

***ViPEr*: uma base de dados de construções léxico-sintáticas de verbos do Português Europeu ¹**

Jorge Baptista

Universidade do Algarve/CECL & INESC ID Lisboa/L2F
jbaptis@ualg.pt

a Christian Leclère

Abstract

This paper presents ViPEr, the lexicon-grammar of European Portuguese verbs, a database with the main relevant syntactic and semantic information on the construction of about 5, 000 most frequent verbs, namely, the number and type of their arguments, their distributional constraints, the most salient transformational properties, coreference constraints and mode selection for argument sub-clauses, and, finally, some reproducible semantic features (semantic roles and polarity). The automatic annotation followed by the manual disambiguation of a corpus with 250, 000 words (39, 000 verbs) made possible to verify that ViPEr shows a good lexical coverage in view of several Natural Language Processing applications requiring verb sense disambiguation.

1. Introdução

O desenvolvimento de aplicações em processamento computacional de linguagem natural (PLN) requer a elaboração de bases de dados linguísticos de grande cobertura e elevada granularidade. Apresentamos nesta comunicação o *ViPEr*², uma base de dados das construções léxico-sintáticas dos verbos mais frequentes do Português Europeu. O quadro teórico de referência é o do Léxico-Gramática (Gross, 1975, 1996; Lamiroy (ed.) 1998), fundado nos princípios da gramática transformacional de operadores de Harris (1991).

Neste quadro teórico, a unidade de análise é a *frase elementar*, que, numa primeira aproximação, pode ser definida como a construção sintática determinada por um predicado semântico na sua extensão máxima, isto é, composta pelo sujeito, o verbo e todos os complementos essenciais expressos. Considera-se que este é o contexto sintático adequado para determinar as propriedades sintáticas e semânticas das construções de um elemento predicativo.

¹ A investigação foi parcialmente suportada pela FCT/MCTES, através do projeto REAP.PT (ref. CMU-Portugal/HuMach/53/2008). Uma versão preliminar deste trabalho foi apresentada em Baptista (2012). O autor agradece os comentários dos avaliadores, que permitiram melhorar a versão final deste artigo.

² <https://string.l2f.inesc-id.pt/w/index.php/Dictionaries#ViPEr>.

No caso deste estudo, trata-se de determinar o conjunto das construções sintáticas associadas a cada verbo. Considera-se que, no caso dos verbos que apresentam diferentes significados, quando estes são claramente distintos e, sobretudo, quando essa distinção se apoia em propriedades formais, reproduzíveis, cada significado corresponde a uma construção sintática distinta, constituindo uma entrada léxico-sintática autónoma. O léxico-gramática é, pois, uma representação formal das unidades de sentido – neste caso, dos predicados verbais – e das suas propriedades.

Tanto quanto sabemos, embora estejam disponíveis várias descrições sintáticas de construções verbais do Português Europeu, nem sempre estas foram realizadas tendo em vista a sua aplicação em processamento computacional da língua, ou foram disponibilizadas à comunidade do PLN. Assim, por exemplo, refira-se os dicionários de Fernandes (2008), Borba (1991) e Busse (1994), ou os estudos linguísticos parciais, desenvolvidos sobretudo ao longo das últimas três décadas, como por exemplo Oliveira (1981), Chaby (1997) ou Rodrigues (1997). Estes dados, mesmo quando suficientemente formalizados para poderem integrar sistemas de PLN, pouco foram utilizados em aplicações concretas. Por outro lado, baseando-se na aplicação massiva de métodos introspectivos (Laporte, 2010), raramente foram feitas tentativas para a sua validação em *corpora* (uma tentativa recente nessa direção é o trabalho de Gomes, 2011).

É ainda de referir o léxico do projeto LE-PAROLE³, cuja camada sintática é descrita com algum pormenor em Marrafa *et al.* (1999), e cujo formalismo, em SGML, se baseia no sistema de codificação dos projetos GENELEX e EAGLES. O léxico é constituído por 20.000 entradas, das quais 3.000 verbos, contendo informações sobre a respetiva estrutura e distribuição. Apesar de se reconhecer “a importância das alternâncias [=transformações] para a determinação da relação entre o significado e a sua relação sintáctica”, devido “às características [d]a ferramenta de gestão utilizada”, que “não permite o tratamento automático das relações” transformacionais (alternâncias), apenas se indicaram a ordem básica dos constituintes de cada construção, enquanto “os diferentes contextos de realização dos argumentos são expressos e associados ‘manualmente’” (idem: 203). Os autores alertam para o facto de adotarem uma perspetiva não derivacional para o tratamento das “alternâncias”, por outras palavras, uma perspetiva não-transformacional, diversa da, pela nossa parte, que seguimos neste estudo.

³ Disponível em: http://catalog.elra.info/product_info.php?products_id=713 [2013-07-12].

O objetivo deste estudo é o de fornecer um inventário sistemático das principais construções sintáticas dos verbos do Português Europeu, identificando as respetivas propriedades formais (estruturais, distribucionais e transformacionais).

1.1. Seleção do vocabulário

Partindo dos dados de frequência do corpus CETEMPúblico⁴ (Rocha e Santos, 2000) depois de processado pela cadeia de processamento STRING (Mamede *et al.*, 2012)⁵, procedeu-se à descrição e classificação dos 5.052 verbos plenos mais frequentes (5 ou mais ocorrências). As razões para esta seleção prendem-se, por um lado, com o objetivo de utilizar esta base de dados em diversas aplicações de PLN, em que a cobertura lexical é um fator determinante; e, por outro lado, com a limitação decorrente da necessidade de recorrer à introspeção para distinguir e descrever os diferentes significados e respectivas construções sintáticas dos verbos, contexto em que apenas os usos frequentes e bem conhecidos podem ser adequadamente descritos por forma a garantir a reprodutibilidade das observações.

Excluíram-se as construções de auxiliares (Baptista *et al.*, 2010) e as construções com verbo-suporte (Ranchhod, 1990; Baptista, 2005), bem como as construções fixas idiomáticas (M. Gross, 1982; G. Gross, 1996; Vale, 2001; Baptista *et al.*, 2004).

