

Integrated Processing of Support Verb Constructions in Portuguese

J Baptista^{1,2}, A Rassi³, C Santos-Turati³, C Barros³, O Vale³, N Mamede^{2,4}

¹ U. Algarve - FCHS/CECL (Portugal), ² INESC-ID Lisboa/L2F-Spoken Language Lab (Portugal)

³ U. Federal São Carlos, SP (Brazil), ⁴ U. Lisboa/Instituto Superior Técnico (Portugal)

Abstract

This paper describes the integration of a large-sized lexicon-grammar of over 10K Portuguese support-verb constructions (Baptista 2005, Barros 2014, Santos-Turati 2012, Rassi et al. 2014) in the grammar of a rule-based dependency parser that is part of a fully-fledged natural language processing chain, *STRING* (Mamede et al. 2012), developed for Portuguese.

Parsing support verb constructions in *STRING*

In spite of its MWE nature, it is not adequate to parse most SVC as a *compound (fixed) word*, but rather as a *special type of collocation* that yields to general syntactic analysis. Therefore, only at the semantic event identification, parsing stage are the SVC adequately tagged.

Integrating support verb constructions in *STRING*

Because of the large size of the lexicon involved, and the fact that SVC constructions are being described by several linguists separately, though coordinately, the following methodology was adopted towards a smooth cumulative procedure: (i) the partial lexicon-grammars (one for each elementary support verb, normally) undergo a process of validation, using a validator to ensure formal consistency; (ii) a script then converts the table data into rules, using the XIP parser syntax — the parsing module of *STRING*; the rules are then integrated in the *STRING* processing chain.

Npred	Variety	subject		1st complement			Vsup-STANDARD			Vsup-CONVERSE			Vop-CAUSE			Example		
		N	O	N	I	I	CDIR	PREDSUBJ	MOD	CDIR	PREDSUBJ	MOD	Nasp	CDIR	MOD			
[lemma:compra,pl]	BP	-	agent	-	-	-	-	-	-	-	-	-	-	-	-	-	A Ana faz compras 'Ana does some shopping'	
abraço	BP	-	agent	a,em	+	-	patient	-	dar	-	-	-	-	-	-	-	Eva deu um abraço em João 'Eve gave a hug to John'	
alergia	BP	-	experien	a	+	-	patient	-	fazer, ter	estar com, ser de	sofrer de	-	-	-	-	-	Eva tem alergia ao pólen 'Eve has allergy to pollen'	
autorização	BP	-	agent	a	+	-	patient	-	dar, conceder	-	-	-	-	-	-	-	O João deu autorização a Eva para sair 'John gave Eva permission to leave'	
caso	BP	-	experien	com	+	-	co-experien	+	ter, manter	-	-	-	-	-	-	-	Eva tem um caso com João 'Eve has an affair with John'	
chute	BP	-	agent	a,em	+	-	patient	-	dar	-	-	-	-	-	-	-	João deu um chute na bola 'John kicked the ball'	
comando	BP	-	agent	de	+	-	patient	-	ter	estar a/em	-	-	estar sob	contar	-	-	Eva está ao comando da empresa 'Eve is at the command of the company'	
conversa	BP	-	agent	com	+	-	co-agent	+	ter	estar a	-	-	-	-	-	-	Eva teve uma conversa com João 'Eve had a talk with John'	
convulsão	BP	-	experien	-	-	-	patient	-	sofrer, ter	estar com	-	-	-	-	-	-	Eva teve uma convulsão 'Eve had a convulsion'	
dor	BP	-	agent	em	-	+	object-ci	-	ter	estar com	-	-	-	-	-	-	Eva tem uma dor no pulso 'Eve has a pain in the wrist'	
estouro	BP	-	agent	-	-	-	patient	-	dar	-	-	-	-	-	-	-	O motor deu um estouro 'The engine gave a bang'	
gemido	BP	-	agent	-	-	-	patient	-	dar, soltar	-	-	-	-	-	-	-	Eva deu um gemido 'Eve gave a moan'	
golpe de judo	BP	-	agent	a,em	+	-	patient	-	dar, fazer	-	-	-	-	-	-	-	João deu um golpe de judo ao Pedro 'John gave a judo kick to Peter'	
guarda	BP	-	experien	de	+	-	patient	-	ter	-	-	-	estar a/sob	-	-	-	Eva tem a guarda das filhas 'Eve has the legal keeping of the children'	
guarda	BP	-	agent	a	+	-	object	-	fazer, montar	estar de	-	-	-	-	-	-	João está de guarda à casa 'John is on guard to the house'	
influência	BP	+	agent	em, sobre	+	-	patient	-	ter, manter	estar com, ser de	-	-	ter	estar sob	contar	-	-	Eva tem influência sobre o João 'Eve has some influence over John'
lua-de-mel	BP	-	experien	com	+	-	co-experien	+	-	-	-	-	-	-	-	-	-	Eva está em lua-de-mel com o João 'Eve is on honeymoon with John'
lua-de-mel	BP	-	agent	PrepLoc	+	-	locative	-	fazer	estar de	-	-	-	-	-	-	-	Eva e João fizeram a lua-de-mel nas Açores 'Eve and John did the honeymoon'
multa	BP	-	agent	a,em	+	-	patient	-	dar, passar	-	-	-	apanhar, ganhar	-	-	-	-	O polícia passou uma multa ao João 'The cop gave a ticket to John'
murro	BP	-	agent	a,em	+	-	patient	-	dar, pregar	-	-	-	-	-	-	-	-	O João deu murro ao Pedro 'John gave a punch to Peter'
raíche	BP	+	experien	-	-	-	patient	-	ter	-	-	-	-	-	-	-	-	O trabalho de Eva teve um certo raíche 'Eve's work had a certain ?'
saúde	BP	-	agent	a	+	-	patient	-	fazer	-	-	-	-	-	-	-	-	Eva fez uma saúde ao João 'Eve made a toast to John'
saúde	BP	-	experien	-	-	-	patient	-	ter	estar com, ser de	-	-	sofrer	-	-	-	-	Eva tem bastante saúde 'Eve has a good health'
sede	BP	-	experien	-	-	-	patient	-	sofrer, ter	-	-	-	-	-	-	-	-	Eva tem sede 'Eve has thirst'
tapo	B	-	agent	em	-	-	patient	-	dar, pregar	-	-	-	apanhar, ganhar	-	-	-	-	Eva deu um tapo no João 'Eve gave a slap to John'
telefonema	BP	-	agent	a	+	-	addressee	-	fazer	-	-	-	-	-	-	-	-	Eva fez um telefonema ao João 'Eve made a phone call to John'

