus is produced out of real texts and it was built to be used in the training of machine-learning based systems perfectly justifies this aspect of the lexical distribution of SVC, even if the documentation is scarce on the sampling method used (if any) to select the sentences therein. In our more lexicographic-oriented and rule-based approach to the automatic identification of SVC in texts, lexical and syntactical diversity is a good feature to put the STRING system to the test.

[4] https://gitlab.com/parseme/parseme_corpus_pt

[5] https://www.ufjf.br/

Linking operator verb (*Vop*)
In some cases, and for theoretical reasons, we do not concur with the description of SVC given to some structures found in the corpus. For example, the concept of linking operator-verb (*Vopl*) construction, proposed by (Gross, 1981, p.30) seems to be ignored by the corpus annotators and it is dealt with as an ordinary SVC construction; for example (*Vopl* construction emphasized):

> *A Cátedra Milton Santos* tem como objetivo *a difusão de informações* 'The Milton Santos Chair *has as its aim* (=aims) to disseminate information(id=pt_br-ud-train-s7942);
> *o especial da TV Globo* terá como tema *a vida de Dolores Duran* 'the TV Globo special *will have as theme* the life of Dolores Duran' (id=diario_gaucho_16311).

These nouns (*objetivo* 'objective' and *tema* 'theme') have a clear support-verb (*Vsup*) construction, with a standard syntactic configuration, where the predicative noun is usually the direct complement of the support-verb e.g. (*Vsup* construction emphasized):

> *(A constituição de) a cátedra* tem *um* objetivo *preciso* '(The constitution of) the chair has a precise purpose;
> *O programa* tem *um* tema *interessante* 'The program has an interesting theme'

Furthermore, the predicative nature of the preposition phrase introduced by *como* 'as' (or, alternatively, by *por* lit.: 'by', 'idem') hints at the existence of the corresponding sentences with copula verbs (Paiva Raposo, 2013), e.g.

> *A difusão de informações* era *o objetivo da cátedra* 'Dissemination of information was the goal of the chair';
> *A vida de Dolores Duran* era *o tema do programa* 'Dolors Duran's life was the theme of the program';

A similar situation occurs with operator verb *ter* on adjectival/participial constructions or on SVC with *estar Prep*:

> *Já Federer* . . . teve *uma* campanha mais perturbada . . . 'Federer [=person] . . . had a more troubled/disrupted campaign'
> *Pelo segundo ano consecutivo, o Cruzeiro* teve *uma* campanha abaixo de_as expectativas. 'For the second year in a row, Cruzeiro [football club] has fallen short of expectations' (lit.: had a campaign below expectations')

These can hardly be thought of as elementary SVC sentences, for they yield to a transformational analysis that recovers the underlying elementary sentence under *ter* 'have': *A campanha foi perturbada / esteve abaixo das expectivas* 'The campaign was disrupted / was below expectations' On the other hand, a clear SVC construction of this predicate noun exists and it is patent in the corpus:

> *Os jovens . . . estão* fazendo a campanha *com a cara e a linguagem deles* 'Young people are making the campaign with their face and their language'

Finally, and again with noun *campanha* 'campaign', there are some cases where the notion of support-verb seems too much stretched:

> *. . . o partido foi vítima de uma intensa* campanha promovida *pela oposição de direita e seus aliados . . .* 'the party was the victim of an intense campaign promoted by the right-wing opposition and its allies'

In this case, the verb *promover* 'promote' has been analysed as a support-verb, probably because the (semantic) **agent** of *campanha* happens to coincide with the (syntactic) subject of the verb (*a oposição de direita e seus aliados* 'the right-wing opposition and its allies'). However, and in our perspective, the verb can not be considered a support-verb for its subject may not be correferent to the semantic agentive argument of the noun in its direct object position: *O Pedro promoveu a campanha do Rui* 'Pedro promoted Rui's campaign' (Notice, by the way, how the corpus ignored the expression *ser vítima de* 'be the victim of'. On SVC with copula verbs, see below).

Causative operator verb (*Vopc*)
Another case of operator verb, also originally described by (Gross, 1981), is the causative operator verb (Vopc), which has, nevertheless, been integrated into the PARSEME Guidelines [6] and represented by the category LVC.CAUSE. In the STRING system, this structures are lexically associated with the predicate noun construction and, when adequately captured, they are represented by a VOPCAUSE dependency (Baptista and Mamede, 2020). So, a simple matter of different notation might seem to be the case, here. Still, the annotation of this category seems inconsistent in the corpus. For example, the following instances were either marked with `LVC.cause`, or missed, or marked as `LVC.full` (=SVC):

> *A ausência do sexo também* traz *uma forte* angústia (marked with `LVC.cause`) 'The absence of sex also brings a strong anguish'
> *Nós . . . estamos ansiosos para montar um*

