O verbo *dar* em português brasileiro:

descrição, classificação e processamento automático

Letraria®

Amanda Pontes Rassi

Amanda Pontes Rassi

O verbo *dar* em português brasileiro:

descrição, classificação e processamento automático

Araraquara
Letraria
2023

Ficha catalográfica

Dados Internacionais de Catalogação na Publicação (CIP)
(Câmara Brasileira do Livro, SP, Brasil)

Rassi, Amanda Pontes
O verbo dar em português brasileiro [livro eletrônico]: descrição, classificação e processamento automático / Amanda Pontes Rassi. - Araraquara, SP : Letraria, 2023
PDF.

Bibliografia.
ISBN 978-65-5434-023-6

1. Língua portuguesa - Conjunções 2. Língua portuguesa - Semântica 3. Língua portuguesa - Sintaxe 4. Língua portuguesa - Verbos I. Título.

22-139996 CDD-469.5

Índices para catálogo sistemático:

1. Verbos : Gramática : Português : Linguística 469.5
Inajara Pires de Souza - Bibliotecária - CRB PR-001652/O

CAPES

Conselho editorial

Aline Fernandes de Azevedo Bocchi – UNIFRAN

Anderson Salvaterra Magalhães – UNIFESP

Claudia Zavaglia – UNESP/Ibilce – *Campus* São José do Rio Preto

Luiz Francisco Dias – UFMG

Marcelo Concário – UNESP – *Campus* Bauru

| SUMÁRIO

Prefácio 11

Introdução 16

1 FUNDAMENTAÇÃO TEÓRICA **28**

1.1 Harris e a gramática transformacional 28

1.1.1 A noção de operadores e argumentos e as restrições de seleção 29

1.1.2 A frase elementar e a noção de *transformação* harrissiana 31

1.1.3 As transformações unárias 34

1.2 O modelo do Léxico-Gramática 35

1.2.1 Premissa 1: Cada unidade lexical tem uma gramática própria 37

1.2.2 Premissa 2: A unidade mínima de análise é a frase elementar 38

1.2.3 Premissa 3: Os testes de aceitabilidade são feitos com recurso massivo à introspecção 39

2 PRINCÍPIOS E PROCEDIMENTOS METODOLÓGICOS **43**

2.1 Obtenção dos dados 44

2.1.1 *Corpus* PLN.Br Full 44

2.1.2 Lista de colocados no AC/DC 45

2.1.3 Lista de *Npred* em PE 45

2.2 Ferramentas de busca e processamento 46

2.2.1 A ferramenta Unitex 46

2.2.2 O concordanciador WebCorp Live 48

2.2.3 A ferramenta CorpusAnnotator 49

2.2.4 A ferramenta ReCal 0.1 Alpha for3+ Coders 50

2.3 Princípio de classificação 52

2.3.1 Da classificação geral 52

2.3.2 Das construções com verbo-suporte (*CVS*) 53

2.4 Método de descrição do Léxico-Gramática 53

2.5 Integração dos dados em *parser* 57

3 ESTADO DA ARTE 59

3.1 Classificação geral das construções verbais 59

3.2 Identificação e descrição das *CVS* 64

3.3 Identificação e/ou processamento automático das *CVS* 68

PARTE I – PROPOSTA TIPOLÓGICA PARA O VERBO DAR 73

4 O VERBO PLENO *DAR* 74

4.1 Verbo pleno com complemento dativo 74

4.2 Verbo pleno com complemento reduzido/apagado 77

4.3 Verbo pleno com complemento locativo 78

4.4 Verbo pleno com complemento quantificador 79

4.5 Verbo pleno com complemento instrumental 80

4.6 Verbo pleno com complemento não-restrito introduzido pela prep 'com' 81

4.7 Verbo pleno transitivo-predicativo 82

4.8 Verbo pleno com complemento não-preposicionado 83

5 O VERBO-OPERADOR CAUSATIVO *DAR* 85

5.1 Propriedades das construções com o *VopC dar* 85

5.1.1 Restrições ao preenchimento lexical do argumento na posição sujeito 86

5.1.2 A comutação do *VopC* com outros verbos causativos 87

