

Descrição do verbo cortar para processamento automático de linguagem natural

Wagner Pacheco, Eric Laporte

► **To cite this version:**

Wagner Pacheco, Eric Laporte. Descrição do verbo cortar para processamento automático de linguagem natural. Laporte, Éric ; Smarsaro, Aucione ; Vale, Oto. Dialogar é preciso. Linguística para processamento de línguas, PPGEL/UFES, pp.165-176, 2013, 978-85-8087-104-3. hal-00804811

HAL Id: hal-00804811

<https://hal-upec-upem.archives-ouvertes.fr/hal-00804811>

Submitted on 26 Mar 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Descrição do verbo *cortar* para processamento automático de linguagem natural

Wagner Lúcio Pacheco¹
Éric Laporte²

Resumo: *Este artigo busca apresentar a evolução dos estudos descritivos do comportamento morfo-sintático-semântico do verbo cortar, observando como este se realiza enquanto verbo pleno, verbo suporte ou presente em expressão fixa, como parte do projeto de pesquisa da Prof^aDr^aAucione Smarsaro, intitulado “Descrição de estruturas lexicais do português do Brasil para processamento automático”. Tal descrição se fez a partir da metodologia léxico-gramática definida por Gross (1975) que possibilita resultados sólidos e reprodutíveis usando um corpus constituído de 33 exemplos retirados da nossa experiência de falante ou da Web, e adaptados para ficarem simples, porém ainda aceitáveis. Tal método considera “a frase elementar” como “unidade mínima de sentido” (Laporte, 2000, p. 5) e possibilita a elaboração de fórmulas a partir da representação do sujeito e de complementos essenciais dessas unidades mínimas.*

Palavras chave: descrição, verbo, sintaxe, morfologia, semântica, recenseamento, codificação.

Introdução

O trabalho de descrição do português que tem por objetivo a aplicação em processamento automático de linguagem natural depende da descrição das estruturas linguísticas dessa língua, de forma que a qualidade dessas aplicações se relaciona proporcionalmente à qualidade e à quantidade de estruturas compostas cobertas pela pesquisa. Nesse sentido, esse tipo de pesquisa se constitui como sendo infundável, ao considerar os problemas que tornam necessária uma contínua operação dos dispositivos de informação, seja na elaboração de sistemas mais sofisticados, seja na atualização das pesquisas que sustentam os sistemas. “Visto que os sistemas abertos evoluem na integração entre sistema e ambiente [...] a atualização das linguagens documentárias é uma operação que visa dar sustentação à evolução dos dispositivos de informação” (Kobashi, 2007, p. 7). Os dicionários eletrônicos e inúmeras outras estruturas dependem dessa descrição para ter um bom funcionamento, e é fato que esse trabalho depende de um grande número de pesquisadores que considerem as diversas particularidades da gramática da língua, como também as suas diferentes formas de uso. Toda descrição de estruturas linguísticas, de qualquer categoria de palavra, é importante para o processamento automático de linguagem natural. Os softwares terão mais qualidade quando tivermos uma maior cobertura de descrição lexical do português para esse fim. Com essa compreensão, um estudo descritivo de verbos também se coloca como uma condição para o processamento automático, tendo em vista o reconhecimento de que um verbo pode apresentar diferentes comportamentos no uso da língua. Borba (1996), Neves (1999), Scher (2004), dentre outros, já apresentaram estudos que abordam essa questão.

A formalização de uma tabela e sua publicação, seja para facilitar “a todos e a cada um julgar se as diferentes precauções tomadas desempenharam seu papel, verificando se os resultados obtidos estão de acordo com os julgamentos de aceitabilidade” (Laporte, 2008, p. 9), seja na oferta de dados linguísticos aos engenheiros da computação e programadores para sua utilização em softwares de tradução automática, de busca e recuperação, e

¹ Graduando do curso Licenciatura em Língua Portuguesa e Literatura de Língua Portuguesa na Universidade Federal do Espírito Santo.

² Universidade Federal do Espírito Santo; Université Paris-Est; eric.laporte@univ-paris-est.fr

dicionários eletrônicos, possui “*um interesse científico e técnico de primeira ordem*” (Laporte, 2008, p. 9).

Nesse sentido esta pesquisa contribui para o processamento Automático de Linguagem Natural ao somar-se a outras já realizadas e a realizar.

A definição de classes como aquelas propostas pelo Léxico-Gramática, por exemplo, se fundamenta em descrições semelhantes à nossa, pois as classes são definidas a partir das propriedades.