[6] `https://parsemefr.lis-lab.fr/parseme-st-guidelines/1.2/?page=050_Cross-lingual_tests/020_Light-verb_constructions__LB_LVC_RB_`

> *time competitivo, que seja divertido e* traga orgulho *para os fãs* (the *Vopc* was missed) 'We are looking forward to building a competitive team that is fun and brings pride to the fans'
> *Caro V., a tua postura é sempre admirável, o que* faz com que *tua lealdade a esta coluna só me* dê orgulho (marked as `LVC.full`) 'Dear V., your posture is always admirable, which makes me proud of your loyalty to this column'

Some of these *Vopc*, such as *fazer com que* 'make' 'lit.: make with that'), in the last example (missed), had already been mentioned in Portuguese literature (Baptista, 1999).

To conclude this topic, some instances of predicate noun constructions corresponding to the concept of causative operator-verb (*Vopc*) have been found to have been annotated as SVC. This distinction is not always made in the literature. For example, (de Athayde, 2000) largely ignores it and assimilates Vopc constructions to SVC with support-verb *fazer* 'do/make', while (Chacoto, 2005) keeps a clear-cut distinction (as we do). Besides, (Gross, 1998), who originally proposed the concept in (Gross, 1981), revised his initial position and proposed to treat Vopc as a special type of support-verb. In our view of the issue, we concur with the PARSEME classification criteria (see 6). The semantic added-value of **cause** introduced by Vopc is distinct from the function of Vsup (and the merely grammatical or stylistic values Vsup convey). Thus, this specific **cause** semantic value must somehow be captured for an adequate representation of the meaning relations among multiple elements within the sentence.

Secondly, the theoretical construct of linking operator-verb (*Vopl*), also introduced by (Gross, 1981), has been by and large been ignored in the corpus annotation. There is no clear indication in the literature on how to deal with this formal variation, though we posit that it may not be adequate to mix it together with standard SVC. Unlike Vopc, no new lexical element of meaning is introduced in the sentence, since the Vopl recuperates one of the arguments of the underlying predicate noun (and often it also structures this noun syntactical construction). It has been proposed (Ranchhod, 1990, p.183 ff.) that the use of Vopl be seen as a type of saliency-inducing device (in a similar as extraposition or clefting). In the Harrissian framework (Harris, 1991) that we adopted, this is a specific type of operation, in much the same way, but with a different meaning-inducing effect, as Vopc. Hence, in the same way as LVC.cause (within PARSEME) or the VOPCAUSE dependency (within STRING), a distinct notation may be required.

SVC with *ser* and *estar*

Also, it is interesting to notice that concerning SVC involving otherwise copula verbs *ser* and *estar* 'be', while they have been, in general, ignored (and eventually excluded from SVC) as per PARSEME Guidelines), they are still showcased in the corpus, where some (very few) of this SVC expressions can be found, both with support-verb *ser* 'be':

> <Title of a song> *é sucesso!* 'is [a] hit (lit.: success)' ... *um comercial da Tim era sucesso na tevê* 'a Tim commercial was a hit on TV'

and with support-verb *estar Prep* 'be Prep' (Ranchhod, 1990):

> *Você está com saudade do frio?* '' *Ela está com febre, dores no corpo e coriza* 'She has a fever, body aches and a runny nose' (id=diario_gaucho_16030); *Quando não está de licença* (id=pt_br-ud-train-s7822)

Standard/Converse SVC

Finally, even though appropriately signalling them as SVC, PARSEME does not distinguish between *standard* (or active-oriented) SVC, where the subject is the **agent** of the semantic predicate expressed by the predicative noun, from *converse* (or passive-oriented) SVC, where the subject is the **patient** (or the **object**) of that same semantic predicate; e.g. *O Pedro deu um abraço ao João = O João recebeu/levou um abraço do Pedro* 'Peter gave John a hug=John got a hug from Peter' This Conversion transformation (Gross, 1981; Gross, 1989) is very productive in many languages and has already received extensive linguistic descriptions in Portuguese, both in the European (Baptista, 1997a; Baptista, 1997b; Baptista, 2005b), and, moore recently, in the Brazilian variety (Calcia, 2016; Calcia and Vale, 2019; Calcia, 2022). These different types of SVC need to be both described and appropriately annotated in corpora, like STRING does, as they have an impact, among other aspects, on the determination of the semantic roles of the predicate noun's argument slots. We have counted many converse-like SVC structures (34) in the testing partition, which signals that this phenomenon is not rare in the corpus. Because of this lack of distinction, in the evaluation of the STRING's output, the standard/converse opposition was not taken into consideration.

