

INTEGRAÇÃO DE PREDICADOS NOMINAIS EM *PARSER*: UMA EXPERIÊNCIA COM AS CONSTRUÇÕES COM O VERBO-SUORTE *DAR* EM PORTUGUÊS BRASILEIRO

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

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

Introdução

As construções com verbo-suporte (*CVS*) são predicados nominais formados por um verbo-suporte (*Vsup*) e um nome predicativo (*Npred*). Nesse sentido, para identificar

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

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

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

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

uma *CVS*, é necessário identificar tanto os verbos que podem funcionar como *Vsup* quanto os nomes predicativos que com eles se constroem. Neste trabalho, adotamos a noção de verbo-suporte da gramática transformacional de operadores de Harris (1991) e da abordagem do Léxico-Gramática (GROSS, 1975, 1981).

Além do conceito de *CVS*, há também diferentes testes que podem ser usados para identificar essas construções (RANCHHOD, 1990; BAPTISTA, 2005). O principal teste, que representa uma propriedade necessária e suficiente das *CVS*, é a estreita relação entre o *Npred* e tipicamente o sujeito¹ da construção (e.g. *Pelé deu um chute na bola*, interditando a construção **Pelé deu o chute do Neymar na bola*). Essa relação é da mesma natureza semântica que a relação existente entre o verbo e seu sujeito, em um predicado verbal (e.g. *Pelé chutou a bola*).

Além desse teste, há outros que podem ser indicativos de *CVS*, como a substituição da construção com *Vsup* por um verbo pleno correspondente (como *dar um abraço* = *abraçar*, ou *dar um beijo* = *beijar*); as restrições sobre os determinantes (como *Ana deu uma passeada no parque*, interditando a construção **Ana deu minha passeada no parque*); a descida do advérbio, que permite que um advérbio de uma construção verbal “desça” para a posição de adjetivo numa construção nominal (e.g. *Rui chutou fortemente a bola* = *Rui deu um chute forte na bola*); dentre outros testes.

Para uma visão mais geral sobre *CVS*, veja-se, entre outros, Gross (1981, 1994, 1998), Giry-Schneider (1978, 1987), Meunier (1981), Vivès (1983), Ranchhod (1990), Baptista (2005). A literatura oferece pelo menos duas abordagens distintas para esse fenômeno: (i) uma que considera as *CVS* como um bloco cujos constituintes são relativamente fixos, como um subtipo de expressões multi-palavras (*multi-word expressions*), tais como as palavras compostas e as expressões idiomáticas (CALZOLARI et al., 2002; SAG et al., 2002; DIAB; BHUTADA, 2009); (ii) e outra que considera as *CVS* como uma estrutura sintática complexa que segue as mesmas regras da gramática geral da língua, mas possui propriedades específicas e admite transformações sintáticas sistemáticas. Essa segunda abordagem, na qual este trabalho se insere, reconhece e descreve as redes de relações sintáticas existentes entre os constituintes de uma *CVS*.

As *CVS*, por serem fenômenos complexos, levantam uma série de desafios para seu processamento automático, como por exemplo as *CVS* nem sempre são resultantes de nominalizações; nem sempre o *Vsup* da *CVS* está explícito na frase; as construções nominais não mantêm necessariamente a mesma rede argumental de suas construções verbais equivalentes *etc.* Em decorrência disso, os analisadores sintáticos, em geral, não tratam esse fenômeno.

Os *parsers* (analisadores sintáticos automáticos) de referência, disponíveis para o Português, tais como o *Palavras* (BICK, 2000) e o *LX-parser* (SILVA et al., 2010), não dispõem ainda de informações sobre predicados nominais formados por *Vsup* e

¹ Diz-se que que a relação se dá entre o *Npred* e o sujeito, tipicamente, porque, nas construções *standard*, como em *A Ana deu um beijo no Rui*, essa relação se verifica entre o *Npred* *beijo* e o sujeito (*Ana*). Por outro lado, em construções conversas, tal como *O Rui recebeu um beijo da Ana*, essa mesma relação se verifica entre o *Npred* *beijo* e o complemento (*Ana*).

Npred. Há diferentes tipos de predicados nominais, mas, neste trabalho, trataremos especificamente das construções nominais cujo núcleo predicador está no nome (chamado nome predicativo) e este nome é auxiliado por um verbo (chamado verbo-suporte).

Nesse sentido, desenvolvemos uma análise linguística sistemática das *CVS*, adotamos uma formalização dos dados baseada na proposta do Léxico-Gramática (GROSS, 1975, 1981), inserimos os dados em uma cadeia de processamento do Português, a STRING (MAMEDE et al., 2012), e avaliamos o resultado do sistema com base na anotação manual de um *corpus*.

A análise, descrição e classificação dos dados foram feitas em três trabalhos diferentes: 1.815 predicados nominais formados pelo verbo-suporte *fazer* (BARROS, 2014); 2.273 predicados nominais com o verbo-suporte *ter* (SANTOS, 2015); e 1.489 predicados nominais com o verbo-suporte *dar* (RASSI, 2015) Todos esses dados vêm sendo sistematicamente analisados, descritos e formalizados em matrizes do Léxico-Gramática (L-G) das construções nominais do Português.

No L-G, a descrição dos fenômenos linguísticos é apresentada no formato de matrizes binárias: as linhas contêm as entradas lexicais (neste caso, os *Npred*) e as colunas representam as propriedades sintático-semânticas de cada entrada. Por exemplo, cada nome predicativo impõe restrições distribucionais ao tipo de argumentos, à preposição que introduz os complementos essenciais e ao determinante do nome predicativo; a matriz também codifica os verbos-suporte *standard* e conversos, bem como suas variantes aspectuais e/ou estilísticas; codifica os papéis temáticos dos argumentos; a possibilidade de aceitar ou não a conversão, a passiva, a simetria, dentre outras propriedades.

Apesar de haver descrições dos predicados nominais com os verbos *fazer*, *ter* e *dar*, este trabalho apresenta apenas os resultados da integração das construções nominais com o verbo-suporte *dar* na STRING.

A STRING é uma cadeia de processamento do Português, com uma estrutura modular, que executa as principais tarefas básicas de Processamento de Língua Natural (PLN), como a tokenização, a segmentação textual, a etiquetagem das partes-do-discurso (*POS tagging*), a desambiguação morfossintática, a análise sintática superficial (*chunking*), a análise sintática profunda, nomeadamente como a extração de dependências (sujeito, complemento *etc.*), dentre outras tarefas. A STRING usa o *parser* XIP (*Xerox Incremental Parser*), que é um analisador sintático baseado em regras e em estatística (MOKHTAR et al., 2002).

Os dados das construções com o verbo-suporte *dar* foram integrados na cadeia de processamento, como um dos componentes da gramática do Português, implementada no XIP. Isso foi feito sob a forma de regras léxico-sintáticas de extração de dependências, a fim de extrair automaticamente a dependência a que chamamos *SUPPORT* entre o *Vsup* e o *Npred* e entre o *Npred* e seus argumentos.

Em trabalho anterior (RASSI et al., 2015), descrevemos uma proposta geral para a extração de eventos e de dependências associados às construções com *Vsup* na

STRING. Naquele trabalho, indicamos a estratégia de implementação das construções com verbo-suporte. Recorde-se que as *CVS* podem formar construções *standard* (*A Ana deu um beijo no Rui* - SUPPORT [vsup-standard]), de orientação semântica ativa, ou construções conversas (*O Rui recebeu um beijo da Ana* - SUPPORT [vsup-converse]), de orientação semântica passiva.