5.2 O estatuto das construções causativas 88

5.3 Tipos de frases causativas 90

5.3.1 Causa sensação ou sentimento 92

5.3.2 Causa doença 93

5.3.3 Atribui propriedade ou característica 95

5.3.4 Verbo causativo-resultativo 97

6 AS FRASES FIXAS **100**

6.1 As Expressões Cristalizadas 101

6.1.1 Classe C1 106

6.1.2 Classe CDN 109

6.1.3 Classe CP1 111

6.1.4 Classe C1PN 113

6.1.5 Classe CNP2 116

6.1.6 Classe C1P2 117

6.1.7 Classe CPP 119

6.1.8 Classe CPPN 122

6.1.9 Classe N=Ø 123

6.2 Os provérbios 124

PARTE II – CONSTRUÇÕES COM O VERBO-SUPORTE *DAR* E NOME PREDICATIVO 130

7 O CONCEITO DE *CONSTRUÇÃO NOMINAL COM VERBO-SUPORTE* **131**

7.1 O conceito de verbo-suporte 131

7.2 O conceito de nome predicativo 134

8 PROPRIEDADES GERAIS DAS *CVS* **137**

8.1 Relação particular entre o *Npred* e um de seus argumentos 137

8.2 Restrições sobre os determinantes 138

8.3 Descida do advérbio 142

8.4 A dupla análise dos complementos preposicionais 144

8.5 Redução do *Vsup* e formação de grupo nominal 145

8.6 Possibilidade de variação do *Vsup* 146

9 PROPRIEDADES SINTÁTICO-SEMÂNTICAS DAS CONSTRUÇÕES COM *DAR* **148**

9.1 Propriedades formais 149

9.1.1 Número de argumentos 149

9.1.2 Inserção de expressões de quantidade 151

9.1.3 Determinantes 155

9.1.4 Preposições 158

9.1.4.1 Especificidades das preposições *a* e *para* 159

9.1.4.2 Especificidades das preposições *com* e *contra* 161

9.1.4.3 Especificidades da preposição *em* 163

9.1.4.4 Especificidades da preposição *de* 164

9.1.4.5 Especificidades da preposição *sobre* 166

9.1.4.6 Especificidades das preposições locativas 166

9.2 Propriedades distribucionais 168

9.2.1 Nomes humanos (*Nhum*) e Nomes não-humanos (*N-hum*) 168

9.2.2 Nomes parte do corpo (*Npc*) 171

9.2.3 Nomes Locativos (*Nloc*) 173

9.2.4 Frases completivas (*QueF*) na posição de argumento 176

9.2.5 Papéis semânticos dos argumentos 178

9.3 Propriedades transformacionais 180

9.3.1 Simetria 180

9.3.2 Conversão 183

9.3.3 Apassivação 186

9.3.4 Nominalização 187

9.4 Desdobramentos lexicais 189

10 CLASSIFICAÇÃO DAS *CVS* **192**

10.1 Classificação sintática das *CVS* 192

10.2 Clusterização semântica das *CVS* a partir dos *Npred* 195

10.2.1 *Npred* de esportes 195

10.2.2 *Npred* de gêneros acadêmicos 197

10.2.3 *Npred* de documentos 197

10.2.4 *Npred* de atos violentos 198

10.2.5 *Npred* de fala 199

10.2.5.1 Nomes predicativos enquanto nomes operadores 201

10.2.5.2 Expressões interjetivas que migram para a classe dos *Npred* 202

10.3 O verbo-suporte de ocorrência 204

PARTE III – PROCESSAMENTO COMPUTACIONAL DOS PREDICADOS NOMINAIS 211

11 O PROCESSAMENTO DAS CONSTRUÇÕES COM VERBO-SUPORTE **212**

11.1 Considerações preliminares 212

11.2 Cadeia de processamento automático do Português - STRING 214

11.3 Estratégia para implementação dos dados na STRING 216

12 A PRODUÇÃO DO *CORPUS* DE REFERÊNCIA **220**

12.1 Extração automática das frases 220

12.2 A seleção da amostra 223

12.3 O processo de anotação 224

12.3.1 A ferramenta de anotação 224

12.3.2 Procedimentos de anotação 225

12.3.3 Normalização da anotação 226

12.4 Cálculo da concordância entre anotadores 227

12.5 Seleção da subamostra 229

13 INTEGRAÇÃO DOS DADOS NA *STRING* **232**

13.1 Resultados da primeira avaliação 232