1. Metodologia

Para esta pesquisa, utilizamos o método do léxico-gramática definido por Gross (1975), que de acordo com Laporte (2008, p. 15) “fornece um exemplo concreto e produtivo de resultados”. Também consideramos a Gramática de Valências de Borba (2007), que assim como Gross, considera que o comportamento de uma estrutura deve ser avaliado num determinado contexto de uso. Também dão suporte ao desenvolvimento desse artigo as pesquisas do linguista francês Éric Laporte, que contatamos através de textos e cursos ministrados na UFES, onde o pesquisador desenvolve estudos teóricos e práticos de descrição linguística.

O corpus é constituído de 33 exemplos, retirados da nossa experiência de falante ou da Web, e adaptados para ficarem simples, porém ainda aceitáveis. O principal método de análise é a aplicação de critérios sintáticos formais: sequências são criadas artificialmente a partir dos exemplos iniciais, por aplicação de operações formais conhecidas; em seguida, a aceitabilidade das sequências é julgada com base da nossa intuição de falante nativo. A partir da aplicação de testes sintáticos formais, tendo como meta a observação da produção de sentido desses verbos em diferentes contextos de uso, e observando sua conceituação na condição de verbo pleno, de verbo suporte e de parte de expressões fixas, os itens lexicais levantados foram reunidos em grupos de exemplos, para codificação de suas propriedades linguísticas e formalização desses dados para fins computacionais. Assim, fez-se uma análise das restrições distribucionais do verbo *cortar*, agrupando os exemplos conforme os itens lexicais observados em 11 grupos. Essas restrições lexicais, sintáticas e semânticas consideraram as posições dos argumentos das frases (Gross, 1975).

2. Propriedades estruturais do verbo *cortar*

A descrição desenvolvida, tendo como princípio norteador o método Léxico-Gramática, buscou inicialmente reconhecer as propriedades estruturais dos verbos, suas características morfológicas, sintáticas e semânticas em diferentes situações de uso da língua portuguesa, levando em consideração as diferentes realizações desse verbo como pleno, suporte ou em expressão fixa. Essas características devem ser observadas no emprego do verbo em diferentes contextos de uso do português, por meio de testes de natureza sintática. As descrições que utilizam critérios formais devem contemplar fatores morfológicos, sintáticos, semânticos e pragmáticos, pois muitas vezes é o uso que define a relação de sentido estabelecida pelo verbo em certas estruturas linguísticas. Os estudos realizados para esta pesquisa mostraram que o verbo *cortar* pode funcionar em seu contexto de uso como verbo pleno, verbo suporte ou como parte de uma expressão fixa.

O Verbo pleno é aquele que funciona como predicado da frase, sendo este o centro de significado do período. Por exemplo, na frase *Pedro cortou o pão com uma faca amolada*. Já os verbos-suporte, de acordo com a teoria de Maurice Gross e a Gramática de Neves (1999), são verbos de significado lexical parcial ou totalmente esvaziado, que formam juntamente com seu complemento um significado global que, muitas vezes, tem

correspondência com um verbo pleno da língua, constituindo-se assim numa classe especial de verbos. Os testes aplicados mostraram que o verbo *cortar* comporta-se raras vezes como suporte, e nesse caso, sempre como variante terminativa de *fazer* ou *dar*:

O governo fez investimentos nessa área.

*O governo **cortou** os investimentos nessa área.*

*João **deu** financiamentos aos estudos do seu filho.*

*João **cortou** o financiamento dos estudos do seu filho.*

Foi possível confirmar também a utilização do verbo *cortar* em expressões fixas, aquelas que, para a manutenção do seu sentido, não permitem a substituição de certos elementos do predicado:

*Jessé **cortou** o coração de Ana*

*Jessé **cortou** o coração de sua amiga.*

*Jessé **cortou** o fígado de Ana³*

3. Codificação das propriedades estruturais do verbo *cortar*

Após a observação do verbo em diferentes contextos de uso, os itens lexicais levantados foram agrupados para codificação de suas propriedades linguísticas e para formalização desses dados, tendo como objetivo sua utilização futura em programas de computador. A codificação das propriedades estruturais do comportamento do verbo analisado foi apresentada neste trabalho em uma tabela de léxico-gramática, que de acordo com Paumier (2002, p. 02) “*são matrizes binárias que descrevem as propriedades de certas palavras. [...] cada palavra tem um comportamento quase único. As tábuas possibilitam estabelecer a gramática de cada palavra*”. Nosso objetivo foi o de utilizar os dados reunidos nessa tabela para formar dados em programas, na forma de dicionário eletrônico.