Neste trabalho, descreveremos especialmente os resultados do processamento automático das *CVS* com o *Vsup dar* na STRING, comparando-os com a anotação manual de uma amostra do *corpus* PLN.Br Full (BRUCKSCHEN et al., 2008). A amostra total conta com 2.646 frases extraídas aleatoriamente do PLN.Br Full, com candidatos a pares de *Vsup* e *Npred*. Neste trabalho, no entanto, referimos apenas 580 frases dessa amostra total, que correspondem às frases com o verbo *dar* e suas variantes.

Estado da arte

Grande parte dos trabalhos que descrevem tarefas automáticas relacionadas às *CVS* tratam da identificação ou extração dessas construções a partir de *corpora*, seja com base em padrões lexicais (por meio de expressões regulares), seja com base na anotação manual de *corpus* e aprendizado de máquina.

Grefenstette e Teufel (1995) apresentam um método de identificação dos verbos-suporte a partir de um *corpus* não-etiquetado, pela comparação dos argumentos ligados às formas verbais e às potenciais formas nominalizadas, ou seja, transfere-se a rede argumental das construções verbais para os potenciais candidatos a construção nominal. Os autores buscam encontrar os verbos-suporte mais prováveis para cada nome predicativo, mas considerando apenas os *Npred* que são nominalizações. Sabe-se que muitos *Npred* são nominalizações de verbos, tais como nos pares *abraço=abraçar*, *apresentação=apresentar*, *chute=chutar* etc., mas há também *Npred* que são chamados *nomes predicativos autônomos*, que não são derivados de verbos, tais como *greve*, *sermão*, *cólica*, etc. Assim, o método apresentado pelos autores não captura os *Npred* autônomos. Naquele trabalho, Grefenstette e Teufel (1995) extraíram de *corpus* em inglês 6.704 sentenças com candidatos a verbos-suporte e candidatos a nominalizações, produzindo uma lista de potenciais construções com verbos-suporte que ocorrem com as formas nominalizadas. Além de desconsiderar os *Npred* autônomos, outro problema dessa abordagem consistiu em se considerar que a construção nominal mantém a mesma rede argumental que a construção verbal equivalente, o que nem sempre se verifica.

Para o espanhol, Páez (2014) extraiu de um *corpus* 81.274 frases com candidatos a *CVS*, dos quais os verbos-suporte mais representativos são *tener* (*ter*), *hacer* (*fazer*) e *dar* (*dar*). A autora extraiu também automaticamente as combinações de qualquer nome e 12 verbos, variantes dos *Vsup tener*, *hacer* ou *dar*. Em seguida, ordenou as principais combinações de verbo e nome por frequência e calculou a probabilidade de co-ocorrência de tal verbo com tal nome, usando três medidas de associação (*log-likelihood*, *Student's T-score* e *Maximum likelihood estimator*). Ao fim da tarefa,

a autora elencou as 15 *CVS* mais recorrentes em espanhol, de acordo com as medidas de associação utilizadas e concluiu que cerca de 69% das *CVS* dessa lista foram corretamente identificadas.

Na literatura, encontramos outros trabalhos, como o de Paez (2014), que partem de uma lista prévia de verbos que podem funcionar como *Vsup* ou uma lista de nomes que podem funcionar como *Npred*. A proposta de Duran et al. (2011) se diferencia dessas abordagens por partir dos padrões sintáticos de combinações de categorias gramaticais para encontrar *CVS*, como por exemplo V N PRP (*abrir mão de*), V PRP N (*deixar de lado*), V DET N PRP (*virar as costas para*), V DET ADV (*dar o fora*), V ADV (*ir atrás*), V PRP ADV (*dar para trás*), V ADJ (*dar duro*)².

Usando esse método, Duran et al. (2011) conseguiram identificar 773 predicados complexos, que foram, em seguida, anotados manualmente. Segundo os autores, esses predicados complexos incluem (mas não se limitam a) *CVS*, a que os autores chamam construções com *verbo leve*³. Consideramos, no entanto, que a utilização de expressões regulares com combinações de categorias gramaticais não será a abordagem mais adequada para a identificação exclusiva de *CVS*, pois as *CVS*, via de regra, são formadas por V (DET) N, o que é sintaticamente idêntico às estruturas dos predicados verbais, compostos por um verbo pleno (V), seguido de um objeto direto (N), que pode ou não ser introduzido por determinante (DET).

Por outro lado, há também trabalhos que objetivam processar (e não apenas identificar) essas construções. Cite-se, por exemplo, Barreiro et al. (2014), que avaliaram dois sistemas de tradução automática, o OpenLogos (baseado em regras) e o Google Translate (baseado em redes neurais), na tarefa de traduzir construções com verbo-suporte em cinco línguas: francês, alemão, italiano, português e espanhol. Para realizar os experimentos e a avaliação, os autores produziram um conjunto de 100 frases que analisaram como *CVS*⁴ e o anotaram manualmente. Como resultado da avaliação dos dois sistemas, os autores concluíram que o Google Translate traduz melhor as *CVS* do que o OpenLogos, atribuindo esse resultado ao seu rico conhecimento lexical.

No presente trabalho, com o intuito de contribuir para as tarefas de processamento das *CVS* e visando a preencher as lacunas para sua identificação automática, apresentamos a metodologia e os resultados da integração das *CVS* com o *Vsup dar* na STRING, por meio do *parser* XIP. Os resultados da performance desse sistema foram avaliados comparativamente com um *corpus* de referência, anotado manual e independentemente.

² ADJ = adjetivo, ADV = advérbio, DET = determinante, N = nome, PRP = preposição e V = verbo.

³ As *CVS* são também referidas na literatura como *construções com verbo leve* (SCHER, 2004; DURAN et al., 2011; TU; ROTH, 2011; BUTT; GEUDER, 2001; ISTVÁN; VINCZE; FARKAS, 2013). Os termos *verbo leve* e *verbo-suporte* são comumente interpretados como sinônimos, porém há diferenças conceituais entre as nomenclaturas. Adotamos o conceito de verbo-suporte por considerarmos que o verbo serve para “suportar” as marcas de número, pessoa, tempo, modo e aspecto.

⁴ Em rigor, nem todas as construções listadas pelos autores são construções nominais com *Vsup*, incluindo-se nelas igualmente construções adjetivais, preposicionais e até frases com verbo-operador.

Construção do *corpus* de referência para as CVS

Nesta seção, explicaremos brevemente os procedimentos adotados para a constituição do *corpus* de referência, sua anotação e a seleção de uma subamostra a ser processada na STRING. Todo o processo de construção e de anotação do *corpus* já foi abordado pormenorizadamente em trabalho anterior (RASSI et al., 2015).

As matrizes do Léxico-Gramática (nomes predicativos e verbos-suporte *fazer, ter e dar*) foram intersectadas com grafos de referência do Unitex⁵, a fim de procurar sistematicamente no *corpus* PLN.Br todas as possíveis combinações de um dos verbos-suporte com um dos nomes predicativos, considerando-se apenas as combinações previstas nas matrizes. Por meio dessa metodologia, foram identificadas, no *corpus*, 121.198 frases com pares de candidatos {*Vsup, Npred*}, isto é, frases em que ocorre simultaneamente um potencial *Vsup* e um nome predicativo.

Selecionamos uma amostra dessas 121.198 frases, proporcional à quantidade de ocorrências de cada par {*Vsup, Npred*}. A amostra é composta por 2.646 frases e corresponde a 2,18% do total de frases. Essa seleção recuperou pelo menos um caso de todos os pares {*Vsup, Npred*} que tenham, pelo menos, 21 ocorrências. A Tabela 1 resume as principais informações sobre o *corpus* e a amostra selecionada.

Tabela 1 – Dados da amostra comparativamente aos dados do *corpus*.

	<i>Corpus</i>	Amostra	%
Pares { <i>Vsup, Npred</i> } (nº de frases)	121.198	2.646	2,18%
Pares { <i>Vsup, Npred</i> } diferentes	4.668	1.130	24,2%

Fonte: Elaboração própria.