1.2. Propriedades representadas

Para cada verbo, indicou-se de forma sistemática, entre outras propriedades:

(i) o número e tipo de argumentos (distinguindo, nomeadamente as construções completivas e não-completivas), considerando apenas actantes do predicado verbal;

(ii) a natureza distribucional dos diferentes argumentos (humano/não-humano, locativo; especificaram-se certas classes de objetos como nomes parte-do-corpo, meios de transporte, plantas, animais, vestuário, etc. sempre que relevantes);

(iii) as propriedades transformacionais consideradas mais relevantes: simetria, passiva com *ser* e/ou *estar*, passiva pronominal e construção neutra;

(iv) restrições de correferência e de seleção de modo em construções completivas;

(v) algumas informações semânticas reprodutíveis (*e.g.* natureza do complemento locativo: *origem/destino/trajeto* (§2.6) ; papel semântico do complemento dativo: *destinatário/fonte/interlocutor*; a polaridade *positiva/negativa* do predicado).

⁴ www.linguateca.pt/cetempublico

⁵ <https://string.l2f.inesc-id.pt>

1.3. Critérios de classificação

Para a classificação, baseamo-nos nas propostas do Léxico-Gramática (veja-se a síntese de Leclère, 2002), tendo sido constituídas cerca de 60 classes léxico-sintáticas (v. apêndice).

Os diferentes sentidos de um mesmo verbo, quando claramente reprodutíveis e associados a propriedades estruturais, distribucionais e/ou transformacionais, foram sistematicamente desdobrados em diferentes entradas. Assim, por exemplo, para um verbo como *apontar*, consideraram-se as seguintes construções⁶:

*O Pedro **apontou** à Joana os defeitos que devia corrigir (9)*

*O bandido **apontou** uma faca ao polícia (36DT)*

*O Pedro **apontou** os números premiados num papel (38LD)*

*O Pedro **apontou** o João como sério candidato ao prémio (39)*

Foi assim possível determinar cerca de 6.224 construções distintas. Um ou mais exemplos ilustram na base de dados cada construção.

O resto deste artigo está organizado do seguinte modo: Na secção §2 descrevemos com algum pormenor os critérios de classificação utilizados e as principais classes de construções verbais assim constituídas. Na secção §3 apresentamos sucintamente os resultados do processo de classificação e uma primeira avaliação deste novo recurso linguístico. Concluimos, em §4, apresentando perspectivas de trabalho de investigação futuro.

2. Princípios gerais de classificação

Nesta secção, apresentamos de forma muito sucinta os principais princípios de classificação aplicados na construção do ViPEr, um léxico-gramática dos verbos do Português Europeu. Por razões de espaço, algumas noções são apenas referidas de forma breve, remetendo-se o leitor para as referências relevantes. Pelo caminho, ilustraremos as diferentes classes assim constituídas.

2.1. Frase elementar

O principal princípio taxonómico adoptado é o conceito de *frase elementar* (ou frase simples), definida como a construção sintática determinada por um predicado semântico com todos os seus argumentos (sujeito e complementos essenciais) expressos. Tal definição exclui

⁶ Os códigos alfanuméricos à esquerda dos exemplos designam as classes de construções verbais (ver apêndice).

todas as construções que envolvem adjuntos e modificadores. O termo *elementar* é aqui entendido num sentido preciso: trata-se da construção que resulta da primeira restrição harrissiana (Harris, 1991) que constitui o *kernel* da gramática: a ordem parcial de entrada das palavras na frase. Esta restrição estabelece uma dicotomia fundamental entre, por um lado, os operadores, ou seja, palavras que requerem outras palavras para construir uma frase; e, por outro lado, os argumentos elementares, isto é, palavras que não podem ser elas mesmas operadores. Neste sentido, considera-se que os operadores de segunda ordem, ou seja, os operadores que requerem que pelo menos um dos seus argumentos seja ele próprio um operador (*e.g. dizer*) fazem parte da base da gramática (*kernel*) ao mesmo título que uma palavra que apenas selecione argumentos elementares (*e.g. cair*). Por essa razão, as construções completivas são integradas na base da gramática, sendo assim consideradas também como frases elementares. Dada a sua importância, falaremos destas construções numa secção à parte (§2.6).

2.2. Projeção máxima

Cada frase elementar é descrita considerando o princípio da projeção máxima do predicado, isto é, com todos os argumentos expressos, incluindo todos os complementos essenciais. De um modo geral, os predicados verbais apresentam entre um e três argumentos. Exceções a esta generalidade, em virtude do seu significado particular, são: (i) as construções meteorológicas e designativas de partes do dia (*e.g. nevar, amanhecer*; classe **31I**) e outras construções impessoais⁷; (ii) as construções com predicados de transferência (*e.g. importar, transferir*; **38LT**) e mais algumas outras (apostar) que não constituem conjuntos semanticamente homogêneos e que são pouco numerosas, podendo ser definidas em extensão.

Os complementos obrigatórios são sempre considerados complementos essenciais. Contudo, como muitos complementos podem ser frequentemente reduzidos, nem sempre é fácil determinar que complementos devam ser considerados como essenciais à construção de um dado predicado. De um modo geral, apenas uma análise caso a caso permite tomar essa decisão, para o que nos socorremos, com frequência, de exemplos do corpus. Por outras palavras, se um verbo ocorrer frequentemente com um determinado tipo de complemento, consideramos que este faz parte da respetiva construção.

Por outro lado, enquanto muitos complementos circunstanciais são frequentemente adjuntos adverbiais, certos verbos podem selecionar esse tipo de complementação, que deve,

⁷ A construção impessoal do verbo *tratar-se de*, *e.g. Trata-se de isso/fazer isso*, é a única encontrada que apresenta um complemento. Encontra-se provisoriamente integrada na classe **02**.

pois, ser considerada como elemento essencial da construção verbal. É o caso, por exemplo, dos complementos locativos de verbos como *viver* (**35LS**), ou os complementos de modo de verbos como *comportar-se* (**33MV**).