13.2 Análise dos erros 234

13.2.1 Análise dos falsos-positivos: erros de classificação do tipo de dependência 234

13.2.3.1 A identificação de expressões fixas como *CVS* 239

13.2.3.2 A identificação de nomes compostos como nomes predicativos simples 241

13.2.3.3 A classificação indevida de construções com o verbo-operador de ligação *ter* como *CVS standard* 241

13.2.2 Análise dos falsos-negativos: Problemas de não-identificação da dependência SUPPORT 245

13.2.3 Análise dos verdadeiros-negativos: casos em que a STRING acertadamente não extraiu dependência 248

13.3 Casos particulares de verdadeiros-positivos 249

13.4 Segunda avaliação da performance do sistema 251

14 VALIDAÇÃO E CONVERSÃO AUTOMÁTICA DOS DADOS EM REGRAS **255**

14.1 Validação formal 255

14.2 Validação linguística 257

14.3 Conversão dos dados em regras de extração de dependência 259

Conclusão 264

REFERÊNCIAS 268

APÊNDICES

Apêndice A – As categorias 281

Apêndice B – O processo de anotação 282

Apêndice C – Matriz do Léxico-Gramática 283

Apêndice D – *Guidelines* para a anotação 284

| Prefácio

A Hora da Estrela Kelly Sacoman

Por Samuel Ponsoni[1] e Roberto Leiser Baronas[2]

> Pois na hora da morte a pessoa se torna brilhante estrela de cinema, é o instante de glória de cada um e é quando como no canto coral se ouvem agudos sibilantes.
>
> Clarice Lispector em *A Hora da Estrela*

Alhures já dizia o bom e velho Aristóteles que *a arte imita a vida*. Todavia, não raramente *a vida [com requintes de crueldade] imita a arte*. Este prefácio dialoga com a segunda parte da asserção, derivada do pensamento aristotélico. Dizemos isso pelo fato de que a inspiração dessa obra é um trágico acontecimento. Muito similar ao vivido pela principal personagem do romance *A Hora da Estrela* de Clarice Lispector. Macabéa sai da cartomante cheia de "saudades do futuro", pronta para começar a sua nova vida. Porém, ao atravessar a rua, é atropelada. Neste romance de Clarice, num dos raros momentos de hesitação do narrador, engendrado pela genialidade da autora, cabe ao leitor definir o futuro de Macabéa: morrer ou vomitar uma estrela de mil pontas.

Kelly Cristina Bognar Sacoman não teve a mesma sorte de Macabéa, pois, no dia 10/06/2013, ao retornar de São Carlos à sua casa, depois de retirar o seu diploma de Mestrado no Programa, "cheia de saudades do futuro", sofreu um trágico acidente automobilístico e faleceu. Porém, um acontecimento, nos diria o bom e velho Michel Pêcheux, possui ainda muitos outros elementos a serem colocados dentro e fora do logicamente estabilizado. O acidente foi causado por uma caminhonete que, em sentido contrário, na Rodovia Luís Augusto de Oliveira (SP-215), na altura da cidade de Ribeirão Bonito-SP, por volta das 19h30, colidiu com o carro de Kelly, que estava acompanhada de sua amiga Lilian Gaiotto de Moraes – a qual, felizmente, sofreu apenas ferimentos leves –, causando o grave acidente que resultou na morte de nossa ex-aluna. Testemunhas disseram, à época, que o motorista da caminhonete estaria supostamente embriagado (de não parar em pé!!!!!) e, pior, evadiu-se covardemente do local, com ajuda de terceiros, especula-se, sem sequer prestar socorro à nossa ex-companheira de UFSCar.