A amostra conta com 1.130 pares diferentes de {*Vsup, Npred*}, o que corresponde a 24,2% do *corpus*, que é composto por 4.668 pares diferentes.

A anotação das 2.646 frases com candidatos a CVS foi feita manualmente por 5 anotadores falantes nativos do Português e especialistas em CVS. Para tal, adaptou-se uma ferramenta de anotação de *corpora* já existente (SUÍSSAS, 2014). A anotação consistiu em assinalar, para cada frase, um código (convencional) que corresponde ao tipo de construção sintática indicada pelo par {*Vsup, Npred*} que aparece entre parênteses no início de cada frase. Cada anotador escolhe uma das etiquetas disponíveis:

CVS-STANDARD - para as construções com verbo-suporte standard

Ex.: (*dar, tapa*) *Ana deu um tapa em Rui.*

CVS-CONVERSE - para as construções com verbo-suporte converso

Ex.: (*levar, tapa*) *Rui levou um tapa da Ana.*

⁵ O Unitex é um conjunto de softwares que permitem o processamento textual de grandes *corpora*, disponível em: <http://www-igm.univ-mlv.fr/~unitex/>

VOPC - para as construções com verbo operador causativo

Ex.: (*dar,medo*) *O vento sombrio deu medo em Ana.*

Ex.: (*fazer,medo*) *O vento sombrio fez com que Ana tivesse medo.*

OTHER - para qualquer outro tipo de construção

Ex.: (*fazer,academia*) *Rui fez (=construiu) uma academia.* [verbo pleno]

Ex.: (*dar, tiro*) *O governo deu um tiro no próprio pé.* [expressão fixa]

Ex.: (*ter, controle*) *Rui tem Max sob seu controle.* [verbo-operador de ligação]

Ao final do processo, as anotações foram tabuladas em 5 colunas e, em seguida, comparadas com auxílio à ferramenta ReCal 0.1 Alpha for3+ Coders⁶, para calcular a concordância entre os anotadores. A concordância média entre os 5 anotadores foi de 80,8%. A ferramenta calcula igualmente o coeficiente Kappa (COHEN, 1960), que é uma medida estatística que avalia a concordância entre pares de avaliadores, também chamada de *inter-annotator agreement* ou *inter-rater agreement*. A média do Cohen's Kappa atribuída à anotação foi de 0.604.

A partir da anotação manual das 2.646 frases, foram selecionadas aquelas que tiveram concordância igual ou superior a 60%, ou seja, aquelas em que 3 ou mais anotadores atribuíram a mesma etiqueta. A Tabela 2 apresenta o número de frases por grau de concordância entre os anotadores, na amostra geral.

Tabela 2 – Distribuição de frases por grau de concordância entre anotadores, na amostra geral.

Concordância	Nº de frases	% de frases
2 anotadores	44	1,7%
3 anotadores	326	12,3%
4 anotadores	627	23,7%
5 anotadores	1574	59,5%
*frases com erro	75	2,8%
TOTAL	2.646	100%

Fonte: Elaboração própria.

Como se verifica, as frases anotadas pela totalidade dos anotadores correspondem a quase 60%. Somando-se as frases anotadas com a mesma etiqueta pela maioria dos anotadores (3 ou mais), a concordância corresponde a 95,5% do total, o que representa um bom resultado.

Selecionamos, do conjunto total da amostra, todas as frases que tinham o verbo *dar* ou uma de suas variantes *standard* (*dar, aplicar, conceder, fazer*) ou conversa (*ter, receber, levar e tomar*) e um nome predicativo e que tinham sido anotadas com a mesma etiqueta por, pelo menos, 3 anotadores. Nesse sentido, foram selecionadas 580 frases

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

(22% da amostra geral) para comporem uma subamostra e serem, posteriormente, analisadas na STRING.

A Tabela 3 apresenta a distribuição das frases dessa subamostra, por nível de concordância entre os anotadores.

Tabela 3 – Distribuição de frases por nível de concordância entre anotadores, na subamostra.

Concordância	Nº de frases	% de frases
3 anotadores	95	16,4%
4 anotadores	137	23,6%
5 anotadores	337	58,1%
*frases com erro	11	1,9%
TOTAL	580	100%

Fonte: Elaboração própria.

Na Tabela 3, não indicamos as frases com concordância entre 2 anotadores porque essas sentenças não foram selecionadas para a subamostra. Como se pode perceber, a distribuição das frases por grau de concordância entre anotadores da subamostra (Tabela 3) é praticamente proporcional à sua distribuição na amostra global (Tabela 2).

Pode-se analisar também a distribuição das frases por categoria (ou etiqueta) atribuída pela maioria dos anotadores. A Tabela 4 mostra a distribuição de frases da subamostra, discriminando as porcentagens por categoria.

Tabela 4 – Quantidade de frases por grau de concordância e por categoria.

	3 anotadores	4 anotadores	5 anotadores	TOTAL	%
STANDARD	42	78	227	347	59,82%
CONVERSE	35	38	77	150	25,86%
OTHER	15	21	33	69	11,89%
VOPC	3	0	0	3	0,51%
*frases com erro	---	---	---	11	1,89%
TOTAL por %	95	137	337	580	100%

Fonte: Elaboração própria.

Como se nota, a categoria de SUPPORT[vsup-standard] é a mais consensual entre os anotadores, correspondendo a quase 60%, seguida da categoria SUPPORT[vsup-converse], com cerca de 25%. Os casos restantes somam-se 14,29%.

Integração dos dados na STRING

Seguindo a estratégia delineada em trabalho anterior (RASSI *et al.*, 2015), foi construído um conjunto de programas que convertem automaticamente as informações constantes na matriz do L-G (das construções com *dar*) em regras de extração de dependências que o XIP usa para determinar as relações sintáticas entre o *Vsup* e o *Npred*. Recorde-se que se consideram dois tipos de relação de dependência: `SUPPORT[vsup-standard]` e `SUPPORT[vsup-converse]`. Essas regras abrangem igualmente os casos com o verbo-operador causativo (dependência `VOP-CAUSE`), contudo, nas construções com *dar*, essa categoria praticamente não ocorre (apenas 3 casos), pelo que não a referiremos aqui. Assim, por exemplo, com base na informação da entrada de *beijo*, o sistema gera a seguinte regra (Fig. 1):

Figura 1 – Primeira regra de extração da dependência `SUPPORT[vsup-standard]` para o *Npred beijo*.

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

Fonte: Elaboração própria.

As regras de dependência são constituídas essencialmente de duas partes: num primeiro momento, enunciam-se as condições (`if`) que se têm de verificar para que seja extraída uma dependência; em seguida, explicita-se a dependência a extrair. Neste caso (Fig. 1), a regra determina primeiro uma variável (`#2`) cujo lema é o do verbo-suporte (`[lemma:"dar"]`); em seguida, um conjunto de condições em alternativa (`||`) são enunciadas:

- a primeira condição (linha 2 da Fig. 1) corresponde à situação em que a *CVS* sofre uma relativização, o que é capturado pela dependência `MOD[post,relat]` entre o verbo-suporte (no caso, um participípio passado) e o nome predicativo que é antecedente do pronome relativo;
- a segunda condição (linha 3) diz que a regra dispara se o nome *beijo* estiver numa relação de complemento direto (cuja dependência é `CDIR`) do verbo *dar*, e este não tiver sido marcado com o traço `[transf-passiva]`, que é introduzido previamente quando uma estrutura passiva é identificada;
- a terceira condição (linha 4) é oposta à anterior, identificando uma relação de sujeito (cuja dependência é `SUBJ`) entre o nome predicativo e o verbo-suporte, quando este se encontra na construção passiva;

- a condição seguinte (linha 5) aplica-se quando o sujeito da construção passiva tem como antecedente o pronome relativo (#4 [pronrel]), o qual é então sujeito do verbo-suporte empregue na passiva.
- Finalmente (na linha 6), a regra verifica se ainda não foi extraída a dependência de SUPPORT [vsup-standard] entre *beijo* e *dar*.