3.1 Recenseamento das Restrições

Para a codificação das propriedades linguísticas de um verbo e a formalização dos dados assim obtidos, fez-se uma análise das restrições distribucionais do verbo *cortar*, agrupando os exemplos conforme os itens lexicais observados. No recenseamento foram notados 11 grupos, com suas construções de base. Essas restrições lexicais, sintáticas e semânticas levaram em consideração a posição que os argumentos ocupavam nas frases (GROSS, 1975). Utilizaremos alguns símbolos: *Nhum* para sintagmas nominais denotando pessoas, *N-hum* para os que não denotam pessoas, *Nconc* objetos concretos, *Nabs* entidades abstratas, *Npc* partes do corpo, *Npl-obl* sintagmas nominais obrigatoriamente no plural.

Grupo 1

1. João **cortou** os dois últimos parágrafos do texto.

2. A nova diretoria **cortou** os funcionários faltosos da folha de pagamento.

3. O conselho de escola **cortou** os alunos indisciplinados do passeio semestral.

4. João **cortou** o açúcar de sua dieta.

5. Ela **cortou** a inveja da sua vida.

6. Ela **cortou** o cigarro definitivamente.

Distribuição: $N_0 =: Nhum$ $N_1 =: N$ $N_2 =: N-hum$

³ Como pode ser observado, a substituição do substantivo *coração* por *fígado* muda completamente o sentido da frase, mesmo sendo de um grupo nominal que designe parte do corpo humano.

N_0 cortar N_1
 N_0 cortar N_1 de N_2
 N_1 ser cortado de N_2 por N_0
 N_1 ser cortado de N_2
 N_1 ser cortado por N_0
 N_1 ser cortado
Construção de base: N_0 **cortar** N_1 **Loc** N_2

Grupo 2

7. A companhia **cortou** a energia de João.
8. A forte chuva **cortou** a energia da rede elétrica da casa de João.
9. A mãe **cortou** a conversa da filha.
10. A presença da mãe **cortou** o medo da filha.
11. O medo das consequências **cortou** a conversa de Pedro com João.
12. A pedra **cortou** a passagem de água na montanha.

Distribuição: N_0 =: N N_1 =: N -hum

N_0 cortar N_1
 N_1 ser cortado por N_0
 N_1 ser cortado
 N_1 estar cortado por N_0
 N_1 estar cortado
Construção de base: N_0 **cortar** N_1

Grupo 3

13. A travessa **corta** a rua principal.
14. O riacho **cortava** toda a extensão do sítio.
15. Rios secos **cortavam** o solo.
16. O navio **cortou** as águas da Baía.
17. Um arrepio **cortou** sua espinha/ **cortou** a espinha de João.

Distribuição: N_0 =: N conc N_1 =: N -abs

N_0 cortar N_1
 N_1 ser cortado por N_0
 N_1 estar cortado por N_0
 N_1 ser cortado
 N_1 estar cortado
Construção de base: N_0 **cortar** N_1

Grupo 4

18. O Médico **cortou** o braço de João⁴.

Distribuição: N_0 =: N hum N_1 =: N pc

N_0 cortar N_1
 N_1 ser cortado por N_0
 N_1 ser cortado

⁴ Neste caso, alguns nomes aplicados a partes do corpo criam certa ambiguidade, que pode ser facilmente resolvida num contexto mais amplo. Por exemplo, *O fígado/a gordura está cortado de João*. João foi submetido a uma dieta em que foi cortado todo tipo de gordura ou fígado, ou será que sofreu uma cirurgia na qual parte ou todo o fígado/gordura foi retirado?