---

1 Professor na Universidade do Estado de Minas Gerais – UEMG – Unidade de Passos e à época da institucionalização do Prêmio Kelly Sacoman, enquanto pós-doutorando no Programa de Pós-Graduação em Linguística da UFSCar – PPGL, coordenou todos os trabalhos atinentes ao concurso. E-mail: samuel.ponsoni@uemg.br

2 Professor no Departamento de Letras e no Programa de Pós-Graduação em Linguística da UFSCar e à época da institucionalização do Prêmio Kelly Sacoman, era o coordenador do PPGL. E-mail: baronas@ufscar.br

Kelly buscara seu diploma naquele dia, uma vez que havia passado em um concurso e utilizaria a certificação para a posse em seu novo trabalho. Kelly também iria de Bauru, sua cidade natal, na qual foi enterrada, a Campinas, visitar seu namorado, já que estávamos às voltas do dia dos namorados daquele ano[3]. Este prêmio é uma homenagem póstuma à nossa estrela *Kelly Cristina Bognar Sacoman*.

Mas, recuperando forças na comoção, ao reviver essa história pela escritura deste prefácio, devemos também falar daquilo que não é trágico, daquilo que deriva, se possível, como algo bom, isto é, os trabalhos que aqui temos à mão. Os trabalhos de pesquisa ora reunidos e publicados derivam do Edital do *Prêmio Kelly Cristina Bognar Sacoman de Teses e Dissertações*, de 2017, promovido pelo Programa de Pós-Graduação em Linguística – PPGL, da Universidade Federal de São Carlos, *campus* São Carlos-SP. Fruto de um longo processo de formação crítica em relação aos saberes sobre linguagem, discurso, ensino-aprendizagem, entre outros temas, através de diversas dissertações e teses, produzidas por alunos e professores do PPGL/UFSCar, entre 2005 e 2016, há nessas pesquisas selecionadas reflexões, apontamentos e compreensões absolutamente instigadoras e de muita qualidade.

Com efeito, podemos asseverar o dito alhures com absoluta certeza diante de trabalhos, tais como:

i) *A constituição da fórmula discursiva "cultura de paz": circulação e produção dos sentidos*, de Helena Maria Boschi da Silva, feito sob a orientação da professora Luciana Salgado Salazar, na qual a pesquisadora discute o conceito de fórmula discursiva, principalmente a partir das perspectivas teóricas de Alice Krieg-Planque, mas também da própria professora Luciana Salazar e de outros autores de renome. Dito de outro modo, a pesquisadora Helena Maria mobiliza o enunciado "cultura de paz" e o observa como fórmula que mantém características de polêmica, cristalização, referente social, entre outros, na circulação de sentidos históricos e ideológicos nos espaços de embate discursivo do Brasil;

ii) *Da repressão à abertura política: processos de espetacularização do discurso político*, de Israel de Sá, cuja orientação esteve a cargo da professora Vanice Maria Oliveira Sargentini, propõe, por sua vez, uma análise de discursos produzidos no período da última (e tomara que última) ditadura militar brasileira (1964-1985); parte de discursos mais repressivos até mais midiáticos (e espetacularizados) acerca do tema e do período, perscrutados em distintos gêneros discursivos (doutrinário, memorialista,

3 Os leitores e as leitoras podem acompanhar esses e outros detalhes nesta reportagem: https://sampi.net.br/bauru/noticias/2357775/geral/2013/06/acidente-em-rodovia-mata-bauruense

jornalístico), sob autoria de sujeitos, quer políticos, quer do campo da mídia. Assim, Israel de Sá compreende que esse movimento discursivo faz emergir à superfície uma certa passagem de uma sociedade repressiva para uma sociedade midiática. Fundamentado em perspectivas teóricas discursivas, ancorado em autores incontornáveis, como Michel Pêcheux e Jean-Jacques Courtine, bem como nos diálogos teóricos entre Michel Foucault e a Nova História, temos um excelente trabalho que parecia falar de um passado, de uma memória quase distante da sociedade brasileira, todavia que os últimos anos, como bem sabemos, não nos deixam tantas certezas assim;