A aplicação de certas transformações sintáticas dá origem a estruturas complexas, que podem suscitar problemas de classificação. Assim, por exemplo, a análise por reestruturação de grupo nominal de construções com nomes parte-do-corpo (Guillet & Leclère, 1992; Leclère, 1995; Baptista, 1997) dá origem a diferentes configurações sintáticas, que não representam, no ViPEr diferentes construções, mas foram tratadas como outras tantas expressões formais do mesmo predicado semântico (classe **32CL**):

O cão mordeu as canelas do carteiro (**32CL**)

= *O cão mordeu as canelas ao carteiro*

= *O cão mordeu o carteiro nas canelas.*

Do mesmo modo, certas construções intransitivas com nomes parte-do-corpo (**31CL**) admitem algumas destas operações:

Os pés do Pedro suam imenso (**31CL**)

= *O Pedro sua imenso dos pés.*

Ns construções neutras (Guillet & Leclère, 1992), isto é, dos verbos diateticamente neutros relativamente à transitividade ou intransitividade do predicado, e em que o sujeito da construção transitiva tem um papel semântico de **causa**, deu-se preferência a esta estrutura mais longa (**32C**)⁸, conquanto a estrutura intransitiva parece representar o predicado semântico numa forma mais simples (isto é, despojado do valor causal). Efetivamente, preferimos analisar a construção transitiva como resultado da fusão de um verbo-operador causativo (*Vop*; M. Gross 1981) com a construção intransitiva:

Os doces fizeram # O Pedro engordou

= *Os doces fizeram o Pedro engordar*

= *Os doces fizeram engordar o Pedro*

= *Os doces engordaram o Pedro* (**32C**)

As formas com sujeito causal sem complemento podem ser derivadas da construção transitiva:

Os doces engordam (toda a gente).

⁸ Tanto a natureza do sujeito causativo como a existência da construção intransitiva estão representadas nas propriedades desta construção.

Finalmente, nas construções intrinsecamente pronominais, *e.g.* :

O Pedro suicidou-se (31H)

O Pedro dá-se com a Ana (35S)

O Pedro queixou-se da Ana à polícia (41)

considerou-se que o pronome reflexo não representa um constituinte autónomo, já que não pode ser reduzido, *e.g.*:

**O Pedro suicidou*

°*O Pedro deu com a Ana*⁹

**O Pedro queixou da Ana à polícia*

nem pode ser substituído por um grupo nominal da mesma natureza distribucional mas não correferente ao sujeito, *e.g.*

**O Pedro suicidou o João*

**O Pedro deu o João com a Ana*

**O Pedro queixou o João da Ana à polícia*

Em alguns casos, frases com diferentes construções da estrutura básica poderão ser aceitáveis mas com uma interpretação irónica (°*O Pedro suicidou o João*), mas este tipo de fenómeno discursivo não foi considerado na classificação, já que corresponde à expressão da criatividade linguística dos falantes, derivada de (e só interpretável por referência a) a construção básica do verbo. Em contrapartida, certos usos figurativos, já completamente generalizados e em que por vezes os falantes nem consciência têm da sua origem metafórica, levaram ao desdobramento do verbo em duas ou mais entradas léxico-sintáticas distintas:

O Pedro ralou o queijo (32C)

Os miúdos ralavam o Pedro (4).

2.3. Complementos preposicionais

A preposição que introduz o complemento foi considerada um critério taxonómico importante, já que cada construção apresenta geralmente apenas uma das preposições básicas, nomeadamente: *telefonar a*, (33; Rodrigues, 1997); *gostar de* (8), *confiar em* (8), *casar com* (35S), *ansiar por* (8), etc.

Nas construções locativas, as preposições locativas foram colapsadas sob a notação *Loc* e o papel semântico do complemento locativo explicitamente indicado nas propriedades da

⁹ A frase poderia ser aceitável mas com um significado diferente (°), tratando-se, portanto, de uma construção autónoma da construção intrinsecamente pronominal.

respetiva entrada lexical. Outras preposições, além das preposições básicas acima, são igualmente indicadas de forma explícita na base de dados.

Um caso especial de construção preposicional consiste nas chamadas construções simétricas (Baptista, 2005), frequentemente empregando a preposição *com*:

O Pedro conversou com o João (35S),

O Pedro confundiu o João com o Miguel (36S1),

O Pedro debateu esse assunto com o João (36S2),

O Pedro concordou com o João em fazer isso (42S).

Este tipo de construção é definido pela relação semântica particular que se estabelece entre dois dos argumentos e o verbo, que permite que os argumentos possam permutar entre si ou serem coordenados numa posição sintática sem que o significado da frase se altere:

O Pedro conversou com o João = O João conversou com o Pedro

Quando os dois argumentos se encontram coordenados, estas construções permitem ainda a inserção facultativa de uma cópia pronominal:

O Pedro e o João conversaram (um com o outro),

o que as distingue das construções recíprocas, que preferimos analisar como resultado da redução de frases coordenadas:

*O Pedro conta com o João <para fazer isso> e o João conta o Pedro <para fazer isso>
= O Pedro e o João contam um com o outro <para fazer isso>.*

Neste último exemplo, a cópia pronominal *um com o outro* não pode ser reduzida,

*°! *O Pedro e o João contam,*

sem que o significado da frase se altere ou a frase se torne inaceitável.

Um caso mais raro de variação da preposição consiste na possibilidade de certos verbos permitirem a redução daquela, sem que o significado da frase se altere:

O Pedro namora com a Sara (35S)

= O Pedro namora a Sara

Nestes casos, foi dada preferência à construção sintática mais longa, com preposição, tendo-se assinalado a propriedade de redução facultativa da preposição. Note-se que, na construção transitiva direta, a formação da passiva não é possível:

**A Sara é namorada pelo Pedro,*

o que parece confirmar ser a construção preposicional a forma de base.

2.4. Propriedades distribucionais

A seguir às propriedades estruturais da construção (número e tipo de complemento, preposições que introduzem os complementos), as restrições distribucionais impostas pelo verbo ao preenchimento lexical das diversas posições argumentais constituem o segundo critério taxonómico mais importante na classificação.