4. Processing SVC in the PARSEME corpus

The test partition of the PARSEME corpus was processed using the STRING natural language processing pipeline (Mamede et al., 2012; Baptista and Mamede, 2020). This system performs all basic text processing tasks, including text segmentation into sentences, tokenization, dictionary-based part-of-speech (PoS) tagging, rule-based and statistical PoS disambiguation, and parsing. The parsing module is a rule-based parser, XIP (Aït-Mokhtar et al., 2002), that, among

other tasks, extracts dependency relations (such as SUBJ[ect], CDIR (direct object), etc.), between the basic constituents (chunks) heads.
Since most SVCs are formally identical to ordinary verbal constructions (the SVC status resulting from the specific verb-noun combination), the overall strategy adopted in STRING consists in, firstly, capturing the syntactic dependencies holding between the predicate noun and the verb, and then extracting a specific dependency SUPPORT_VSUP linking them. The system can be configured to output only the desired dependencies. Fig. 1 shows the result from parsing the sentence *A Ana marcou dois gols* 'Ana has scored two goals', where only some dependencies have been shown.

```
SUPPORT_VSUP-STANDARD(gols,marcou)
SUBJ_PRE(marcou,Ana)
CDIR_POST(marcou,gols)
0>TOP{NP{A Ana} VF{marcou} NP{dois gols}}
```

Figure 1: A sentence parsed by STRING

The -STANDARD suffix in the SUPPORT_VSUP dependency indicates that this is a *standard* (or active-oriented) SVC (in the case of a *converse* construction, a -CONVERSE suffix would be used instead). Fig. 2 shows the dependency rule used to extract the SUPPORT_VSUP dependency shown in the previous Figure.

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

Figure 2: Parsing rule for predicate noun *golo* 'goal'

Briefly, this rule matches the lemma of the verb *marcar* 'score' and checks whether there is a direct complement whose lemma is *golo* 'goal'. The rule also takes into consideration the situation where the predicate noun is the pivot of a relative clause, or the subject of a passive sentence. All these rules are automatically generated (Baptista and Mamede, 2020) from the database that encodes the linguistic (structural, syntactic, semantic, and transformational) properties of the predicative nouns' lexicon (which is therefore called a *lexicon-grammar*). In its current state, the lexicon-grammar of PT-PT SVC contains 5,800 entries (ambiguous predicate nouns, with multiple word senses, constitute several, separate entries), an ongoing description is being done for another 3,320 nouns. One of the purposes of this paper is to gauge the current lexical coverage of this linguistic resource and the system using it, on a previously unseen corpus of data, and, furthermore, on a corpus from a different variety of the language.

In order to allow for the semi-automatic comparison between the PARSEME SVC annotations and the STRING's output, a program was built in-house that detects the PARSEME SVC.full tag, and retrieves the two related elements, returning for that sentence the SUPPORT_VSUP dependency in the STRING's format. This is, in no way, a trivial task, since per PARSEME conventions, the SVC.full tag can be marked either on the line with the support-verb, or on the line with the predicative noun (for example, in the case of passive sentences). Also, as the text segmentation criteria is not exactly the same, some sentences in the PARSEME corpus were split by the STRING, and had to be manually adjusted to avoid mismatches.
The evaluation of the STRING's performance, thus, consists in the comparison between two parallel files aligned at the sentence level (as it is illustratyed in Fig.3-4). The sentence parsing output is the same in both figures, since it was performed by STRING. In sentence 1, we find the SVC *fazer uma aparição* 'make an appearance', while in sentence 2 *tomar uma providência* 'make a provision', which is in the passive.

```
SUPPORT_VSUP-STANDARD(aparição,fez)
1>TOP{PP{PP{Em 2} PP{de outubro} PP{de 2009}} , PP{
em o 10º aniversário} PP{de a SmackDown} , NP{The
Rock} VF{fez} NP{uma aparição} AP{especial} PP{em
um vídeo} AP{pré-gravado} .}

SUPPORT_VSUP-STANDARD(providência,tomada)
2>TOP{NP{Se} NP{nenhuma providência} VCOP{for} VCPART
{tomada} , NP{a população} VTEMP{vai} VASP{voltar a}
VINF{usar} NP{lamparinas} ADVP{ADV{a a noite}} e NP
{geladeira} PP{a querosene} .}
```

Figure 3: PARSEME corpus (Reference)

```
TP:SUPPORT_VSUP-STANDARD(aparição,fez)
1>TOP{PP{PP{Em 2} PP{de outubro} PP{de 2009}} , PP{
em o 10º aniversário} PP{de a SmackDown} , NP{The
Rock} VF{fez} NP{uma aparição} AP{especial} PP{em
um vídeo} AP{pré-gravado} .}

FN:SUPPORT_VSUP-STANDARD(providência,tomada)
2>TOP{NP{Se} NP{nenhuma providência} VCOP{for} VCPART
{tomada} , NP{a população} VTEMP{vai} VASP{voltar a}
VINF{usar} NP{lamparinas} ADVP{ADV{a a noite}} e NP
{geladeira} PP{a querosene} .}
```