iii) *Formação de professores a distância e o acesso de crianças de seis anos à linguagem escrita através de projetos de letramento,* Ariane Ranzani, orientada pela professora Maria Sílvia Cintra Martins, fez e publica aqui um trabalho de pesquisa em que se tem uma excelente proposta de aperfeiçoamento qualificativo para a formação de professores e professoras de ensino fundamental, principalmente em relação ao ensino de produção escrita de crianças de seis anos. Propõe, para tanto, um projeto integrado de letramento, em que a coleta de dados, em observação pela pesquisa, em diferentes gêneros discursivos, se fez em ambiente virtual, por atividades de processo de escrita dado aos jovens alunos(as). Utiliza uma fundamentação teórica bastante pertinente, tal como a Teoria da atividade, de Leontiev, portanto, a linguagem vista sob o prisma de uma atividade sócio-histórica. Acresce, ainda, o papel da psicologia cultural-histórica, de Vigotski, no processo de ensino-aprendizagem do letramento e da escrita, o que nos dá, portanto, pelo todo do trabalho, diversos e consistentes caminhos e apontamentos na dimensão histórica entre linguagem, ensino e escrita;

iv) *Descrição, classificação e processamento automático das construções com o verbo* dar *em português brasileiro*, de Amanda Rassi, orientada por Oto Araújo Vale, nos traz à luz uma robusta descrição formal do verbo "dar", do português brasileiro, para interessantes aplicações computacionais. Analisou, dessa forma, diversas tipologias e ocorrências do verbo "Dar", em exemplos categorizados em, sobretudo, quatro formas de aparição no português brasileiro, a saber, como verbo pleno, operador causativo, verbo-suporte e constituinte de frase fixa. Assim, é uma pesquisa de muita qualidade no entrecruzamento da Linguística e do Processamento de Língua Natural (PLN), pesquisa essa em que temos uma grata contribuição científica ao nosso alcance;

v) *O paradoxo a partir da teoria dos blocos semânticos*: *língua, dicionário e história*, de Júlio César Machado, com orientação de Soeli Maria Schreiber da Silva, pesquisa, e nos apresenta, um estudo sólido e denso sobre o conceito de paradoxo semântico, demonstrando o funcionamento da interdependência entre significâncias opostas.

A análise desse funcionamento se propõe a compreender aspectos orgânicos da língua e, na mesma medida, trabalhar aspectos argumentativos da enunciação em língua. Pode-se dizer, ainda, que a análise da significação está ancorada no cubo argumentativo, que se imbuí, por seu turno, de um desenvolvimento do quadrado argumentativo, ou seja, das identidades analíticas do modelo de blocos semânticos, concebido por Carel/Ducrot no final dos anos 1990 e início dos anos 2000. Portanto, temos um instigante avanço teórico, à brasileira, sobre consagradas teorias linguísticas de terras francesas;

vi) *A estética da cultura popular na folia de momo do recife: questões de alteridade, corporeidade e transgressão,* de Hélio Pajeú, de orientação amparada no professor Valdemir Miotello, faz uma compreensão e análise da cultura popular como o espaço de alteridades nas construções de identidades, à luz das clássicas teorias de Mikhail Bakhtin acerca de cultura popular. Pensando a esfera de atividade do Frevo, sobretudo em relação a outras manifestações e expressões festivas, principalmente da cena pernambucana, tais como, Boi de Carnaval, Maracatu, Urso de Carnaval e Galo da Madrugada, Pajeú reflete questões estéticas na construção do outro, através das principais personagens que se encenam e encenam (n)essas festividades.

Dessa forma, é desse longo e diversificado conjunto de atividades de pesquisas que buscamos o chamado algo de bom, dito antes. Este algo que se iniciou, como dissemos, de uma história trágica, mas que nos brinda, agora, com a publicação desses excelentes trabalhos, pesquisas na mais justa e precisa definição do termo, de jovens pesquisadores do campo da linguística e que passaram pelo PPGL/UFSCar.

Temos certeza de que todos os autores e as autoras aqui mencionados(as) e publicados(as) sentem-se honrados e orgulhosos de fazer parte dessa justa e merecida homenagem à Kelly. Temos certeza de que Kelly se sentiria honrada e homenageada com este livro, cercada de bons companheiros e boas companheiras de PPGL e de bons trabalhos.