Os principais critérios distribucionais utilizados foram os seguintes:

- oposição **nome humano/não-humano** (*Hum/nHum*). Certas classes são definidas pela natureza obrigatoriamente humana de um ou mais dos argumentos do verbo, como é caso de *falecer* (**31H**), *violar* (**32H**), *irritar* (**4**) e *agradar* (**5**). Estas construções têm prioridade na classificação dos verbos em que a oposição humano/não-humano não é relevante (*morrer*, **31R**) ou em que apenas um nome não-humano é admitido (*orbitar*, **32R**). Os nomes de animais foram tratados como nomes não-humanos, não tendo sido levados em consideração na classificação, excepto em certos verbos especializados como o caso das vozes de animais (*coaxar*, **31R**).

- **nome plural** (*Npl*). Trata-se de uma noção abstrata, que procura capturar os casos em que o verbo impõe um argumento de natureza conceptualmente plural a uma dada posição sintática, tanto como sujeito (*abundar*, **31PL**) como no complemento (*dispersar*, *coleccionar*, **32PL**), como ainda numa posição de complemento preposicional (*dividir*, **38PL**; Nascimento, 1997). Note-se que estas construções podem admitir nestas posições não apenas um plural morfológico (*Os peixes abundam no mar*), como um coletivo (*A multidão dispersou*) ou até um singular com valor genérico (*A truta abunda no rio*).

- **nomes parte-do-corpo** (*Npc*). Este conceito pretende ajudar a descrever predicados que selecionam para uma dada posição sintática um nome que designa uma parte do corpo e que, por essa razão, implicam uma relação metonímica intrínseca ou de posseção inalienável com um nome (geralmente de tipo humano); este tipo de construção pode ocorrer tanto na posição de sujeito (*suar*, **31CL**) como de complemento direto (*pentear*, **32CL**) e dá origem, como já vimos atrás, a certas reestruturações do grupo nominal complexo com *Npc* selecionado pelo verbo (Baptista, 1997, 2000).

- **nomes locativos** (*Nloc*), ou seja, nomes envolvidos na expressão de diferentes tipos de predicados locativos. Dada a complexidade (e relevância) do sistema léxico-gramatical dos verbos locativos, trataremos dele numa secção à parte (§2.6).

- **objecto restrito** (**R**). Certas construções verbais, sem que possam ser consideradas construções idiomáticas, apresentam tão fortes restrições ao preenchimento lexical de algumas das suas posições sintáticas que os conceitos atrás são manifestamente insuficientes para dar

conta dessa distribuição particular (Dias *et al.*, 2006; Leclère, 2000). Trata-se, na maior parte dos casos de construções transitivas diretas em que o complemento direto é fortemente restrito a um pequeno conjunto de nomes, frequentemente apenas um:

O Pedro estrelou uns ovos (32R).

Neste exemplo, apenas *ovo* pode ocorrer como complemento de um verbo como *estrelar*; o significado global da frase é claramente composicional, isto é, pode ser derivado do significado do verbo (processo culinário) e do objeto, e este pode apresentar variação livre em número ou quanto à sua determinação. Este tipo de restrição é bastante mais difícil de descrever, pelo que foi constituída uma classe residual, que, à medida que o trabalho de levantamento e classificação for avançando, deverá ser progressivamente completada.

2.5. Propriedades transformacionais

Algumas propriedades transformacionais foram consideradas definitórias de classes de construções verbais. É o caso das construções simétricas (ver acima) ou das construções transitivas diretas sem passiva, geralmente associadas à expressão de grandezas mensuráveis (Baptista e Ranchhod; 1998):

O portátil pesa 1, 2 Kg (32NM)

**1, 2 Kg são pesados pelo portátil*

No mesmo sentido, foi isolado um conjunto de construções causativas, associadas a construções adjetivais através da operação de *Fusão* (M. Gross, 1981):

O Pedro clarificou a sua posição (32TA)

= O Pedro tornou (mais) clara a sua posição

Além destas propriedades, foi sistematicamente descrita para as construções transitivas a possibilidade de formação da passiva com auxiliar *estar* (e variantes), a construção passiva pronominal e a construção passiva neutra (v. acima):

O Pedro / O vento partiu a janela (32C)

A janela foi partida pelo Pedro; A janela (está, ficou) partida

A janela partiu-se <com o vento>

O Pedro / O vírus avariou o computador (32C)

O computador (está, ficou) avariado

O computador avariou-se; O computador avariou

Outras transformações deverão ser, no futuro, consideradas.

2.6. Sistema de construções completivas

As construções completivas, na medida em que são estruturalmente mais complexas, têm prioridade na classificação sobre as construções não completivas. Frequentemente, os verbos admitem que uma dada posição, a completiva comute com um grupo nominal, o que torna por vezes mais difícil determinar se se trata de uma construção completiva ou não. Quando esse grupo nominal tem como cabeça um nome predicativo (Ranchhod, 1990; Baptista, 2005), apresenta uma interpretação proposicional, facto que leva a integrar o verbo numa classe de construções completivas:

O Pedro criticou que o João tivesse tido esse comportamento (6)

= O Pedro criticou o João pelo comportamento que ele tinha tido

= O Pedro criticou o comportamento do João

O mesmo princípio foi utilizado nas construções ditas transitivas-predicativas, associadas à redução de construções adjetivais/atributivas:

O Pedro considera que o João é muito inteligente/um bom pai (6)

= O Pedro considera o João muito inteligente/um bom pai

O estudo deste tipo de relações, em particular as reduções a que pode ser sujeita a completiva deverá prosseguir noutro momento. Noutros casos, a alternância entre nome humano/não-humano (ver abaixo) e completiva não modifica um valor semântico **causal** associado a essa posição sintática, como sucede em:

= O Pedro/essa notícia entristeceu o João

= Que o Pedro tivesse feito isso entristeceu o João (4)

As construções completivas seguem, no mais, os critérios gerais de classificação, nomeadamente o número e tipo de complementos, as preposições que os introduzem e a natureza distribucional de alguns complementos. São de salientar as classes de construções completivas seguintes, que representam a maioria destas construções:

- os verbos **psicológicos** (Oliveira, 1981; classe **4**, exemplo acima)
- os verbos de **percepção mental** e **volitivos** (classe **6**, exemplo acima), e
- os verbos **declarativos** (e.g. *dizer*, classe **9**; Baptista, 2010)

Por uma questão de falta de espaço, remetemos o leitor para o quadro geral de classificação.