Figure 4: STRING output

In the STRING's output 4, when an identical result was obtained, this was marked as a true-positive (TP). When no output was obtained, the dependency in reference was copied to the STRING's output file and marked as a false-negative (FN). There are also cases of sentences that were not marked as SVC, neither in the corpus annotation nor in the STRING's output. These are also considered FN and could only be detected by manual inspection. The manual analysis of the output also allowed for the correction of some cases as false-positives (FP) or as true-negative (TN). These cases will be presented and discussed in the next section.

5. Results and Discussion

Table 1 shows the results from the processing of the testing partition of the PARSEME corpus by STRING. This partition consists of 2,770 sentences (as segmented by STRING), where 311 instances of LVC.full had been manually annotated in the PARSEME corpus. The TP (true-positive cases) correspond to instances where the STRING adequately marked a SVC; the FP (false-positives) are instances where a SUPPORT dependency was incorrectly extracted; and FN (false-negatives) are instances of SVC ignored by the system.

Table 1: Results from comparing the annotations in the testing partition of PARSEME corpus and the annotation of the same corpus as performed by STRING.

TP	**FP**	**FN**
197	20	270
P	**R**	**F**
0.91	0.42	0.58

Considering the 2,770 phases of the corpus, the system's precision was high, though recall is low. From the 311 sentences marked in the PARSEME corpus, the STRING system correctly identified 154 as SVC, missed 149 and incorrectly marked 8. If one strictly considers the annotation in the PARSEME corpus as the reference (golden standard), these partial results from STRING correspond to an accuracy of 49%.

As a first comment on these results, and considering that the construction of the lexicon-grammar of PT-PT SVC is still a work in progress, one could say that these figures are promising, but that there is still much room for improvement. Next, we provide some error analysis and discuss the problematic results.

False-negatives (FN) cases

Four major situations can be distinguished: (i) the predicate nouns are still under description in the lexicon-grammar of STRING; (ii) the predicate nouns have not yet been included in the lexicon-grammar; (iii) the support-verb has not been associated with the predicate noun in the lexicon-grammar; and, (iv) some parsing issue prevented the system from capturing the SVC. We briefly present each one of these situations.

(i) nouns under description

Some of the instances not recognized by STRING concern predicate nouns that are already in the system's lexicon but are still undergoing linguistic description, so they were not used in the parsing. We did not consider these results to be a major problem, rather the natural consequence of a work in progress. These are the following predicate nouns (notice that some are repeated): *baixa*, *convenção*, *convivência*, *dano*, *decisão*, *dever*, *disponibilidade*, *extorsão*, *facilidade*, *grandeza*, *hábito*, *maneira*, *medida*, *obra*, *perda*, *rachadura*, *reclamação*, and *validade*.

(ii) nouns missing in the lexicon-grammar

On the other hand, there is an expressive number of predicative nouns that are not in the STRING's lexicon. Some of these nouns are used in quite usual constructions, hence the urgency in integrating them into the lexicon and to make an adequate description of them: *aniversário*, *antecedentes*, *área*, *autonomia*, *características*, *chefe*, *crime*, *endereço*, *equívoco*, *êxito*, *favoritismo*, *grafitagem*, *homicídios*, *índice*, *lar*, *lembranças*, *matchpoints*, *moleza*, *padrão*, *passado*, *população*, *potencial*, *prazer*, *presença*, *problema*, *procedimento*, *propriedade*, *repertório*, *significado*, *subvenção*, *tempo*, *tratado*, *treinamento*, *turnê* (PT-PT: *turné*, from French: *tournée*), *video-chamada* (orthographic variant of: *videochamada*), *vínculo*.

(iii) support-verb is not associated with the predicate noun in the lexicon-grammar

There is an important number of cases where the SVC has not been identified because the support-verb had not been associated with that particular predicate noun in the lexicon-grammar.

realizar: *ação*, *apresentação*, *audiência*, *concorrência*; ***cometer***: *assalto*; ***assinar***: *acordo*, *contrato*; ***ter***: *cura*, *marcação*, *relação*; ***possuir***: *excentricidade* (Astron.), experiência, *inclinação* (Astron.), *poder*; ***apresentar***: *sinal*.

Several support-verbs, often occurring in converse constructions, have also been missed in the lexicon-grammar: ***sofrer***: *acidente*; ***levar***: *advertências*, *medo* (only in PT-BR), *tombo*; ***passar por*** (=*sofer*): *cirurgia*; ***tomar***: *cuidado*, gols (BR), *precaução*, *providência*; ***chegar a***: *orgasmo*.