Se alguma dessas condições se verificar, a regra dispara e a dependência SUPPORT [vsup-standard] é extraída.

Além dessa regra, a STRING gera também regras para o caso de construções passivas reduzidas (como em *O beijo dado por Rui à Ana*), em que o *Vsup* está apagado:

Figura 2 – Segunda regra de extração da dependência SUPPORT [vsup-standard] para o *Npred beijo*.

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

Fonte: Elaboração própria.

Essa regra difere da anterior pela declaração de contexto, assinalada entre barras verticais (|), na primeira linha. Essa regra acrescenta ao participio passado de *dar* o traço de passiva com *ser* (pass-ser=+). Esse participio deve estar modificando o nome predicativo. Essa regra dispara apenas se a anterior não tiver disparado.

As duas regras anteriores servem para extrair a dependência de SUPPORT [vsup-standard] das *CVS*. Nos casos em que a *CVS* admite conversão, informação que está codificada na matriz do L-G, a STRING também gera as regras de extração da dependência SUPPORT [vsup-converse] (Fig. 3) para capturar frases como *A Ana recebeu um beijo do Rui*.

Figura 3 – Primeira e segunda regras de extração da dependência SUPPORT [vsup-converse] para o *Npred beijo*.

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

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

Fonte: Elaboração própria.

A diferença entre essas duas regras (Fig. 3) e as anteriores (Fig. 1 e 2) é basicamente o lema do verbo, que passa a ser *receber* e, conseqüentemente, a dependência passa a ser SUPPORT [vsup-converse] em vez de SUPPORT [vsup-standard]. Para mais informações sobre o funcionamento das regras de dependência do XIP, veja-se Mamede *et al.* (2012).

Avaliação

A subamostra das frases (com *dar* e suas variantes) selecionada a partir do conjunto de *CVS* manualmente anotadas foi processada na cadeia STRING e a saída do sistema foi analisada por comparação com o *corpus* de referência. Os resultados são apresentados na Tabela 5:

Tabela 5 – Primeira avaliação da performance do sistema⁷.

VP	FP	FN	VN	Precisão	Cobertura	Acurácia	Medida-F
350	91	114	25	79%	75%	65%	77%

Fonte: Elaboração própria.

Utilizam-se na avaliação as métricas habituais de *precisão*, que mede a fração de instâncias relevantes encontradas corretamente ($VP/(VP+FP)$), *cobertura*, que mede a fração de instâncias relevantes que foram encontradas ($VP/(VP+FN)$), *acurácia*, que calcula a combinação da precisão com os erros sistemáticos ($(VP+VN)/(VP+VN+FP+FN)$) e *medida-F*, que é a média harmônica entre a precisão e a cobertura ($2PR/(P+R)$).

Das 580 frases analisadas, a STRING extraiu corretamente a dependência de 350 frases (VP), extraiu incorretamente a dependência de 91 frases (FP) e não extraiu dependência alguma de outras 139 frases, das quais 114 deveriam ter sido extraídas (FN) e 25, de fato, não deveriam ter sido extraídas (VN).

Além desses resultados, o sistema capturou outras 47 dependências que não tinham sido anotadas na referência, já que envolvem pares de palavras que não eram o alvo da frase extraída do *corpus*, como por exemplo: *O varejo, em contrapartida, pode dar descontos no valor cobrado à indústria por determinado espaço na loja.* O par-alvo a ser anotado era {*dar, valor*}, porém, neste caso, não há qualquer relação entre *dar* e *valor*, pelo que os anotadores não o assinalaram. Por outro lado, o verbo *dar* é suporte de *desconto*, o que a STRING capturou bem. Como a dependência (*dar, desconto*) não estava na referência, o *corpus* de referência foi corrigido, acrescentando-lhe as dependências em falta, para serem levadas em conta em um segundo momento da avaliação.

⁷ Considere-se VP = verdadeiros-positivos, FP = falsos-positivos, FN = falsos-negativos e VN = verdadeiros-negativos

A seguir, apresenta-se a análise dos principais problemas identificados na saída do sistema. Após proceder à análise, corrigimos os dados do Léxico-Gramática e processamos o *corpus* novamente. Os resultados da segunda avaliação serão apresentados posteriormente.

Análise de erros

a) Falsos-Positivos

Os Falsos-Positivos (FP) correspondem aos casos em que: (i) o sistema extraiu a dependência errada, como nos casos de ambiguidade entre *SUPPORT*[*vsup-standard*] e *SUPPORT*[*vsup-converse*], ou (ii) extraiu a dependência entre um par de palavras que não possui relação de suporte, por problemas de processamento sintático ou de desambiguação morfossintática. Os dois casos serão analisados em pormenores:

(i) Ambiguidade entre *CVS-standard* e *CVS-conversa*

Tipicamente, o verbo-suporte *dar* é selecionado para formar construções *standard* (de orientação ativa). Por outro lado, o verbo-suporte *receber* é selecionado pelos mesmos nomes predicativos para formar construções tipicamente conversas (de orientação passiva); daí o *Vsup dar* ser considerado um verbo-suporte *standard* e o *Vsup receber* ser considerado um verbo-suporte converso. Há, no entanto, outros verbos - como *ter*, por exemplo - que tanto podem entrar em construções *standard* quanto em construções conversas.

(1)⁸ [**Exemplo construído**]: *O Jô Soares (deu + teve) sua participação no Programa da Hebe.*

(2) [**Exemplo construído**]: *O Programa da Hebe (recebeu + teve) a participação do Jô Soares.*

A primeira frase é tipicamente uma construção *standard*, ao passo que a segunda frase é tipicamente uma construção conversas. Ambas podem ser formadas pelo verbo *ter* e o mesmo nome predicativo *participação*. Portanto, o verbo *ter* estava codificado na matriz do L-G como uma variante *standard* de *dar* e também como uma variante conversas de *receber*. Decidiu-se adotar sistematicamente a dependência

⁸ A maioria dos exemplos apresentados neste artigo foram retirados da subamostra do *corpus*, ou seja, são textos reais. Em situações específicas, utilizamos exemplos construídos/fabricados, seguindo os moldes do L-G, para tornar certos fenômenos mais evidentes. Os exemplos construídos serão antecedidos da notação [**Exemplo construído**]. Nos casos de simplificação de exemplos reais, usaremos a notação [**Exemplo simplificado**]. Já os exemplos reais retirados do *corpus* serão antecedidos do par-alvo que foi anotado manualmente (*Vsup, Npred*).

SUPPORT [vsup-standard], em detrimento da SUPPORT [vsup-converse], nos casos de ambiguidade de classificação, em que o verbo tanto pode funcionar como suporte *standard* quando como suporte converso de um mesmo *Npred*. Isso é feito no XIP por uma regra de “limpeza” que remove as dependências duplicadas.

Essa decisão conduziu a alguns erros de classificação, como por exemplo nos casos apresentados a seguir, que foram etiquetados pela STRING como SUPPORT [vsup-standard], e foram marcados pela maioria ou totalidade dos anotadores como SUPPORT [vsup-converse].