Table 1. Sample of the Lexicon-Grammar of Portuguese predicative nouns

Support verb constructions and Multiword expressions

- *multiword expressions (MWE)* are a hot topic in *natural language processing (NLP)*, both for the linguistic and the computational issues they raise (Sag et al. 2002);
- in the literature, main focus has been in the *automatic identification* (in the sense of or 'automatic discovery' or 'lexical acquisition') of MWE and their extraction from texts (Páez 2014); fewer works deal with their processing;
- MWE are usually considered as equivalent to compound words for NLP purposes, and they are processed as text units, in as much the same way as a single-word lexical item; this, however, is not always adequate;
- *support verb constructions (SVC)*; M. Gross 1981) can be considered a type of MWE: e.g. *give a kiss to sb., take a nap, do a report on sth., make a comment about sth., throw an accusation to sb., get an answer from sb., be of interest to sb., be at ease, etc.*;
- SVC are verb-noun combinations where the nucleus is a *predicative noun (Npred)*
- *Npred* is responsible for the syntactic structure, the selection restrictions and the transformational properties (or 'alternations') of the sentence.
- In SVC the *verb* is just a sort of an auxiliary, a grammatical device, mostly devoid of meaning, and whose function is to *support* the noun by conveying the grammatical features of tense-modality and person-number, which the noun can not express; hence the term *support-verb (Vsup)*.
- many SVC can be considered as *nominalizations* of verbal or adjectival structures
- however, SVC are defined by the particular syntactic-semantic properties of the construction of the verb-noun combination, irrespective of there being a nominalization or not;
- many more predicative nouns, morphosyntactically unrelated to verbs or adjectives, are seen to also form SVC (*sb. be in/suffer a crisis*); the concept of SVC thus allow for the adequate integration in the grammar of these *autonomous Npred*, by providing the natural base sentences these predicates determine.
- in many languages, the number of SVC is similar to (or even larger than) that of full (or distributional) verbs and predicative adjectives, thus they constitute an important subset of the meaning units of those languages.