2.7. Sistema de construções locativas

O sistema léxico-gramatical dos verbos locativos pode ser organizado num conjunto variado de construções sintático-semânticas, considerando, entre outros aspetos, a oposição entre predicados estativos e dinâmicos. A natureza (ou papel semântico) do complemento

locativo é determinada não só pelo verbo mas também pela preposição locativa que introduz o complemento¹⁰. Assim, os predicados **locativos estativos** correspondem a casos como

O Pedro vive em Lisboa (35LS).

No caso dos predicados **locativos dinâmicos**, consideramos os locativos de **origem**:

O Pedro veio de Lisboa (35LD)

os locativos de **destino**:

O Pedro vai para Lisboa (35LD)

e os locativos de **trajeto**:

O Pedro passou por Lisboa (35LD).

Certos verbos admitem mais do que um tipo de complemento locativo, razão por que os diferentes papéis semânticos que a construção verbal autoriza são explicitamente representados no léxico-gramática. Além das construções preposicionais, acima ilustradas, os verbos que exprimem predicados locativos podem apresentar o *Nloc* num grupo nominal, geralmente em posição de complemento direto:

Os alunos invadiram a sala (locativo de **destino**, **38L1**)

Os deputados abandonaram o hemiciclo (locativo de **origem**, **38L1**)

O Pedro atravessou a praça (locativo de **trajeto**, **38L1**).

Nestes casos, na ausência de preposição locativa, apenas o papel semântico codificado no léxico-gramática permite determinar a relação do *Nloc* do complemento com o verbo.

As construções locativas incluem também os casos dos predicados de transferência de objeto, que foram organizados de acordo com a natureza dinâmica/estativa do processo e o papel semântico predominante do complemento de lugar:

O Pedro colocou o livro (em, sobre) a mesa (locativo de **destino**, **38LD**)

O Pedro retirou o livro da estante (locativo de **origem**, **38LS**)

Nos casos em que não foi possível determinar o papel semântico predominante, constituiu-se uma classe específica:

O Pedro transferiu o dinheiro de uma conta para a outra (**transferência**, **38LT**)

Encontramos ainda um conjunto residual de verbos locativos com objecto que designam predicados estativos:

O Pedro situou aquele local no mapa (locativo **estativo**, **38LR**)

¹⁰ Cada preposição locativa é, portanto, caracterizada quanto à natureza semântica (estativo, destino, origem, trajeto) do complemento locativo que introduzem.

São também casos especiais destas construções com objeto as transitivas locativas que envolvem a noção de *Fusão* (M. Gross, 1981). Assim, encontramos as seguintes construções que podem ser analisadas pela fusão de um operador (*pôr, tirar*) e um predicado locativo:

O Pedro enjaulou o leão (= *O Pedro pôs o leão na jaula*; loc. **destino, 38L2**)

O Pedro envenenou a bebida (= *O Pedro pôs veneno na bebida*; loc. **destino, 38L4**)

O Pedro desengordurou a panela

(= *O Pedro tirou gordura da panela*; loc. **origem, 38L5**)

A estes casos, junta-se uma classe residual de predicados com sujeito locativo estativo:

A jaula encerrava o leão (sujeito loc. **estativo, 38L3**).

3. Avaliação e Resultados

A classificação dos 5.037 lemas diferentes mais frequentes no corpus do CETEMPúblico permitiu a identificação de 6.224 construções sintáticas diferentes, os quais constituem o léxico do ViPEr. Foram igualmente identificados 259 construções de verbo-suporte e verbos auxiliares, que não fazem parte do recurso. Do ponto de vista da ambiguidade, ou seja, número de classes diferentes em que o mesmo verbo pode ser classificado, os verbos distribuem-se pelas seguintes classes de ambiguidade (Tabela 1):

Nº Classes	Nº Lemas	(%)	
1	4.267	(84,71%)	
2	515	(10,22%)	
3	159	(3,16%)	
4	62	(1,23%)	
5	19	(0,38%)	
6	8	(0,16%)	<i>abrir, acabar, adiantar, alinhar, apanhar, arrancar, disparar, ligar</i>
7	2	(0,04%)	<i>bater, levar</i>
8	3	(0,06%)	<i>contar, cortar, servir</i>
10	1	(0,02%)	<i>dar</i>
11	1	(0,02%)	<i>passar</i>
Total	6.224		

Tabela 1. Distribuição dos lemas do ViPEr por classe de ambiguidade

Como se pode verificar, a maioria (84,7%) dos verbos representados no ViPEr pertence apenas a uma classe, ou seja, considerou-se que têm apenas um significado. Tal deve-se ao fato de termos adotado uma abordagem muito conservadora no processo de classificação, tendo-nos limitado a considerar apenas as construções mais comuns ou frequentes dos verbos.

A fim de determinar a sua cobertura lexical, o ViPEr foi aplicado a um subconjunto do corpus LE-PAROLE (Nascimento *et al.*, 1997; Ribeiro, 2004), constituído por textos de natureza variada (sobretudo literários e jornalísticos) e contendo cerca de 250.000 palavras,

dos quais aproximadamente 39.000 são verbos, correspondendo a 2.665 lemas (10.314 formas) diferentes. Destas 39.000 ocorrências, 21.500 constituem verbos plenos, 486 são expressões verbais idiomáticas e 17.012 são construções de auxiliares ou afins: 6.365 verbos copulativos, 5.136 verbos-suporte, 4.608 auxiliares verbais (temporais, modais e aspectuais), 630 construções de verbo-operador e 210 usos do pro-verbo *fazer*.

O corpus foi automaticamente anotado e desambiguado (Diniz *et al.*, 2010; Diniz & Mamede, 2011), utilizando a cadeia de processamento STRING (Mamede *et al.*, 2012). Destes 2.665 lemas de verbos, apenas 22 lemas (>1%) não constavam no ViPEr. Trata-se de: algumas lacunas óbvias (*atear, afiançar, atrever, cobiçar, eriçar*); verbos derivados (*auto-declarar, constitucionalizar, redividir*); certos erros de análise morfológica (*titular* (adj.), *combóio* (verbo)); e, sobretudo, verbos relativamente pouco usuais (*digladiar, embatucar, rechaçar, tardar*), quase todos com frequências bastante baixas. O único verbo que o sistema não reconheceu como tal foi a forma *autopropuseram*, para o qual sugeriu um verbo com o lema **autopropuserir*. A extensão do módulo morfológico (Diniz & Mamede, 2011) da cadeia STRING, entretanto desenvolvida (Vicente, 2012), já permite analisar adequadamente estas formas, atribuindo-lhes as etiquetas ViPEr da forma de base.