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

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

(5) (**ter, voto**) *O PMDB conta com 5 integrantes, mas terá um voto a menos se Juvêncio estiver na presidência.*
SUPPORT[vsup-standard](voto,terá)

(6) (**ter, prejuízo**) *Pará deve ter prejuízo com jogo do Brasil.*
SUPPORT[vsup-standard](prejuízo,ter)

Por outro lado, há também os casos que a STRING classificou como SUPPORT [vsup-converse], ao passo que a maioria dos anotadores humanos considerou como SUPPORT [vsup-standard]. É o caso de construções com o *Vsup ter* e os seguintes nomes predicativos: *acordo, alvará, apelido, apresentação, cargo, conhecimento, destino, dica, explicação, financiamento, importância, informação, início, liberdade, limitação, motivo, nome, nota, orientação, ponto, prioridade, privilégio, redução, renda, sinal, título e treinamento.*

Esses nomes predicativos, associados ao *Vsup ter*, ao mesmo tempo em que podem formar construções *standard* (7), também podem ser o resultado de uma conversão de outra construção com o verbo-suporte *standard dar* (8):

(7) [Exemplo construído]: *A Ana tem um vasto conhecimento sobre geografia.*

(8) [Exemplo construído]: *A Ana deu conhecimento neste documento.*

[Conversão] = *Este documento teve o conhecimento da Ana.*

Como se pode perceber, (7) e (8) são construções *standard* diferentes; trata-se de dois nomes predicativos diferentes: uma se refere a uma qualidade ou capacidade intelectual, e a outra construção se refere ao ato de assinar, visar um documento. Por serem construções diferentes, a única que está alistada na matriz do L-G utilizada neste trabalho é a construção com *dar*, cuja conversão se faz com *ter*. A construção (7) também é uma construção de base, mas não com o verbo *dar*, por isso, deverá ser descrita em outra matriz que leve em consideração as construções nominais de base com o *Vsup ter*.

Quando ambas as construções estão disponíveis na matriz, temos os casos de ambiguidade, que faz com que duas regras disparem. Nesses casos, atua a regra de “limpeza” referida anteriormente.

Há ainda outras frases que também foram marcadas pela STRING como SUPPORT [v_{sup}-standard], e que foram etiquetadas pela maioria ou totalidade dos anotadores como construção com verbo-operador causativo (VOP-CAUSE):

(10) (**dar, sorte**) *Colocar roupa branca e pular sete ondas dão sorte porque são rituais para atrair coisas boas e, se você acredita, funcionam.* “.

(11) (**dar, prejuízo**) *Fraude on-line dá prejuízo de R\$ 100 mi.*

(12) (**dar, voto**) *Motivo: administra o orçamento para construção de casas populares, que é polpudo e dá votos.*

As construções do tipo *dar sorte*, *dar prejuízo* e *dar voto* são naturalmente aceitáveis, em outras situações, como construções *standard*, daí terem sido assim classificadas na matriz do L-G:

(13) [**Exemplo construído**]: *A Ana deu sorte na loteria.*

(14) [**Exemplo construído**]: *A empresa da Ana deu prejuízo durante todo o ano.*

(15) [**Exemplo construído**]: *A Ana deu seu voto para o candidato da oposição.*

Nesses casos, a identificação dos papéis temáticos dos argumentos poderia auxiliar na desambiguação das regras de dependência (*standard* ou *converso*) a extrair. Em (10), se o sujeito (*colocar roupa branca e pular sete ondas*) fosse corretamente etiquetado com o papel temático de causa, isso seria um indicativo de que a construção é causativa, e não uma *CVS*. Já o exemplo (11) é ambíguo, apesar de ter sido anotado pela maioria dos anotadores humanos como uma construção causativa. Ele é ambíguo porque permite duas interpretações diferentes: (i) a fraude on-line teve um prejuízo; ou (ii) a fraude on-line causou um prejuízo a alguém. Em (12), o sujeito de *dá votos* é o *orçamento*, o que significa que o orçamento é a causa que faz com que alguém tenha votos.

Apesar de os papéis semânticos terem sido codificados na matriz do L-G, essas informações não foram usadas no processo de extração da dependência SUPPORT porque o módulo de extração automática de papéis temáticos (ou papéis semânticos) da STRING não apresenta ainda bons resultados.

(ii) Problemas de processamento sintático ou de desambiguação morfossintática

Algumas frases foram incorretamente etiquetadas pela STRING como SUPPORT [vsup-standard], por problemas de processamento sintático. Esperava-se que o sistema reconhecesse uma relação de dependência entre um verbo e um nome, e ele reconheceu a relação entre outro verbo e/ou outro nome, como em:

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

SUPPORT [vsup-standard] (aulas,dá)

SUPPORT [vsup-standard] (interpretação,dá)

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

SUPPORT [vsup-standard] (vantagem,teria)

SUPPORT [vsup-standard] (rentabilidade,teria)

O par-alvo, cuja dependência deveria ter sido extraída é (*dar, aula*) e (*ter, vantagem*), respectivamente em (16) e (17). Além de extrair corretamente essas duas dependências, o sistema reconheceu também a dependência entre *dá* e *interpretação* (16) e entre *teria* e *rentabilidade* (17). Nesses casos, há um problema de coordenação de grupos nominais e de extração apropriada da dependência de CDIR. A cadeia analisou (16) como uma coordenação entre *aulas* e *interpretação*, e não entre *voz* e *interpretação*, considerando-se a coordenação entre dois complementos diretos do verbo *dar*: *ela dá aulas de voz e ela dá interpretação*. Da mesma forma, a cadeia analisou (17) como uma coordenação entre *vantagem* e *rentabilidade*, e não entre *rapidez* e *rentabilidade*, considerando-se a coordenação entre dois complementos diretos do verbo *ter*: *essa solução teria a vantagem de rapidez e essa solução teria rentabilidade*.

Em outros casos, o problema resulta de uma incorreta atribuição das etiquetas de categorias gramaticais ou da sua inadequada desambiguação. A STRING também extraiu incorretamente a dependência de SUPPORT [vsup-standard] de frases como:

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

SUPPORT [vsup-standard] (volta,dá)

Nessa frase, a forma diminutiva de *saída* (*saidinha*) não foi reconhecida pelo sistema, pelo que, embora lhe tenha sido atribuída a categoria nome, não foi possível extrair a dependência de SUPPORT, a qual requer a identificação do lema. Por outro lado, a desambiguação de *volta* não foi feita adequadamente e a palavra foi etiquetada

como nome, quando deveria ter sido escolhida a etiqueta verbo. Ora, como se verifica a conjunção aditiva (*e*) entre *saidinha* e *volta*, e este último foi marcado como nome, então o sistema analisou essa sequência como a coordenação de dois nomes. No segundo momento, a dependência de complemento direto (CDIR) entre *dar* e *saidinha* é percolada para o nome coordenado com este último, o que desencadeia a regra que extrai a dependência SUPPORT entre *dar* e *volta*.

Também consideram-se problemas de processamento os casos em que a CVS é parcialmente idêntica a uma expressão fixa e a cadeia extrai duas dependências para os mesmos constituintes. A STRING possui um módulo de análise das expressões fixas (BAPTISTA *et al.*, 2014), que contém alguns milhares de expressões idiomáticas e as regras de extração da dependência FIXED que capturam o verbo e os elementos fixos da construção.

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

SUPPORT [vsup-standard](volta,dando)

FIXED(dando,volta,cima)

Nessa frase, a expressão *dar a volta por cima* foi analisada pela STRING de duas formas diferentes, e foram extraídas duas dependências diferentes para os mesmos constituintes: (i) como SUPPORT [vsup-standard], semelhante ao predicado *dar um passeio*; e (ii) como uma construção fixa (FIXED). Recorde-se que a maioria dos anotadores atribuiu, para essa frase, a etiqueta OTHER, que pode corresponder a uma expressão fixa.

Para corrigir esses problemas, foi criada uma regra geral de “limpeza” que dá preferência à extração da dependência FIXED e exclui a dependência SUPPORT. No final do processamento, deverá permanecer apenas a segunda dependência (como expressão fixa), que é a análise correta.

Outras expressões fixas foram analisadas incorretamente tanto pelo sistema STRING quanto pela maioria dos anotadores.