Support verb constructions' challenges to parsing

SVC pose a serious challenge to NLP:

- SVC have syntactic structure, identical to that of an ordinary verb;
- SVC can undergo several, very general, syntactic *transformations* (passive, pronouncing, clefting, and others);
- a transformation specific to SVC is *Conversion* (G. Gross 1989):
e.g. *O Pedro deu um conselho ao João* 'Peter gave an advice to John'
= *O João recebeu um conselho do Pedro* 'John got an advice from Peter'
- most SVC can be defined by a *basic (or elementary) support-verb*, which has a very broad distribution across the lexicon;
- still, many SVC allow this verb to be replaced by *aspectual or stylistic variants*, that is, verbs that are themselves support-verbs, but introduce *aspectual/stylistic* nuances when compared against the elementary *Vsup* (e.g. *have / lose / keep / maintain faith in sth./sb.*)
- under certain syntactic conditions, SVC allow for the zeroing of the *Vsup*, the predicative noun keeps all its arguments and become an argument of another predicative element (verb, adjective, noun): e.g. *[John's advice to Peter] matters/was important/has importance*;

This paper describes the integration of a large-sized lexicon-grammar of over 10,000, commonly used, Portuguese SVC (Baptista 2005, Barros 2014, Santos-Turati 2012, Rassi et al. 2014) in the grammar of a rule-based dependency parser developed for Portuguese (Mamede et al. 2012).

The lexicon-grammar contains for each *Npred* the corresponding *Vsup*, the distributional constraints on the noun's arguments, the prepositions introducing the complement(s) and the transformational properties the sentence can undergo, such as *Passive*, or *Conversion* (Table 1).

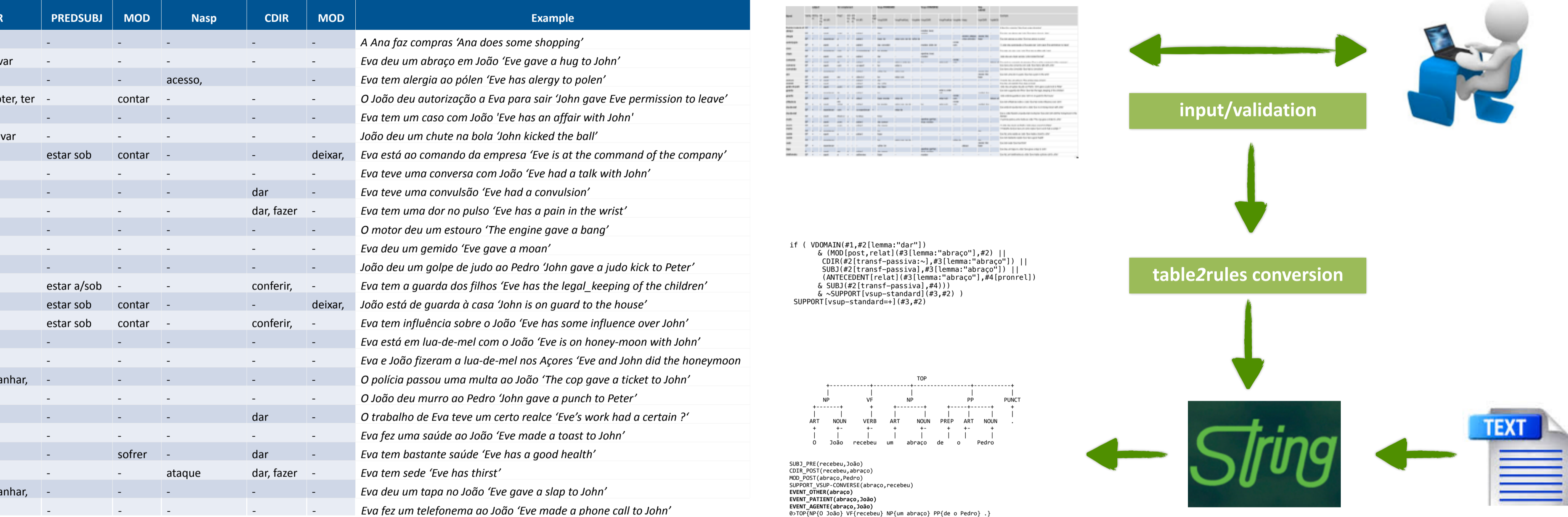


Fig 1. Parsing support verb constructions in *STRING*: example from parsing two SVC dar/receber abraço 'give/take hug', linked by linked by Conversion.