Procedeu-se em seguida à desambiguação manual e sistemática dos 573 lemas ambíguos do corpus, correspondendo a um conjunto inicial de 876 construções, tendo-se verificado a necessidade de corrigir algumas construções e acrescentar 134 outras novas, geralmente pouco frequentes, que não tinham sido inicialmente consideradas.

Tais resultados permitem considerar satisfatória a cobertura lexical do ViPEr, tendo em vista diversas aplicações em PLN que requerem a desambiguação e identificação do sentido dos verbos (Travanca, 2012).

Conclusões e trabalho futuro

A cobertura lexical e a granularidade da informação linguística representada no ViPEr torna-o num recurso lexical, tanto quanto sabemos único, no panorama dos trabalhos em linguística computacional no Português Europeu. Num futuro imediato, pretendemos disponibilizar à comunidade este recurso, a fim de que o mesmo possa ser avaliado por terceiros, sobretudo em contexto de aplicações práticas de PLN ou de ensino de Português. Pretendemos, igualmente, comparar o recurso com o léxico do projeto LE-PAROLE e validar os critérios de classificação medindo o acordo entre anotadores.

Entretanto, e em paralelo, a informação constante do *ViPEr* já está a ser utilizada em diversas aplicações e estudos: (1) a fim de desambiguar os pronomes pessoais (*me, te, se, nos e vos*); (2) na desambiguação automática de sentido de verbos do Português Europeu (Travanca, 2013); (3) na definição a atribuição de papéis semânticos a construções verbais (Talhadas, 2012), e na extração de eventos.

Agradecimentos

O autor gostaria de agradecer aos alunos que, de diferentes formas e em diversos momentos contribuíram para este trabalho: Daniel Santos, David Monteiro, Graça Fernandes, Rui Santos, Tiago Travanca, e ainda a Cláudio Diniz (L2F/INESC-ID Lisboa). Um agradecimento especial a Nuno Mamede (IST/INESC ID Lisboa), pelo seu constante apoio e amizade.

Referências

- Baptista, J. (1997) Conversão, nomes parte-do-corpo e reestruturação dativa. in Castro, Ivo (ed.) *Actas do XII Encontro da APL*, Lisboa, 1996, pp. 51-59. Lisboa: APL/Colibri.
- Baptista, J.; Ranchhod, E (s/d) Propriétés des phrases élémentaires associées à l'expression de grandeurs mesurables. Exemples du Portugais. In Klein, J., B. Lamiroy e J.-M. Pierret (Eds.) *Téorie Linguistique et applications informatiques*. Actes du 16ème Colloque Européen sur la Grammaire et le Lexique Comparés des Langues Romanes (Louvain-la-Neuve, 24-27 Septembre 1997). *Cahiers de l'Institut de Linguistique de Louvain* 24 : 3-4 (Numéro special, 183-184) : pp. 49-61. Louvain-la-Neuve: Peeters.
- Baptista, J. (2000) Body-part nouns and local grammars. *Révue d'Informatique et Statistiques en Sciences Humaines*, 36, pp. 53-66. Liège: Univ. Liège.
- Baptista, J. (2005)a *Sintaxe dos predicados nominais com SER DE*. Lisboa: FCG/FCT.
- Baptista, J. (2005)b Construções simétricas: argumentos e complementos. in Figueiredo, O; and Rio-Torto, Graça; and Silva, F. (eds.). *Estudos de Homenagem a Mário Vilela*, pp. 353-367. Porto: Campo das Letras.
- Baptista, J. (2010) *Verba dicendi*: a structure looking for verbs. in Nakamura *et al.* (eds.), pp. 11-20.
- Baptista, J. (2012) *Verb Classification Guidelines*. Technical Report. Lisboa: L2F-Spoken Language Lab/INESC-ID Lisboa.
- Baptista, J. (2013) ViPEr: A Lexicon-Grammar of European Portuguese Verbs, In *Proceedings of the 31st International Conference on Lexis and Grammar*, pp. 10-16, Czech Republic, September 2012.

- Baptista, J.; Correia, Anabela; Fernandes, G. (2004) Frozen Sentences of Portuguese: Formal Descriptions for NLP. *Workshop on Multiword Expressions: Integrating Processing*. Barcelona: ACL.
- Baptista, J.; Mamede, N.; Gomes, F. (2010) Auxiliary verbs and verbal chains in European Portuguese. In: Pardo, T. *et al.* (eds.). *Computational Processing of the Portuguese Language. PROPOR 2010*, LNCS/LNAI 6001, pp. 110-119. Berlin: Springer.
- Boons, J-P.; Guillet, A.; Leclère, Ch. (1976a). *La structure des phrases simples en français: Les constructions intransitives*. Paris: Droz.
- Boons, J-P.; Guillet, A.; Leclère, Ch. (1976b). *La structure des phrases simples en français: Les constructions transitives*. Rapport de Recherches 10. Paris: LADL.
- Borba, F. (1991) *Dicionário Gramatical de Verbos do Português Contemporâneo do Brasil*. São Paulo, Brasil: UNESP.
- Busse, W. (1994) *Dicionário Sintático de Verbos*. Coimbra: Livraria Almedina.
- Chaby, Teresa. (1997) *Construções de segmentação. Propriedades léxico-sintáticas*. MA Thesis). Universidade de Lisboa - Faculdade de Letras.
- Dias, M.; Laporte, E. & Leclère, Ch. (2006) Verbs with very strictly selected complements. In: *Collocations and Idioms: The First Nordic Conference on Syntactic Freezes*. Finland: University of Joensuu.
- Diniz, C; Mamede, N. (2011) *LexMan - Lexical Morphological Analyser*. Technical Report. Lisboa: L2F/INESC-ID Lisboa.
- Diniz, C.; Mamede, N.; Pereira, J. (2010) *RuDriCo2* - a faster disambiguator and segmentation modifier, In *II Simpósio de Informática (INForum)*, pp. 573-584, Braga: Univ. Minho.
- Fernandes, F. (2008) *Dicionário de Verbos e Regimes* (45th ed). São Paulo, Brasil: Globo.
- Gomes, F. (2011) Validation of Lexical-Syntactical Matrices. MSc Dissertation. Instituto Superior Técnico, Universidade Técnica de Lisboa.
- Gross, G. (1996a). *Les expressions figées en français*. Paris: Ophrys.
- Gross, M. (1975) *Méthodes en Syntaxe*. Paris: Hermann.
- Gross, M. (1981) Les bases empiriques de la notion de prédicat sémantique. *Langages*, 63, 7-52.
- Gross, M. (1982) Une classification des phrases «figées» du français. *Revue Québécoise de Linguistique*, 12-2, pp-pp.16
- Gross, M. (1996) Lexicon-Grammar. In Brown, Keith & Miller, J. (eds.), *Concise Encyclopedia of Syntactic Theories*. Cambridge: Pergamon, pp. 244-259.