N_1 estar cortado por N_0

N_1 estar cortado

Construção de base: N_0 **cortar** N_1

Grupo 5

19. João **cortou** uma folha ao meio / João **cortou** uma folha em três partes.

20. A serra **cortou** a árvore ao meio.

21. O Mágico **cortou** a mulher ao meio.

22. João **cortou** uma fatia da pizza.

Distribuição: N_0 =: N -abs N_1 =: N -abs N_2 =: N pl-obl

N_0 cortar N_1 / N_0 cortar N_1 em N_2

N_1 ser cortado por N_0 / N_1 ser cortado por N_0 em N_2

N_1 ser cortado / N_1 ser cortado em N_2

N_1 estar cortado / N_1 estar cortado em N_2

Construção de base N_0 **cortar** N_1 em N_2

Grupo 6

23. João **cortou** as unhas.

24. João **cortou** a grama.

Distribuição: N_0 =: N hum N_1 =: N conc

N_0 cortar N_1

N_1 ser cortado por N_0

N_1 ser cortado

N_1 estar cortado

Construção de base: N_0 **cortar** N_1

Grupo 7

25. João **cortou** o vestido com um molde da revista.

26. A máquina **cortou** o vestido no modelo que Pedro desejava.

Distribuição: N_0 =: N -abs N_1 =: N conc

N_0 cortar N_1

N_1 ser cortado por N_0

N_1 ser cortado

N_1 estar cortado por N_0

N_1 estar cortado

Construção de base: N_0 **cortar** N_1

Grupo 8

27. Pedro **cortou** seu adversário em poucos segundos.

28. O carro de João **cortou** a moto de Pedro com dificuldades.

Distribuição: N_0 =: N -abs N_1 =: N -abs

N_0 cortar N_1

N_1 ser cortado por N_0

N_1 ser cortado

Construção de base: N_0 **cortar** N_1

Grupo 9

29. Àquela notícia **cortou** o coração de João.

30. A saudade **cortou** o coração de Maria.

Distribuição: $N_0 =: N\text{-conc}$ $N_1 =: N\text{hum}$

N_0 cortar o coração de N_1

O coração de N_1 ser cortado por N_0

O coração de N_1 ser cortado

O coração de N_1 estar cortado por N_0

O coração de N_1 estar cortado

Construção de base: N_0 **cortar o coração de** N_1

Grupo 10

31. João **cortou** caminho da igreja até o rio pela floresta.

Distribuição: $N_0 =: N\text{hum}$ $N_1 =: N\text{conc}$ $N_2 =: N\text{conc}$ $N_3 =: N\text{conc}$

N_0 cortar caminho de N_1 até N_2 Loc N_3

N_0 cortar caminho de N_1 até N_2

N_0 cortar caminho de N_1

N_0 cortar caminho até N_2 Loc N_3

N_0 cortar caminho até N_2

N_0 cortar caminho Loc N_3

N_0 cortar caminho

O caminho de N_1 até N_2 Loc N_3 ser cortado por N_0

O caminho Loc N_3 ser cortado por N_0

Construção de base: N_0 **cortar caminho de** N_1 **até** N_2 **Loc** N_3

Grupo 11

32. O governo **cortou** os investimentos nessa área.

33. João **cortou** o financiamento dos estudos do seu filho.

Distribuição: $N_0 =: N\text{hum}$ $N_1 =: N\text{abs}$

N_0 cortar N_1

N_1 ser cortado por N_0

N_1 ser cortado

N_1 estar cortado por N_0

N_1 estar cortado

Construção de base: N_0 **cortar** N_1

3.2 Tabela de Léxico-Gramática

Na tabela de Léxico-Gramática foram listados somente os itens lexicais onde o verbo *cortar* funciona como verbo pleno, relacionando os itens lexicais com suas propriedades sintático-semânticas.

Tabela 1. Tabela de Léxico-Gramática do verbo *cortar*

VERBO <i>CORTAR</i>																								
TABELA LÉXICO GRAMÁTICA																								
<i>CORTAR</i>	EXEMPLO	N0 =: Nhum	N0 =: Nconc	N0 =: Nabs	N1 =: Nhum	N1 =: Nconc	N1 =: Nabs	N1 =: Npc	N2 =: Nhum	N2 =: Nconc	N2 =: Nabs	N2 =: Npl-obl	N0 V	N0 VN1	N0 VN1 em N2	N0 V N1 Loc N2	N1 ser Vpp	N1 ser Vpp por N0	N1 ser VppLoc N2	N1 ser VppLoc N2 por N0	N1 estar Vpp	N1 estar Vpppor N0	N1 estar VppLoc N2	N0 VN1 em N2
<i>Cortar</i>	João <i>cortou</i> os dois últimos parágrafos do texto.	+	-	-	+	+	+	-	-	+	+	-	-	+	-	+	+	+	+	+	+	-	+	-
<i>Cortar</i>	A companhia <i>cortou</i> a energia elétrica de João.	+	+	+	-	+	+	-	-	-	-	-	-	+	-	-	+	+	-	-	+	+	-	-
<i>Cortar</i>	A travessa <i>corta</i> a rua principal.	+	+	-	-	+	-	-	-	-	-	-	-	+	-	-	+	+	-	-	+	+	-	-
<i>Cortar</i>	O médico <i>cortou</i> o braço de João.	+	+	-	+	+	-	+	-	-	-	-	-	+	-	-	+	+	-	-	-	-	-	-
<i>Cortar</i>	João <i>cortou</i> uma folha/João <i>cortou</i> uma folha em três partes.	+	+	-	+	+	-	+	-	+	-	+	+	+	+	-	+	+	-	-	+	-	-	+
<i>Cortar</i>	João <i>cortou</i> as unhas.	+	+	-	-	+	-	+	-	-	-	-	-	+	-	-	+	+	-	-	+	-	-	-
<i>Cortar</i>	João <i>cortou</i> o vestido com o molde da revista.	+	+	-	-	+	-	-	-	-	-	-	-	+	-	-	+	+	-	-	+	-	-	-
<i>Cortar</i>	O carro de João <i>cortou</i> a moto de Pedro.	+	+	-	+	+	-	-	-	-	-	-	-	+	-	-	+	+	-	-	+	+	-	-