(iv) parsing issues

A large number cases correspond to situations where the system failed to recognize the SVC. It would not be possible to go through all those cases in this paper, and a detailed debug of the system's performance is underway.

Some situations, however, can already be reported, for they are clear. For example, in the next sentence, the predicate noun *ajuste* 'adjustment' has been incorrectly PoS-tagged as a verb (*ajustar* 'adjust'), so the SVC was not captured.

> *A prática de fazer* ajuste *no superávit com os dividendos tem sido comum nos últimos anos* 'The practice of adjusting the surplus with dividends has been common in recent years'

In other cases, a particular syntactic construction prevented the parsing from extracting the key dependency required to capture the SVC. This is the case of sentence:

> *Três integrantes de um bando que fez* um dos maiores *ataques a banco dos últimos anos*

> *no Estado* . . . 'Three members of a gang that carried out one of the biggest bank attacks in recent years in the State . . . ',

where the partitive determiner *um dos Adj ataques* 'one of the Adj attacks' precluded the extraction of the direct object dependency between the support-verb *fazer* and the predicate noun *ataque*.

In some of the cases above, as the SVC detection is carried out at a later stage of the rule-based parsing process, accumulated errors in the previous stages impede the correct identification and extraction of the SUPPORT _VSUP dependency. This particularly obvious in the case of PoS-tagging errors. Other situations may involve some development and further refinement of the STRING's underlying rule-based grammar.

True-negatives (TN) cases
Several predicate nouns (such as *checagem* 'checking'), or, else, specific support-verb-noun combinations (e.g. *levar medo* 'be afraid'), are exclusive of the PT-BR variety, so they could not have been previously included the PT-PT lexicon-grammar. This is the case of: ***dar***: *olhada*; ***levar***: *bola*, *medo*, ***realizar***: *checagem*; ***receber***: *premiação*; ***sofrer***: *pane*, ***ter***: *contato* (em PT-PT *contacto*); ***tomar***: *gols*.
In other cases, the PT-BR shows a specific lexical (*registro*) or orthographical variant (*pênalti*, in PT-PT: *penálti*) of the word, that have not been properly lemmatized in the STRING's lexicon.
A few spelling errors also prevented the system from matching. For example, the hesitant use of the hyphen is one of those cases: *video-chamadas*.

False-positives (FP) cases
False-positive (FP) cases, though in a smaller number, correspond to the situation where, for some reason, the system failed the parsing. The following is an interesting example:

> *Seria uma boa surpresa e uma prova de que amor não tem hora nem dia marcados.*

In this case, the system extracts a direct complement dependency between *tem* 'have' and *prova* 'proof', hence triggering the extraction of the SUPPORT dependency: `SUPPORT_-VSUP-STANDARD(prova,tem)`.

6. Conclusions and future work

In this paper, we analyzed the support-verb constructions manually annotated in a publicly available corpus of Brazilian Portuguese (PT-BR) multiword expressions (MWE), which was originally built within the scope of the project PARSEME. We emphasize that one of the reasons for using this corpus is the fact of it being publicly available and having been independently annotated. We parsed this corpus using the natural language processing system STRING, purposefully developed for Portuguese, whose SVC lexicon-grammar has been specifically built for the European variety (PT-PT). The construction of this linguistic resource is still ongoing. Both the system and its rule-based parser, as well as the SVC lexicon-grammar of European Portuguese were briefly described. The goal of this paper was to gauge the performance of the STRING system on this corpus with texts from the Brazilian variety (PT-BR), manually annotated for SVC. We briefly described the PARSEME corpus and the way it was processed by STRING, to retrieve the dependencies corresponding to the syntactic relation between the support-verb and the predicate noun. We compare the corpus SVC annotations with the STRING output. Results are encouraging, as precision is high (91%), but there is still much room for improvement, since recall is relatively low (42%). Many of the predicative nouns in the corpus that were not recognized by STRING were already included in the system's lexical-syntactical database, but had not yet undergone a full linguistic description, so they had been left out of the parsing. Similarly, for many predicate nouns, though they had been already described in the lexicon-grammar and were used by the parser, the full range of the variants of the basic or elementary support-verb had not been yet encoded in the lexicon-grammar. In some cases, this is due to the PT-BR language variety, for example, in cases where *possuir* 'possess' is often used in PT-BR instead of the elementary support-verb *ter* 'have', more common in PT-PT. The same seems to happen with many instances of *realizar* 'perform', a common variant of *fazer* 'do/make'. On the use of these (so-called) *stylistic* variants (Ranchhod, 1990; Baptista, 2005b) and the extension of elementary support-verbs, our approach to SVC is similar to that of the PARSEME Directives, namely, we also:

> take a broader scope than what is usually considered in the literature by taking in cases in which the verb has light semantics *per se* (it only bears morphology, such as the tense and mood, in any case), which hence cannot be described as "bleached" as is usually said of support-verbs.