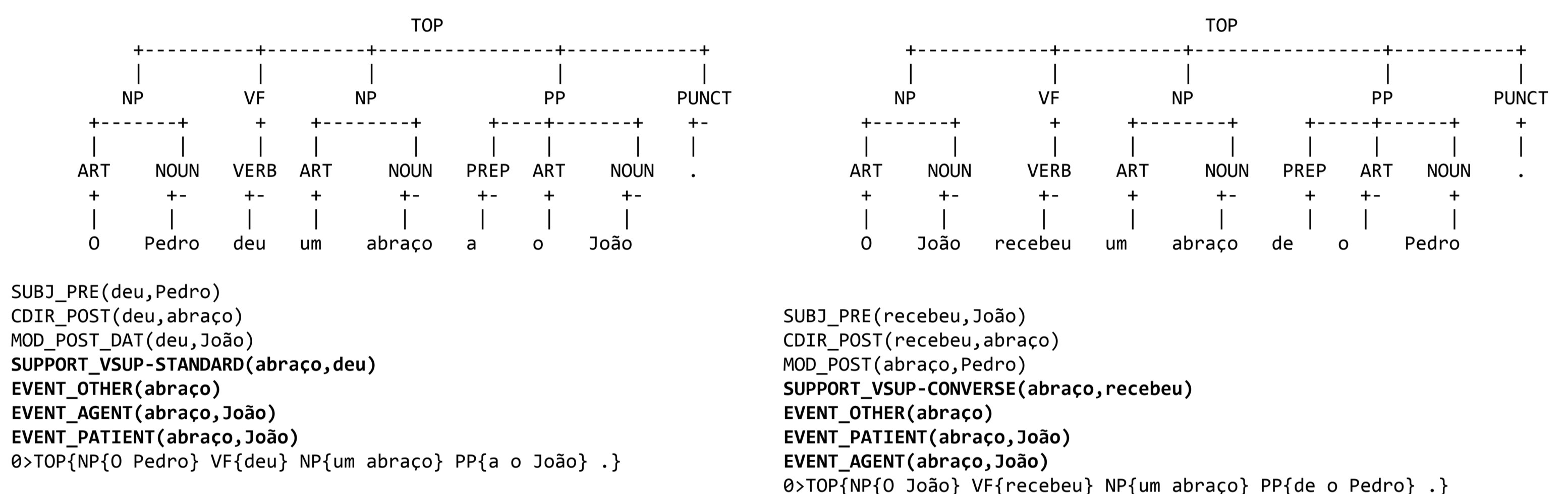


Fig 1 illustrates the parsing of two sentences, linked by Conversion. The *chunking tree* (top) identifies basic syntactic constituents (*chunks*) while the dependencies (center) link the chunks' heads. The chunked sentence (bottom) is also shown. For lack of space, several auxiliary dependencies were omitted. SUBJ = Subject, CDIR = Direct Complement, MOD = umbrella-dependency for Modifiers and Prepositional Complements, SUPPORT = links the predicative noun (governor) with its support verb, the feature *_VSUP-STANDARD* indicates this to be an active-like SVC (agent aligned with subject) while *_VCONVERSE* indicates the passive-like SVC; the EVENT dependency identifies the semantic predicate, while the features *_AGENT* and *_PATIENT* indicate the semantic roles of the predicate's arguments.

Strategy for parsing support verb constructions in *STRING*

The strategy adopted consists in: (i) letting the basic grammar build the parse as for any ordinary sentence; and then, (ii) using the lexicon-grammar information, extracting the adequate *syntactic-semantic dependencies* between the predicative noun, the support-verb and the noun's arguments. The goal is to capture the *semantic predicate* expressed by the *Npred*. This semantic predicate is captured by a generic dependency *EVENT*, on which the arguments' semantic roles will hang: So: besides having the syntactic dependencies between the sentence's constituents, one would also have the semantic structure of the predicate being expressed, as illustrated in Fig.1, above.