- Guillet, A.; Leclère, Ch. (1981) Restructuration du group nominal. *Langages* 63, 99–125.
- Guillet, A.; Leclère, Ch. (1992) *La structure des phrases simples en français: 2 - Les constructions transitives locatives*. Paris: Droz.
- Harris, Z. S. (1991) *A Theory of Language and Information. A Mathematical Approach*. Oxford: Clarendon.
- Lamiroy, B. (ed.) (1998) *Le Lexique-Grammaire. Travaux de Linguistique 37*. Bruxelles: Duculot.
- Laporte, E. (2010) Le lexique-grammaire est-il exploitable pour le traitement des langues? in Nakamura *et al.* (eds.): pp. 207-218.
- Leclère, Ch. (2002) Organization of the Lexicon-Grammar of French Verbs. *Linguisticae Investigationes*, 25-1, 29-48.
- Leclère, Ch. (1995) Sur une restructuration dative. *Language Research*, 179–198.
- Leclère, Ch.. (2002) Organization of the Lexicon-Grammar of French Verbs. *Linguisticae Investigationes*, 25-1, 29–48.
- Mamede, N.; Baptista, J.; Diniz, C. (2012) STRING - An Hybrid Statistical and Rule-Based Natural Language Processing Chain for Portuguese. In *PROPOR 2012 – Demo*: <http://www.propor2012.org/demos/DemoSTRING.pdf> (05-09-2012).
- Marrafa, P.; Gonçalves, J.; Mendes, A. (1999) A Sintaxe do LE-PAROLE. in Marrafa, P. e Mota, A. (Orgs.) *Linguística Computacional: Investigação Fundamental e Aplicações*, pp. 191-205. Lisboa: APL/Colibri.
- Nakamura, T.; Laporte, E.; Dister, A.; Fairon, C. (eds.) (2010) *Les tables. La grammaire du français par le menu*. Mélanges en hommage à Christian Leclère. *Cahiers du CENTAL* 6. Louvain-la-Neuve: Presses Universitaires de Louvain.
- Nascimento, M.; Bettencourt, J.; Marrafa, P.; Ribeiro, R.; Veloso, R.; Wittmann, L. (1997) LE- PAROLE – Do corpus à modelização da informação lexical num sistema multi-função. In *Actas do XIII Encontro da APL*, Lisboa, Portugal.
- Oliveira, M.E. (1981) *Syntaxe des verbes psychologiques du Portugais*. Lisboa: INIC.
- Ranchhod, E. (1990) *Syntaxe dos predicados nominais com estar*. Lisboa: INIC.
- Ribeiro, R. (2004) *Anotação Morfossintáctica Desambiguada do Português* (Tese de Mestrado). Lisboa: IST/UTL.
- Rocha, P & Santos, D. (2000 November) CETEMPúblico: Um corpus de grandes dimensões de linguagem jornalística portuguesa. *Actas do V Encontro para o processamento computacional da língua portuguesa escrita e falada, PROPOR'2000*, pp. 131–140.

- Rodrigues, R. (1997) Determinação das propriedades sintáticas das frases simples com a forma N0 V a N1. Master thesis, Universidade de Lisboa - Faculdade de Letras.
- Talhadas, R. (2012) *Automatic Semantic Role Labeling for European Portuguese*. Relatório de Projeto. Faro: U. Algarve.
- Travanca, T. (2013) *Disambiguation of Verb Senses*. Tese de Mestrado. Lisboa: IST/INESC-ID Lisboa.
- Vale, Oto. (2001) *Expressões Cristalizadas do Português do Brasil: Uma Proposta de Tipologia*. Araraquara: UEJMF(PhD Thesis).
- Vicente, A. (2012) *LexMan: um Segmentador e Analisador Morfológico com transdutores*. Relatório de Projeto. Lisboa: IST/INESC-ID Lisboa.