On the other hand, we adopt the general Lexicon-Grammar approach, as posited by (Fotopoulou et al., 2021) for *aspectual* variants of support-verbs, which, in our view, would improve the granularity of SVC description within the PARSEME framework.
In other cases, the PT-PT lexicon-grammar lacks sufficient coverage in the determination of lexical variants of support-verbs. For example, the variants *passar/passar por* 'pass, pass through' had been entirely left out. Naturally, as a work in progress, the description of the lexicon-grammar is yet to be concluded. In this sense, a non-negligible number of predicate nouns had not yet been even listed in the database, and

some of them are quite usual/commonly used predicates. Their integration in lexicon-grammar and subsequent linguistic description is, therefore, urgent.

A more complex situation arises when the SVC construction, in principle, should have been recognized by STRING, but was not (false-negative). The detailed analysis of these cases is still underway but it may be due to a variety of causes. Paramount among them is the fact that the SVC detection and the extraction of the SUPPORT_VSUP dependency is performed by the STRING rule-based parser at the final steps of processing, hence it suffers from the accumulation of errors in the previous analysis stages. A major factor in this sub-optimal performance comes from the PoS-tagging phase. In other cases, the syntactic context is such that, in the previous stages of parsing, it prevents the key dependencies required for SVC extraction (e.g. a direct object, a subject or even a simple noun modifier dependency) from being adequately extracted, thus hindering the SVC extraction phase. For example, the specific syntactic structure of a noun phrase headed by the predicate noun with the support-verb participle as its modifier (e.g. *ação realizada* 'action performed'), a structure derived from the SVC passive sentence (v.g. *realizar uma ação*), though it had been mentioned in (Baptista and Mamede, 2020; Barreiro et al., 2022), does not seem to be working properly in the parser. Since this is a common structure used in PT-BR news, a relevant part of the missing SVC seems to be due to this case.

The partition of the PARSEME annotated corpus, as well as the reference built for this paper are to be made available in the STRING project site to the NLP (and especially to the MWE/SVC) interested community.

Besides the completion of the PT-PT SVC lexicon-grammar, several venues are open to future work. We envisage the analyse the entire corpus fully parsed with STRING, once the lexicon-grammar has been deemed satisfactorily complete, in order to retrieve: new (missed) instances of SVC, as well as instances of operator verbs. The lexicon-grammar and the new annotation will then be made available to the community.

7. Acknowledgements

Research for this work was supported by national funds through Institution, under grant ref. UIDB/50021/2020.

8. Bibliographical References

Aït-Mokhtar, S., Chanod, J.-P., and Roux, C. (2002). Robustness beyond shallowness: Incremental deep parsing. *Natural Language Engineering, 8*, pages 121–144.

Baptista, J. and Mamede, N. (2020). Syntactic Transformations in Rule-Based Parsing of Support Verb Constructions: Examples from European Portuguese. In Alberto Simões, et al., editors, *9th Symposium on Languages, Applications and Technologies (SLATE 2020)*, volume 83 of *OpenAccess Series in Informatics (OASIcs)*, pages 11:1–11:14, Dagstuhl, Germany. Schloss Dagstuhl–Leibniz-Zentrum für Informatik.

Baptista, J. and Ranchhod, E. (1998). Propriétés des phrases élémentaires associées à l'expression de grandeurs mesurables. exemples du portugais. *Cahiers de l'Institut de Linguistique de Louvain*, 24(3-4):49–61, September 24-27, 1998.

Baptista, J., Rassi, A. P., Santos-Turati, C., de Barros, C. D., Vale, O. A., and Mamede, N. (2015). Integrated processing of support verb constructions in Portuguese, 19-20 March 2015.

Baptista, J. (1997a). Conversão, nomes parte-do-corpo e restruturação dativa. In Ivo Castro, editor, *Actas do XII Encontro da Associação Portuguesa de Linguística*, volume 1, pages 51–59, Lisboa, 30 de Setembro a 2 de Outubro de 1996, Braga-Guimarães, Portugal. Associação Portuguesa de Linguística, APL/Colibri.

Baptista, J. (1997b). *Sermão*, *tareia* e *facada*: uma classificação das expressões conversas *dar–levar*. *Seminários de Linguística 1*, pages 5–37.

Baptista, J. (1999). *Fazer/Fazer com*: Um verbo-operador do português. *Seminários de Linguística*, 3:163–171.

