

Integrated Processing of Support Verb Constructions in Portuguese

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

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

³ U. Fed. São Carlos, SP – Brasil, ⁴ Instituto Superior Técnico, U. Lisboa, Portugal

For a long time, but particularly after Sag *et al.* (2002), *multiword expressions* (MWE) have become a hot topic in *natural language processing* (NLP). The main focus in the literature has been, so far, the automatic identification (in the sense of discovery or automatic lexical acquisition) of MWE and their extraction from texts (Páez 2014). For most authors, most MWE are equivalent to compound words, and are to be processed as text units, in as much the same way as a single-word lexical item.

On the other hand, support verb constructions (SVC; M. Gross 1981), *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.*, are verb-noun combinations, where the nucleus of the semantic predicate is not the verb but the noun, hence it is called a *predicative noun* (*Npred*). This noun 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*).

SVC pose a serious challenge to NLP, as they have syntactic structure, often identical to that of an ordinary verb, and can undergo several, very general, syntactic transformations (passive, pronouncing, clefting and others). Furthermore, 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.*). Besides, under certain syntactic conditions, SVC allow for the zeroing of the *Vsup*. In these reduced structures, the predicative noun, while keeping all its arguments, can then become an argument of another predicative element (verb, adjective, noun): *e.g. [John's advice to Peter] matters/was important/has importance*. Without the verb, though, these complex noun phrase structures are more difficult to capture in parsing. Finally, many SVC can be considered as nominalizations of verbal or adjectival structures. However, as SVC are defined by the particular syntactic-semantic properties of the construction of the verb-noun combination, irrespective of there being a nominalization, many more predicative nouns, morpho-syntactically unrelated to verbs and adjectives, are seen to also form SVC (*sb. be in/suffer a crisis*). The notion 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.

This paper describes the integration of a large-sized lexicon-grammar of over 10K 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* (G. Gross 1989; e.g. *O Pedro deu um conselho ao João = O João recebeu um conselho do Pedro* 'give advice to/get advice from').

In spite of its MWE nature, it is not adequate to parse most SVC as a compound (fixed) word combinations, but rather as a special type of collocation that yields to general syntactic analysis. Having this in mind, the strategy here adopted consists in: first, letting the basic grammar build the parse as for any ordinary sentence; and then, 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 conveyed by the *Npred*, which is expressed by a generic dependency *EVENT*, so that, besides having the syntactic dependencies between the sentence's constituents, one would also have the semantic structure of the predicate being expressed, as illustrated below (Fig.1):

O Pedro fez um comentário ao livro 'Pedro made a commentary to the book'

SUBJ(feiz, Pedro)	SUPPORT_STANDARD(comentário, fez)	EVENT_OBJECT(comentário, livro)
CDIR(feiz, comentário)	EVENT_LEX(comentário)	
MOD(feiz, livro)	EVENT_AGENT(comentário, Pedro)	

Fig. 1. Dependency parsing of a SVC: SUBJ[ect]; MOD[ifier], umbrella tag for prepositional phrases, CDIR, direct complement; EVENT, generic dependency for representing semantic predicates; features: _LEX, the word that functions as the predicative node; _AGENT and _OBJECT, semantic roles of the EVENT arguments; SUPPORT, dependency holding between the predicative noun and its support verb; features: _STANDARD, for standard SVC, and _CONVERSE for converse SVC (not in the example).

Based on the properties formalized in the lexicon-grammar, a program generates the rules that describe the patterns produced by each construction, including standard and converse, active and passive SVC. The parser, XIP – Xerox Incremental Parser (Ait-Mokhtar *et al.* 2002), then applies these rules to texts. Preliminary evaluation on SVC with *Vsup dar* (Rassi *et al.* 2014), using a manually annotated corpus of 2,600 SVC sentences extracted from journalistic text, presented promising results.

References

- Ait-Mokhtar, S. ; Chanod, J. P.; Roux, c. 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. D. *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. *et al.* 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, At Dublin, Ireland.
- Ranchhod, E. *Sintaxe dos predicados nominais com Estar*. Lisboa: INIC - Instituto Nacional de Investigação Científica, 1990.
- Sag, I. A. *et al.* *Multiword 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.