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

SUPPORT[vsup-standard](tiro,dando)

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

SUPPORT[vsup-standard](passo,deu)

Em (20), os elementos da CVS (*dar* e *tiro*) formam um subconjunto da expressão fixa *dar tiro no pé*. Em (21), a expressão fixa *dar um passo à frente* é completamente

ambígua com a *CVS dar um passo*. Do ponto de vista semântico, ambas podem significar literalmente “mover uma perna em direção à frente” ou metaforicamente “seguir adiante, superar algum desafio”. O mesmo problema ocorre com outras expressões, tais como *dar o primeiro passo, dar um passo decisivo, dar passos firmes etc.*, em que o nome *passo* está, provavelmente, na origem dessas construções idiomáticas.

Tanto a anotação manual quanto a classificação automática desses casos deverão ser revistas, a fim de manter a consistência da descrição linguística. As expressões fixas do Português Brasileiro foram descritas por Vale (2001) e muitas delas já foram classificadas também no Português Europeu (BAPTISTA et al., 2004) e inseridas no léxico da STRING (BAPTISTA et al., 2014).

Outro caso de falso-positivo diz respeito à classificação de construções com verbo-operador de ligação como se fossem construções com verbo-suporte. Os dois exemplos que se seguem são casos de construções com verbo-operador de ligação:

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

SUPPORT [vsup-standard](nomes,tiveram)

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

SUPPORT [vsup-standard](nome,tiver)

Nos dois casos, o *Npred nome* foi identificado como complemento direto (CDIR) do verbo *ter*, pelo que dispararam as duas regras de extração da dependência SUPPORT [vsup-standard=+] e SUPPORT [vsup-converse=+]. Por causa da regra de seleção da dependência *standard* nos casos ambíguos, o sistema extrai apenas a dependência SUPPORT [vsup-standard=+]. Esses casos, porém, correspondem a construções com o verbo-operador de ligação, conforme foi anotado manualmente pela maioria dos anotadores.

O verbo-operador de ligação é um conceito introduzido por Gross (1981) e, posteriormente, estudado por Ranchhod (1990), para o Português Europeu, para se referir a verbos que operam sobre construções de base, acrescentando-lhes um argumento o qual já se encontra presente na frase de base na posição de complemento.

(24) [*Exemplo construído*]: *O Rui tem # A Ana está sob o controle do Rui.*

(24a) *O Rui_i tem a Ana sob o (seu_i + *meu + *teu) controle.*

Há alguns casos que haviam sido classificados tanto pelos humanos quanto pela STRING, na primeira avaliação, como SUPPORT [vsup-standard]. Após uma revisão sistemática das anotações do *corpus* de referência, porém, consideramos que trata-se de outros casos de verbo-operador de ligação. Em geral, o problema está

relacionado a nomes como *notícia*, *orientação*, *informação*, *explicação*, *opinião*, *solução*, *resposta*, *exemplo*, *definição*, *dica*, *pista*, *sugestão*, *argumento*, *parecer etc.*, associados ao *Vsup ter*, que admitem duas interpretações diferentes, uma de sentido passivo, em (25), e outra de sentido ativo, em (26), como se observa em:

(25) [Exemplo construído]: *Zé teve uma notícia ruim <quando Ana lhe contou sobre sua doença>*.

(26) [Exemplo construído]: *Zé tem uma notícia ruim <para dar à Ana>*.

Os tempos verbais, em rigor o aspecto “pontual” do pretérito perfeito, em (25), e o aspecto “durativo” do presente (ou do imperfeito), em (26), permitem distinguir esses dois empregos. O exemplo (25) é menos controverso, sendo claramente considerado como uma *CVS* conversa, já que tem como contraparte a construção *standard*:

(25a) \equiv *A Ana deu uma notícia ruim ao Zé*.

O estatuto dessa construção conversa não gerou dúvidas entre os anotadores. Em contrapartida, o mesmo par (*ter*, *notícia*), em (26), parece ter um estatuto especial, pois se assemelha a uma construção de orientação ativa (e.g. *Ana deu uma notícia ruim ao Zé*), mas a ação não chega a se concretizar (aspecto imperfectivo).

O predicado de base em (26) é *dar uma notícia*, já que pode ser reconstituído na oração infinitiva introduzida por *para* (e.g. *Zé tem uma notícia ruim para dar à Ana*). O verbo *ter*, nesse sentido, serve apenas para ligar o argumento (*Zé*) ao predicado *dar uma notícia*. Esse argumento (*Zé*) não é novo, ele já existia na frase de base. Nessas condições, o verbo *ter*, em (26), também tem o estatuto de um verbo-operador de ligação.

O mesmo fenômeno pode ser observado em vários outros *Npred* associados ao verbo *ter*. As frases de (27) a (32), retiradas do *corpus* de referência, exemplificam o problema.

(27) (*ter*, *notícia*) *Segundo Zeca, o Estado vizinho de Mato Grosso tem “quase uma dezena de usinas instaladas na bacia do Paraguai sem que se tenha tido notícia de um único acidente ambiental”.*

SUPPORT [vsup-standard] (*notícia,tenha*)

(28) (*ter*, *informação*) *A delegada diz que é importante que os passageiros que sejam furtados ou roubados registrem a ocorrência na delegacia do aeroporto, para que a polícia tenha mais informações sobre o modo como os bandidos agem.*

SUPPORT [vsup-standard] (*informações,tenha*)

(29) (*ter*, *solução*) *A disputa entre juízes e a direção da liga, que aparentemente teria uma solução rápida, deve durar algumas rodadas.*

SUPPORT [vsup-standard] (*solução,teria*)

(30) (**ter, notícia**) O “The Wall Street Journal” **tem** boas **notícias** para todos vocês, ratos de sofá.

SUPPORT [vsup-standard] (notícias,tem)

(31) (**ter, informação**) A página **tem informações** sobre o clube, fotos e os nomes dos membros.

SUPPORT [vsup-standard] (informações,tem)

(32) (**ter, solução**) Quem ousaria dizer que **tem a solução** para o caso?

SUPPORT [vsup-standard] (solução,tem)

Para todos esses casos, a STRING extrai a dependência SUPPORT [vsup-standard]. Os três primeiros exemplos (27), (28) e (29), no entanto, são CVS conversas e os três últimos (30), (31) e (32) não deveriam ter sido extraídos, já que correspondem a construções com verbo-operador de ligação. Como o próprio *corpus* de referência também estava incorreto em relação a esses casos, ele foi posteriormente revisado e corrigido.

b) Falsos-Negativos

Conforme explicitado na Tabela 5, não foi extraída a dependência de SUPPORT em 114 frases. As principais causas da não-identificação da dependência pretendida é a distância existente entre o verbo-suporte e o nome predicativo ou falha em alguma etapa anterior da cadeia de processamento.

(33) (**dar, mostra**) Para fortalecer essa espécie de revolução democrática que, iniciada com a decisão de desalojar um usurpador do poder Executivo, **dá mostras** agora de que deseja ir fundo na moralização e na republicanização do Poder Legislativo.