5 – Conclusão

O desenvolvimento de estudos descritivos das categorias de palavras para processamento automático de linguagem natural é um dos recursos que poderão contribuir para melhoria da qualidade de aplicações no PLN. Isso pode, possivelmente, solucionar muitos problemas no processamento das línguas (por exemplo, o reconhecimento de formas numa sequência linear sem o comprometimento do sentido das informações, eliminando ambiguidades, redundâncias, repetições e informações agramaticais). Apesar de suas limitações⁵, trata-se de um estudo de fundamental importância, conforme observa Laporte (2008, p. 09): “As tabelas do léxico-gramática, [...] mesmo não sendo totalmente isentas de erros [...] [permitem] corrigir os erros e construir as tabelas que faltam para outras partes do léxico ou então para outras línguas”. A codificação e a identificação de um padrão geral para o verbo *cortar* se inserem neste contexto como mais uma estrutura a ser descrita, somando-se a outras já anteriormente realizadas.

6 – Referências Bibliográficas

BORBA, F. da S. Propriedades sintáticas, semânticas e pragmáticas do léxico. *Revista (com) textos linguísticos*. Vitória-Es. n.1. 55-68, 2007.

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

KOBASHI, Nair Yumiko. Fundamentos semânticos e pragmáticos da construção de instrumentos de representação de informação. *Revista de Ciência da Informação* - v.8 n.6 dez/07. Artigo 01. Disponível em: http://www.dgz.org.br/dez07/Art_01.htm

LAPORTE, Éric. A Linguística para o processamento das línguas. **1o Colóquio de Estudos Linguísticos**, Vitória (Espírito Santo): Brésil (2000). Disponível em: <http://hal.archives-ouvertes.fr/docs/00/36/94/10/PDF/vitoria.pdf>

LAPORTE, Éric. Exemplos atestados e exemplos construídos na prática do léxico-gramática. *Revista (Con)textosLinguísticos* 2 (2008) pg 26-51. Tradução do francês: Francisco Antonio P. Léllis. 2008. Disponível em: <http://halshs.archives-ouvertes.fr/docs/00/32/59/26/PDF/obsManipPort.pdf>

NEVES, Maria Helene Moura. **Gramática de usos do português**. São Paulo: UNESP, 1999

PAUMIER, Sébastien. **Unitex – Manual de utilização**. Université Marne-la-Valée. 2002. Tradução Oto Vale e Alexis Neme. Disponível em <http://infolingu.univmlv.fr/brasil/Ferramentas/ManualUnitex-0-1-2-3-4.zip>.

SCHER, A. P. As construções com o verbo leve DAR e nominalizações em –ada no português do Brasil. Tese (Doutorado em Letras). UNICAMP, Campinas, São Paulo, 2004.

SMARSARO, Aucione. O léxico e o processamento de linguagem natural. Aucione Smarsaro. R. (con) *Textos Linguísticos*. Vitória nº 1 p.49 - 54 - 170 2007. Disponível em: <http://www.linguistica.ufes.br/sites/www.linguistica.ufes.br/files/Revista%20Contextos%20Linguisticos%20-%20Vol1.pdf>

⁵ A respeito das limitações do PLN, cabe citar o seguinte comentário de Smarsaro (2007, p. 49): “Até os textos produzidos pelos melhores sistemas de ajuda à tradução necessitam de um sistema para tradutores humanos, devido a erros de tradução que, aliás, tomam os resultados da tradução automática quase um gênero literário cômico”.