Baptista, J. (2004). Instrument nouns and fusion. predicative nouns designating violent actions. In Christian; Leclère, et al., editors, *Lexique, Syntaxe et Lexique-Grammaire (Syntax, Lexis and Lexicon-Grammar). Hommage à Maurice Gross.*, Linguisticae Investigationes Supplementa, pages 31–40. John Benjamins Publishing Co., Amsterdam/Philadelphia.

Baptista, J. (2005a). Construções simétricas: argumentos e complementos. In Olga Figueiredo, et al., editors, *Estudos de homenagem a Mário Vilela*, pages 353–367. Faculdade de Letras da Universidade do Porto.

Baptista, J. (2005b). *Sintaxe dos Predicados Nominais com* ser de. Fundação para a Ciência e a Tecnologia & Fundação Calouste Gulbenkian, Lisboa.

Barreiro, A., Mota, C., Baptista, J., Chacoto, L., and Carvalho, P. (2022). Linguistic resources for paraphrase generation in Portuguese: a Lexicon-Grammar approach. *Language Resources and Evaluation*, pages 1–35.

Borillo, A. (1971). Remarques sur les verbes symétriques. *Langue Française*, 11(1):17–31.

Calcia, N. P. and Vale, O. (2019). Construções conversas do português do Brasil. Descrição e classificação iniciais. *Linguamática*, 10(2):13–20, January.

Calcia, N. (2016). Descrição e classificação das construções conversas no português do Brasil. Master's thesis, Universidade Federal de São Carlos, São Carlos-SP, Brasil.

Calcia, N. P. (2022). *Dar e receber um abraço: uma análise da conversão em português brasileiro*. Ph.D.

thesis, Universidade Federal de São Carlos, São Carlos-SP, Brasil, março.

Chacoto, L. (2005). *O Verbo FAZER em Construções Nominais Predicativas*. Ph.D. thesis, Universidade do Algarve, Faro.

Constant, M., Eryigit, G., Monti, J., van der Plas, L., Ramisch, C., Rosner, M., and Todirascu, A. (2017). Multiword expression processing: A survey. *Computational Linguistics*, 43(4):837–892.

Cook, P., Mitrović, J., Escartín, C. P., Vaidya, A., Osenova, P., Taslimipoor, S., and Ramisch, C. (2021). Proceedings of the 17th Workshop on Multiword Expressions (MWE 2021). In *Proceedings of the 17th Workshop on Multiword Expressions (MWE 2021)*.

de Athayde, M. F. M. Q.-P. (2000). *A estrutura semântica das construções com verbo-suporte preposicionadas do português e do alemão*. Ph.D. thesis, Faculdade de Letras da Universidade de Coimbra, Coimbra.

Dias de Barros, C., Vale, O. A., and Baptista, J. (2013). Description and classification of nominal predicates with the support verb fazer (make/do) in Brazilian Portuguese. In Jorge Baptista et al., editors, *Proceedings of the 32nd International Conference on Lexis and Grammar (CLG'2013)*, pages 165–170, Faro, Portugal, September. Universidade do Algarve – FCHS.

Dias de Barros, C. (2014). *Descrição e classificação de predicados nominais com o verbo-suporte* fazer*: especificidades do Português do Brasil*. Ph.D. thesis, Universidade Federal de São Carlos, São Carlos-SP, Brasil.

Fotopoulou, A., Laporte, E., and Nakamura, T. (2021). Where do aspectual variants of light verb constructions belong? In *Proceedings of the 17th Workshop on Multiword Expressions (MWE 2021)*, pages 2–12. Association for Computational Linguistics.

Giry-Schneider, J. (1978). *Les nominalisations en Français: l'opérateur faire dans le lexique*. Librairie Droz, Genova.

Gross, M. (1981). Les bases empiriques de la notion de prédicat sémantique. *Langages*, 15(63):7–52.

Gross, G. (1989). *Les construction converses du français*. Droz, Genève.

Gross, M. (1996). Lexicon-grammar. In Keith Brown et al., editors, *Concise Encyclopedia of Syntactic Theories*, pages 244–259. Pergamon, Cambridge.

Gross, M. (1998). La fonction sémantique des verbes supports. *Travaux de Linguistique: Revue Internationale de Linguistique Française*, 37(1):25–46.

Harris, Z. S., (1955). *Co-occurrence and transformation in linguistic structure*, pages 143–210. D. Reidel Publishing Company.

Harris, Z. (1964). The elementary transformations. In Henry Hiz, editor, *Papers on Syntax*, pages 211–235. D. Reidel Publishing Company.

Harris, Z. (1976). *Notes du Cours de Syntaxe*. Editions du Seuil, Paris.

Harris, Z. S. (1982). *A Grammar of English on Mathematical Principles*. New York: John Wiley & Sons.

Harris, Z. S. (1991). *A Theory of Language and Information. A Mathematical Approach*. Clarendon Press, Oxford.