Fig 2. Identifying support verbs and extracting *SUPPORT* dependency in *STRING*

```

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

Figure 2 shows the rule, automatically generated by the *table2rule* converter, that extracts the *SUPPORT* dependency for sentence in Fig.1(a), based on the previously extracted syntactic dependencies. The rule uses the XIP parser (Ait-Mokhtar et al. 2002) syntax, integrated in *STRING*. The matched string is highlighted and read as follows: the *VDOMAIN* dependency is used to include auxiliary verb constructions of the support verb; the predicative noun is the direct object (*CDIR*) of the verb, which is not in a passive construction (*TRANSF-PASSIVA:~*); no standard support dependency has yet been extracted (*~SUPPORT*); then, the rule produces the *SUPPORT* dependency with the *vsup-standard* feature. The same rule also identifies the SVC passive construction, as well as the relative subclauses (active and passive)

Acknowledgments
Research for this paper was partially funded by FCT — Fundação para a Ciência e a Tecnologia (Portugal), under project ref. PEst-OE/EEI/LA0021/2014, by FAPESP (Brazil), and by CAPES/PDSE (Brazil), under processes BEX 6561/13-6 and BEX 12751/13-8.

In these two parses (left), notice that the layer of *syntactic dependencies* (SUBJ, CDIR, MOD) is complementary to the *semantic layer*, which identifies the *EVENT* or semantic predicate, its arguments, and their semantic roles (AGENT, PATIENT), which are the same, irrespective of the changes introduced by the Conversion operation.

Evaluation of parsing support verb constructions in *STRING*

Preliminary evaluation on 580 SVC with support verb *dar* 'give' (Rassi et al. 2014), from a corpus of 2,600 SVC sentences taken from journalistic text, manually annotated by 5 experts (Fleiss' Kappa 0.604: "substantial" *apud* Landis and Koch (1977)) gave interesting results (Table 2).

TP	FP	FN	TN	P	R	F	A
325	56	83	116	0.85	0.79	0.82	0.76

Table 2. Preliminary evaluation of dar 'give' SVC construction
Total of sentences: 580; TP = true positive, FP = false positive, FN = false negative, TN = true negative, P = Precision, R = Recall, F = F-measure, A = Accuracy.

Conclusions and future work

A framework for integrating support verb constructions in the fully-fledged NLP chain *STRING* has been implemented. Preliminary results are promising. Future work involves: (i) completion of the Lexicon-grammar integration (Ranchhod 1990, Baptista 2005, Barros 2012, Santos-Turati 2012) and systematic semantic roles' encoding; (ii) semi-automatic populating some fields: support verbs and their variants, prepositions, operator-verbs; (iii) the systematic comparison of the structures and properties of European and Brazilian variants of Portuguese (Rassi et al. 2014): lexical variants (*parada*^{BP}/*paragem*^{EP} *cardíaca* 'heart arrest'; *carona*^{BP}/*boleia*^{EP} 'hitchhiking'; *chamada* 'rebuque'^{BP,EP} / phone call^{EP}); variation of prepositions; and different sets of support verbs for the same noun in each variety.

References

Ait-Mokhtar, S., J. P. Chanod, C. Roux. Robustness beyond shallowness: incremental dependency parsing. *Natural Language Engineering*, 2002, pp.121-144.
Baptista, J. *Sintaxe dos Predicados Nominais com SER DE*. Lisboa: F. Calouste Gulbenkian/ FCT, 2005.
Barros, C. *Descrição e classificação dos predicados nominais com o verbo-suporte fazer em Português do Brasil* (PhD Thesis) Faculdade de Letras, Universidade Federal de São Carlos, São Carlos-SP, 2014.
Gross, G. *Les construction converses du français*. Genève: Librairie Droz, 1989.
Gross, M. Les bases empiriques de la notion de prédicat sémantique. *Langages*, 63 (15), 1981, pp. 7-52.
Mamede, N., J. Baptista, C. Diniz, V. Cabarrão. *STRING: An hybrid statistical and rule-based Natural Language Processing chain for Portuguese*. In: *Proceedings of PROPOR'12. International Conference on Computational Processing of Portuguese (Demo session)*, Coimbra, Portugal, April, 2012.
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, pp. 419-424.
Rassi, A.P., C. Santos-Turati, J. Baptista, N. Mamede, O. Vale. The fuzzy boundaries of operator verb and support verb constructions with DAR 'give' and TER 'have' in Brazilian Portuguese. *Workshop on Lexical and Grammatical Resources for Language Processing (LG-LP 2014)*, COLING 2014, Dublin, Ireland.
Rassi, A.P., C. Santos-Turati, J. Baptista, N. Mamede, O. Vale (submitted). Um corpus anotado de construções com verbo-suporte em Português.
Ranchhod, E. *Sintaxe dos predicados nominais com ESTAR*. Lisboa: INIC - Instituto Nacional de Investigação Científica, 1990.
Sag, I. A. et al. *Multword Expressions: A Pain in the Neck for NLP*. In: Gelbukh, A. (ed.), In: *Proceedings of the Third International Conference, CICLing - Computational Linguistics and Intelligent Text Processing*, Mexico City, Mexico, February, 2002, pp. 1-15.
Santos-Turati, M. C. A. *Descrição da estrutura argumental dos predicados nominais com o verbo-suporte TER*. In: *Seminário do GEL - Grupo de Estudos Linguísticos do Estado de São Paulo*, 60, São Paulo, Brasil, 2012, pp. 20-21.