Apêndice – ViPer : Classes de construções verbais do Português Europeu

Classe	#	Estrutura	verbo	exemplo
01I	13	QueF ₀ V	adiantar	Não adianta fazer isso
01T	61	QueF ₀ V QueF ₁	evitar	Fazer isso evitava ter de fazer aquilo
02	15	QueF ₀ V Prep ₁ QueF ₁	obrigar	Isto obriga a fazer aquilo
03	1	N ₀ V N ₁ (Loc ₁ Nloc ₁) Vinj ^A	mandar	O Pedro mandou o João à loja comprar café
04	330	Nnr ₀ V Nhum ₁	irritar	Isso irrita o Pedro
05	24	Nnr ₀ V a Nhum ₁	agradar	Isso agrada ao Pedro
06	221	Nhum ₀ V QueF ₁	pensar	O Pedro pensa que o João é inteligente
07	46	Nhum ₀ V a (Vinj ^B) ₁	aprender	O Pedro aprendeu a fazer isso
08	96	N ₀ V Prep ₁ QueF ₁	depende	O Pedro dependia da autorização do João
09	162	Nhum ₀ V QueF ₁ a Nhum ₂	dizer	O Pedro disse ao João que está feliz
10	4	Nhum ₀ V QueF ₁ Prep ₂ Nhum ₂	apostar	O Pedro apostou com o João que ganhava a corrida
11	43	N ₀ V N ₁ a QueF ₂	obrigar	O Pedro obrigou o João a fazer isso
12	34	N ₀ V N ₁ de ₂ (Vinj ^A) ₂	impedir	O Pedro impediu o João de fazer isso
13	21	N ₀ V N ₁ de ₂ QueF ₂	informar	O Pedro informou o João de que ia fazer isso
14	11	N ₀ V N ₁ Prep ₂ N ₂ (a Nhum ₃)	pagar	O Pedro pagou 20€ por isso ao João
16	9	N ₀ V QueF ₁ Prep ₂ QueF ₂	deduzir	O Pedro deduziu isso daquilo
31CL	20	Nbp ₀ V	suar	Os pés do Pedro suam
31H	287	Nhum ₀ V	espirrar	O Pedro espirrou
31I	24	0 V	chover	Chove
31PL	11	Npl ₀ V	proliferar	As bactérias proliferam
31R	275	(Nhum+Nnhum) ₀ V	morrer	O Pedro morreu
32A	35	N ₀ V Nnhum ₁ {apparition}	preparar	O Pedro preparou o almoço
32C	1.148	N ₀ V Nnhum ₁ * {apparition}	ler	O Pedro leu um livro
32CL	193	N ₀ V Nbp ₁	partir	O Pedro partiu um braço
32CV	13	N ₀ N-v N ₁ [= converter N ₁ em N]	cristalizar	Isso cristalizou o açúcar
32H	440	Nhum ₀ V Nhum ₁	amar	O Pedro ama a Ana
32NM	29	N ₀ V Nmeas ₁	medir	O Pedro mede 1, 80 m
32PL	55	N ₀ V Npl ₁	ordenar	O Pedro ordenou os alunos
32R	243	N ₀ V Nc ₁	estrelar	O Pedro estrelou um ovo
32TA	289	N ₀ Adj-v N ₁ [V=tornar Adj N ₁]	amaciar	O sabonete amacia a pele
33	77	N ₀ V a N ₁	telefonar	O Pedro telefonou ao João
33MV	4	N ₀ V Advmanner	portar-se	O Pedro portou-se mal
33NM	1	N ₀ V Prep ₁ Nmeas ₁	ascender	O PIB ascende a 1B\$
34	53	N ₀ V Prep ₁ N ₁ Prep ₂ N ₂	saber	O Pedro sabe muito de futebol
35LD	273	N ₀ V-dyn Loc ₁ Nloc ₁	entrar	O Pedro entrou na sala
35LS	40	N ₀ V-stat Loc ₁ Nloc ₁	viver	O Pedro vive em Lisboa
35R	173	N ₀ V Prep ₁ N ₁	confiar	O Pedro confia no João
35S	112	N ₀ V com N ₁	conversar	O Pedro conversou com o João
36DT	108	Nhum ₀ V Nobj ₁ a Nhum ₂	dar	O Pedro deu um livro ao João
36R	89	N ₀ V N ₁ Prep ₂ N ₂	transformar	O Pedro transforma barro em arte
36S1	82	N ₀ V Nobj ₁ com ₂ Nobj ₂	misturar	O Pedro mistura o açúcar com a farinha
36S2	15	Nhum ₀ V Nobj ₁ com ₂ Nhum ₂	combinar	O Pedro combinou com o João uma ida ao cinema
36TA	6	N ₀ Adj-v N ₁ Prep ₂ N ₂ [V=tornar Adj N ₁]	adequar	O Pedro adequa o discurso ao público
38L1	193	N ₀ V Nloc ₁	invadir	O Pedro invadiu a sala
38L2	35	N ₀ Nloc-v Nobj ₁ [V=put in Nloc]	enjaular	O Pedro enjaulou o leão
38L3	9	Nloc ₀ V Nobj ₁	encerrar	A jaula encerrava a fera
38L4	86	N ₀ Nobj-v Nloc-d ₁ [V=pôr Nobj]	envenenar	O Pedro envenenou a bebida
38L5	10	N ₀ Nobj-v Nloc-s ₁ [V=tirar Nobj]	desengordurar	O Pedro desengordurou o prato
38LD	73	N ₀ Vdyn N ₁ Loc-d ₂ Nloc ₂	pousar	O Pedro pousou o livro na mesa
38LS	73	N ₀ Vdyn N ₁ Loc-s ₂ Nloc ₂	retirar	O Pedro retirou o livro da mesa
38LT	41	N ₀ Vdyn N ₁ Loc-s ₂ Nloc ₂ Loc-d ₃ Nloc ₃	transferir	O Pedro transferiu o livro daqui para ali
38LR	5	N ₀ Vstat N ₁ Prep ₂ N ₂	situar	O Pedro situou a casa no mapa
38PL	57	N ₀ V N ₁ Prep ₂ Npl ₂	dividir	O Pedro dividiu o bolo em fatias
38R	7	N ₀ V N ₁ Loc-d ₂ N ₂	remeter	O Pedro remeteu o João para a Ana
38TD	9	N ₀ V N ₁ Loc-s ₂ N ₂	receber	O Pedro recebeu uma prenda do João
39	58	N ₀ V N ₁ (Prep ₂) N ₂	nomear	O Pedro nomeou o João (como) seu representante
40	11	N ₀ V Prep ₁ N ₁ Prep ₂ N ₂	dar	O Pedro deu com um livro na cabeça do João
41	11	N ₀ V Prep ₁ N ₁ Prep ₂ QueF ₂	apelar 'appeal'	O Pedro apelou ao João para que fizesse isso
42S	5	N ₀ V com N ₁ Prep ₂ N ₂	comungar	O Pedro comungava com o João dos mesmos ideais
Total: 6.224				

Notações: N₀, N₁, N₂, N₃: sujeito e complementos; Prep: preposição; Adj: adjetivo; Adv: advérbio; Nhum: nome humano; Nnhum: nome não-humano; Npc: nome parte-do-corpo; Npl: nome plural; Nloc: nome locativo; Nnr: nome não-restrito; Nobj: “objecto” (semantic role); QueF: oração completiva; Loc: preposição locativa; V: verbo; Vdyn: verbo locative dinâmico; Vstat: verbo locative estativo; -v: terminação verbal.