Mamede, N., Baptista, J., Diniz, C., and Cabarrão, V. (2012). STRING - A Hybrid Statistical and Rule-Based Natural Language Processing Chain for Portuguese. In Alberto Abad, editor, *Int. Conf. on Computational Processing of Portuguese (PROPOR 2012) - Demo Session*. http://www.inesc-id.pt/ficheiros/publicacoes/8578.pdf.

Mota, C., Baptista, J., and Barreiro, A. (2018). The lexicon-grammar of predicate nouns with *ser de* in Port4NooJ. In *International Conference on Automatic Processing of Natural-Language Electronic Texts with NooJ*, pages 124–137. Springer.

Paiva Raposo, E. (2013). Predicações secundárias. In Eduardo et al. Paiva Raposo, editor, *Gramática do Português*, volume 2, pages 1340–1356. Fundação Calouste Gulbenkian.

Picoli, L., Vale, O. A., and Laporte, E. (2021). Aspecto verbal nas construções com verbo-suporte. *Revista do GEL*, 18(1):204–229.

Ramisch, C., Savary, A., Guillaume, B., Waszczuk, J., Candito, M., Vaidya, A., Barbu Mititelu, V., Bhatia, A., Iñurrieta, U., Giouli, V., Güngör, T., Jiang, M., Lichte, T., Liebeskind, C., Monti, J., Ramisch, R., Stymne, S., Walsh, A., and Xu, H. (2020). Edition 1.2 of the PARSEME shared task on semi-supervised identification of verbal multiword expressions. In *Proceedings of the Joint Workshop on Multiword Expressions and Electronic Lexicons*, pages 107–118, online, December. Association for Computational Linguistics.

Ranchhod, E. (1990). *Sintaxe dos predicados nominais com "estar"*. Instituto Nacional de Investigação Científica (INIC).

Rassi, A., Santos-Turati, C., Baptista, J., Mamede, N., and Vale, O. (2014). The fuzzy boundaries of operator verb and support verb constructions with dar "give" and ter "have" in Brazilian Portuguese. In *Proceedings of the Workshop on Lexical and Grammatical Resources for Language Processing (LG-LP 2014), COLING 2014*, Dublin, Ireland, August. Workshop on Lexical and Grammatical Resources for Language Processing (LG-LP 2014), COLING 2014, Dublin, August 24, 2014, COLING 2014.

Rassi, A., Mamede, N., Baptista, J., and I, O. V. (2015a). Integrating support verb constructions into a parser. In *Proceedings of the Symposium in Information and Human Language Technology (STIL'2015)*, pages 57–62.

Rassi, A. P., Baptista, J., and Vale, O. A. (2015b). Um corpus anotado de construções com verbo-suporte em português. *Gragoatá*, 39(1):207–230, Junho, 2015.

Rassi, A., Calcia, N., Vale, O. A., and Baptista, J.

(2016). Estudo contrastivo sobre construções conversas em PB e PE. In O. Nadin, et al., editors, *Léxico e suas interfaces: descrição, reflexão e ensino*, pages 199–218. Cultura Acadêmica.

Rassi, A. P. (2015). *Descrição, classificação e processamento automático das construções com o verbo* dar *em português brasileiro*. Ph.D. thesis, Universidade Federal de São Carlos, São Carlos-SP, Brasil.

Reis, S., Mamede, N., and Baptista, J. (2021). Predicados de comunicação em português europeu: nominalizações e nomes autónomos. *Revista da Associação Portuguesa de Linguística*, (8):237–259, Out.

Sag, I. A., Baldwin, T., Bond, F., Copestake, A., and Flickinger, D. (2002). Multiword Expressions: A Pain in the Neck for NLP. In Alexander Gelbukh, editor, *Proceedings of CICLing*, pages 1–15, Mexico City, Mexico, February.

Santos, C. (2015). *Construções com verbo-suporte* ter *no Português do Brasil*. Ph.D. thesis, Universidade Federal de São Carlos, São Carlos-SP, Brasil.

Savary, A., Sailer, M., Parmentier, Y., Rosner, M., Rosén, V., Przepiórkowski, A., Krstev, C., Vincze, V., Wójtowicz, B., Losnegaard, G. S., Parra Escartín, C., Waszczuk, J., Constant, M., Osenova, P., and Sangati, F. (2015). PARSEME–PARSing and Multiword Expressions within a European multilingual network. In *7th Language & Technology Conference: Human Language Technologies as a Challenge for Computer Science and Linguistics (LTC 2015)*.

Vaza, A. (1988). Estruturas com nomes predicativos e verbo-suporte *DAR*. Master's thesis, Faculdade de Letras da Universidade de Lisboa, Lisboa, Portugal.