O predicado nominal *dar mostras* encontra-se numa oração relativa restritiva, mas o pronome relativo *que* encontra-se separado do verbo-suporte por uma oração encaixada (oração reduzida de participio). Dada a complexidade da frase e a interação das regras da gramática, o *parser* analisa *mostras* como sujeito e não como complemento direto (CDIR) de *dar*. Por essa razão, a dependência SUPPORT não foi extraída. Note-se que, numa frase mais simples, a análise já é adequada:

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

SUPPORT [vsup-standard] (mostras,dá)

Em outros casos, o sistema não extraiu adequadamente a dependência SUPPORT, devido a falha no processamento de algum tipo de anáfora, que ocorre em uma etapa da cadeia anterior à etapa de extração da dependência SUPPORT:

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

O *Npred declaração* é o sujeito de uma construção passiva que se encontra subordinada à oração principal. Contudo, o sujeito (*já que essa declaração não foi dada*) encontra-se elidido, pois já ocorre como sujeito da oração principal. Apesar de a STRING conter um módulo de resolução desse tipo de anáforas (PEREIRA; ZAC, 2010), neste caso o sistema não conseguiu capturar adequadamente o sujeito elíptico e, por isso, não extraiu a dependência SUPPORT da construção com verbo-suporte. No entanto, o sistema captura adequadamente a construção nominal passiva em uma frase cujo sujeito esteja explícito:

(34a) [**Exemplo simplificado**] *A declaração não foi dada em um depoimento formal.*
SUPPORT [vsup-standard] (declaração,dada)

Os 114 predicados nominais foram testados individualmente na cadeia de processamento, utilizando-se frases simples como exemplos (exemplos simplificados). Para todos eles, foi possível obter a análise adequada, o que significa que o problema não está nos dados linguísticos do L-G, mas resulta do complexo processo de análise do *parser*.

Em alguns casos em que a conjunção coordenativa não está explícita, tendo sido substituída por uma vírgula, o sistema não conseguiu extrair a dependência:

(35) (**dar, amasso**) *Ah, Lorena, você só dá uns beijinhos nele, uns amassos e pronto.*
SUPPORT [vsup-standard] (beijinhos,dá)

Nessa frase, apenas a dependência entre *dar* e *beijinhos* foi extraída. A dependência entre o par-alvo (*dar, amasso*) não foi identificada pela STRING. Acontece que, nesse exemplo, a palavra *pronto* foi analisada como um adjetivo e não está em um contexto formal que permita formar um grupo nominal que ficaria então coordenado com os nomes *beijinhos* e *amassos*, pelo que as regras de coordenação não disparam; por isso, também não foi extraída a coordenação entre *beijinhos* e *amassos*.

A coordenação, tal como está implementada na STRING neste momento, é tratada como um fenômeno estritamente local, ligando grupos nominais e/ou preposicionais, incluindo os casos de enumeração de 3 ou mais elementos, em que há elipse das conjunções intermédias (*e.g. laranjas, bananas e maçãs*). Se explicitarmos a conjunção coordenativa da frase (35) entre os grupos nominais *uns beijinhos* e *uns amassos*, então o sistema extrai as duas dependências corretamente:

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

SUPPORT [vsup-standard] (amassos,dá)

SUPPORT [vsup-standard] (beijinhos,dá)

Note-se que, em outros casos de coordenação entre nomes predicativos com o mesmo verbo-suporte, a STRING extraiu a dependência de SUPPORT corretamente, como em:

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

SUPPORT [vsup-converse] (aprovação,tiveram)

SUPPORT [vsup-converse] (leitura,tiveram)

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

SUPPORT [vsup-converse] (confirmação,recebeu)

SUPPORT [vsup-converse] (resposta,recebeu)

Nos dois exemplos, as dependências de complemento direto (CDIR) que foram extraídas entre {ter, leitura} e {receber, resposta} foram percoladas para os outros grupos nominais com que estes nomes se encontram coordenados, nomeadamente {ter, aprovação} e {receber, confirmação}, respectivamente.

c) Verdadeiros-Negativos

Dos 139 casos em que a STRING não extraiu dependência SUPPORT, há 25 frases das quais o sistema, de fato, não deveria ter extraído a dependência, pois não há relação sintática entre o *Vsup* e o *Npred*. Os anotadores humanos também não consideraram que, nessas frases, houvesse uma construção com verbo-suporte; são, portanto, verdadeiros-negativos.

(38) *(receber, título)* Por exemplo, no documento dizia que foi **recebido** a **título** de horas extras CR\$ 200 mil.

(39) *(levar, ponto)* Roteiro de um dia **leva** aos **pontos** altos de San Francisco.

(40) *(ter, sorte)* Sua **sorte** foi **ter** sido socorrido com rapidez.

(41) *(ter, aula)* Quando estiver pronto será a sede da administração do campus e também **terá** salas de **aula**.

Como dissemos acima, a extração das frases a partir do *corpus* PLN.Br apenas se baseou no fato de o par *Vsup Npred* se encontrar presente na frase e não levou em conta quaisquer relações sintáticas. Por essa razão, tais frases selecionadas para a amostra constituem contra-exemplos que contribuem para a aferição da qualidade do processamento.

Casos particulares de Verdadeiros-Positivos

A seção anterior apresentou os principais problemas de extração automática da dependência SUPPORT nas frases retiradas do *corpus*. Além daqueles, deve-se destacar ainda o conjunto de dependências que, mesmo não sendo o alvo pelo qual a frase fora extraída do *corpus*, ainda assim o sistema detectou uma relação de SUPPORT.

Nos exemplos abaixo, indica-se em negrito tanto o par-alvo quanto o par extraído pela STRING. O par-alvo (extraído automaticamente do *corpus*, usando Unitex) está inserido no início do exemplo, entre parênteses; já o par {*Vsup,Npred*}, analisado adequadamente pela STRING, encontra-se sublinhado no corpo do exemplo.

(42) (**dar, show**) *Os ingressos custam R\$ 30,00 e darão direito a diversos **shows** de dança e música e a um jantar típico com especialidades da China, Japão, Coréia, Tailândia, Indonésia e Índia.*

SUPPORT [*vsup-standard*] (direito,darão)

(43) (**dar, nome**) *Segundo Vassoureiro, esse costume, que ainda persiste, deu origem ao nome „papangu“.*

SUPPORT [*vsup-standard*] (origem,deu)

(44) (**ter, nó**) *Pollack - Os europeus têm as mesmas **informações** que nós temos.*

SUPPORT [*vsup-converse*] (informações,têm)

SUPPORT [*vsup-converse*] (informações,temos)

(45) (**receber, comissão**) *Sua candidatura ao COI recebeu o aval da **comissão** executiva da entidade, que é formada por 11 pessoas, entre elas o presidente, Juan Antonio Samaranch.*

SUPPORT [*vsup-converse*] (aval,recebeu)

Como a STRING executa um processamento sintático tanto superficial (*chunking*) quanto profundo (com extração de dependências entre constituintes), o sistema reconhece os constituintes que, de fato, possuem alguma relação e ignora os que não possuem.

Ressalte-se ainda que, em (44), a cadeia extraiu corretamente duas dependências: uma em que *informações* é complemento direto de *têm*, e outra em que *informações* é o antecedente do pronome relativo *que*, o qual é complemento direto de *temos*.

Por meio da integração dos dados do L-G na STRING e da análise sintática automática, também foi possível identificar outros pares de *Vsup* e *Npred* que não haviam sido considerados anteriormente, como, por exemplo:

(46) (*ter, condição*) Como o MEC não **tem condições** de fiscalizar todos os 5.506 municípios brasileiros, pretende **contar com a ajuda** dos Estados.

SUPPORT [vsup-converse] (ajuda, contar)

Além do par-alvo (*ter, condição*) que permitiu que essa frase fosse extraída automaticamente do *corpus* com recurso ao Unitex, a STRING identificou também o par {*contar com, ajuda*} para o qual extraiu corretamente a dependência SUPPORT [vsup-converse].

Segunda avaliação da performance do sistema

Após a análise de erros, os dados linguísticos do Léxico-Gramática foram corrigidos na matriz e as frases da subamostra foram processados novamente.

A Tabela 6 apresenta e compara os resultados da primeira com a segunda avaliação.

Tabela 6 – Resultado da primeira e segunda avaliações da performance da STRING.

	TP	FP	FN	TN	Precisão	Cobertura	Acurácia	Medida-F
Primeira avaliação	350	91	114	25	79%	75%	65%	77%
Segunda avaliação	325	56	84	115	85%	87%	80%	86%

Fonte: Elaboração própria.

Como se nota, a performance do sistema melhorou na segunda avaliação: a precisão aumentou em 6%, a cobertura aumentou em 12%, a acurácia subiu 15% e, consequentemente, a medida-F também subiu 9%.

Vale destacar o aumento significativo no número de verdadeiros-negativos da primeira (TN=25) para a segunda avaliação (TN=115). A primeira avaliação considerou como *golden standard* a anotação da maioria ou unanimidade dos anotadores, sem verificar se aquela anotação estava consistente ou não. Ao revisar sistematicamente as anotações, identificamos todos os casos em que o verbo *ter* funciona como verbo-operador de ligação e corrigimos esses dados no *corpus* de referência, por isso aumentou significativamente o número de frases em que a STRING acertadamente não extraiu suas dependências de SUPPORT.

A melhoria na performance do sistema se deve também à correção dos dados do L-G, pois as dependências que estavam sendo extraídas como SUPPORT [vsup-converse] passaram a ser SUPPORT [vsup-standard].

Além de corrigir o *corpus* de referência e os dados linguísticos do L-G, inserimos, no dicionário (PB.dic) usado pela STRING, as informações de flexão/derivação de grau dos nomes terminados em *-ada/-ida*. As dependências de nomes como *saidinha*, *fugidinha* e *arrumadinha* não estavam sendo extraídas porque o sistema não reconhecia os nomes no diminutivo como formas flexionadas/derivadas dos nomes predicativos *sáida*, *fugida* e *arrumada*, respectivamente. Ao associar o paradigma de flexão/derivação a esses nomes, a STRING passou a extrair corretamente suas dependências.

Ademais, foi criada uma regra de “limpeza” para os casos de expressões fixas. A STRING extraía, ao mesmo tempo, as dependências de *FIXED* e de *SUPPORT* para as construções cujos constituintes ora são fixos ora formam uma *CVS*. A regra criada para a cadeia dá preferência à extração da dependência *FIXED* e exclui a dependência *SUPPORT*, nos casos de duplicação de dependências entre os mesmos constituintes.

Conclusão e trabalhos futuros

Como resultado desse trabalho, produzimos um *golden standard* de construções com o verbo-suporte *dar* e suas variantes para o Português. Esse *golden standard* consiste em um *corpus* anotado automaticamente com as dependências *SUPPORT*[*vsup-standard*] e *SUPPORT*[*vsup-converse*] pela STRING e revisado manualmente por uma equipe de linguistas.

Os resultados da tarefa indicam ganhos na diferente parametrização das regras. Ressalva-se, no entanto, que os experimentos foram feitos para um pequeno conjunto de *CVS*, envolvendo apenas um *Vsup* elementar e suas variantes.

Num futuro próximo, pretende-se integrar também à STRING as matrizes do Léxico-Gramática referentes às construções nominais com os *Vsup* *fazer* e *ter* e avaliar, de forma mais abrangente, a performance do sistema, utilizando na íntegra as 2.646 frases anotadas manualmente.

Uma das especificidades do Português Brasileiro em relação ao Português Europeu é a grande produtividade de nomes predicativos que podem ser criados por derivação com o sufixo *-ada/ida*. Praticamente todos os verbos de ação e muitos verbos de processo podem dar origem a nomes predicativos com o sufixo *-ada/ida* (SCHER, 2004), os quais selecionam prioritariamente o verbo-suporte *dar* (como *abandar* = *dar uma abanada*, *enxugar* = *dar uma enxugada*, *enrugar* = *dar uma enrugada*, *crescer* = *dar uma crescida*, *sumir* = *dar uma sumida* etc.). Da mesma forma, alguns nomes de objetos, instrumentos e nomes parte-do-corpo também podem ser acrescidos do sufixo *-ada* para formar nomes predicativos (BAPTISTA, 2004) (*bater com uma cadeira* = *dar uma cadeirada*, *bater com o cinzeiro* = *dar uma cinzeirada*, *bater com o ombro* = *dar uma ombrada* etc.). Esse fenômeno é bastante produtivo nas construções nominais com o *Vsup dar*. Nesse sentido, pretende-se, em trabalhos futuros, expandir a lista dos nomes predicativos e integrá-los aos dicionários da STRING para que mais construções possam ser identificadas em textos reais.

Contudo, esses *N-ada* são muitas vezes ambíguos com verbos no particípio passado, o que levanta diversos problemas de processamento, em especial pela dificuldade na desambiguação de *part-of-speech* (POS).

Agradecimentos

Os autores agradecem a contribuição de Cláudia Dias de Barros e Maria Cristina Andrade dos Santos, na tarefa de anotação do *corpus*, bem como pela disponibilização de seus dados, e ainda ao apoio financeiro da Capes - Coordenação de Apoio à Pesquisa, sob o processo BEX 12751/13-8 e do Fundo Nacional da FCT - Fundação para a Ciência e a Tecnologia, sob o projeto PEst-OE/EEI/LA0021/2013.

RASSI, A.; BAPTISTA, J.; VALE, O.; MAMEDE, N. Integration of nominal predicates in parser: an experiment in constructions with the support verb 'give' in Brazilian Portuguese. *Alfa*, São Paulo, v.62, n.3, p.543-571, 2018.

- *ABSTRACT: This paper describes the methodology for the integration of nominal predicates, like support verb constructions (SVC), in XIP parser, which is used by the processing chain for Portuguese named STRING. It presents, more specifically, 580 SVC with the verb ,give', which were manually annotated and then integrated into the rule-based XIP grammar in order to extract the SUPPORT dependency between the predicative noun (Npred) and the support verb (Vsup). It is necessary to analyze automatically SVC because they are different from full verb constructions, they have complex syntactic structures, specific syntactic-semantic properties and they admit systematic syntactic transformations, although lexically determined. The concept of SVC and the lexical-syntactic approach adopted follow the theoretical and methodological principles of Lexicon-Grammar. As a result of integration of such data to the XIP parser, the system achieved 85% precision, 87% recall, and 80% accuracy and 86% F-measure.*
- *KEYWORDS: Support verb. Predicative noun. Support verb constructions. Causative operator verb. XIP parser.*

REFERÊNCIAS

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

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

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

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

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

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

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

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

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

CALZOLARI, N.; FILLMORE, C. J.; GRISHMAN, R.; IDE, N.; LENCI, A.; MACLEOD, C.; ZAMPOLLI, A. Towards best practice for Multiword Expressions in Computational Lexicons. In: **Third International Conference on Language Resources and Evaluation**, LREC. Las Palmas, Canary Islands – Spain, May, 2002. p.1934-1940.

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

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

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

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

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

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

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

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

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

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

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

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

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

MAMEDE, N.; BAPTISTA, J.; DINIZ, C.; CABARRÃO, V. String: an hybrid statistical and rule-based natural language processing chain for Portuguese. **International Conference on Computational Processing of Portuguese (Propor 2012), Demo Session**. Coimbra, Portugal, April, 2012.

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

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

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

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

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

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

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

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

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

SAG, I. A.; BALDWIN, T.; BOND, F.; COPESTAKE, A. A.; FLICKINGER, D. Multiword Expressions: A Pain in the Neck for NLP. In: GELBUKH, A. (Ed.) **Proceedings of the Third International Conference**, CICLing - Computational Linguistics and Intelligent Text Processing. Mexico City, Mexico, February 17-23, 2002. p.1-15.

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

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

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

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

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

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

VIVÈS, R. **Avoir, prendre, perdre**: Constructions à verbe support et extensions aspectuelles. 1983. 388 f. Tese (Thèse de Troisième cycle), Laboratoire Automatique Documentaire et Linguistique, Université Paris 8, Paris, 1983.

Recebido em 1 de novembro de 2017

Aprovado em 7 de maio de 2018