Resta ainda mais um gesto de afeto, desta vez um Agradecimento muito especial aos/às pareceristas que foram determinantes para a escolha da Tese e Dissertação, agraciadas com o *Prêmio Kelly Cristina Bognar Sacoman de Teses e Dissertações*, de 2017.

Pareceristas das dissertações

Diana Vasconcelos Lopes – Universidade Federal Rural de Pernambuco – UFRPE;
Maria Inês Pagliarini Cox – UFMT;
João Kogawa – Universidade Federal de São Paulo – Unifesp;
Solange Aranha – UNESP/Ibilce – *Campus* – São José do Rio Preto-SP

Pareceristas das teses

Claudia Zavaglia – UNESP/Ibilce – *Campus* São José do Rio Preto-SP;
Marcelo Concário – UNESP – *Campus* Bauru;
Anderson Salvaterra Magalhães – UNIFESP;
Luiz Francisco Dias – FALE/UFMG;
Aline Fernandes de Azevedo Bocchi – UNIFRAN.

São Carlos, primavera, nos estertores do turbulento, complexo, ameaçador..., mas com aparente alívio e respiro, ano de 2022.

Os prefaciadores

Introdução

Os verbos *ser*, *ter*, *estar*, *fazer*, *dar*, *ir* e *haver* são em geral os verbos mais produtivos das línguas românicas. Fauci e Mirto (2003) já haviam indicado *ser*, *estar*, *dar*, *ter* e *fazer* como os mais recorrentes em línguas latinas, mas, recorrendo-se a uma busca breve por frequência de verbos em Português Brasileiro e Europeu[4], encontram-se os dez verbos mais frequentes, nesta ordem: *ser*, *ter*, *estar*, *fazer*, *poder*, *ir*, *dizer*, *haver*, *dar* e *dever*.

A grande produtividade desses verbos reflete também uma complexidade maior na análise e classificação de suas construções. As construções com o verbo *dar*, por ser um dos mais produtivos e complexos da língua, serão investigadas e descritas neste livro. Esse verbo pode ocorrer em Língua Portuguesa como verbo pleno, verbo-suporte, operador causativo, verbo auxiliar, construção modal, verbo impessoal, elemento formador de expressão cristalizada, elemento constituinte de estruturas proverbiais, dentre outras classificações.

Há diferentes trabalhos na literatura que descrevem uma classe ou outra, mas não foi identificada ainda nenhuma descrição completa capaz de abranger todas as categorias das construções em que o verbo *dar* aparece. A título de exemplo, citem-se os trabalhos de Vaza (1988), Baptista (1997) e Scher (2004) sobre o verbo-suporte *dar*; de Salomão (2008) e Auwera e Plungian (1998) sobre as construções modais com o verbo *dar*; de Baptista, Mamede e Gomes (2010) e Duran e Aluísio (2010) sobre os verbos auxiliares; de Vale (2001) e Baptista, Correia e Fernandes (2004) sobre as construções fixas, dentre outros.

A literatura mais vasta de que dispomos, no entanto, é sobre as construções nominais com verbo-suporte. Há descrições bastante refinadas para as construções nominais em Francês (GIRY-SCHNEIDER, 1978; LABELLE, 1984; GIRY-SCHNEIDER, 1987; LECLÈRE, 1990; LAMIROY, 1998), em Italiano (FAUCI; MIRTO, 2003; CANTARINI, 2004; BRATÁNKOVÁ, 2013), em Português Europeu (RANCHHOD, 1990; BAPTISTA, 1997; BAPTISTA, 2005b; CHACOTO, 2005), dentre outras línguas. Há também descrições de construções com verbos-suporte em Português Brasileiro (NEVES, 1996; NEVES, 2002; SCHER, 2004; SCHER, 2005; SCHER, 2006; CONEJO, 2008; ASSIS, 2009; DAVEL, 2009; ABREU, 2011), mas convém ressaltar que nenhum deles visa à exaustividade nem a uma formalização aplicável ao Processamento de Linguagem Natural (PLN).

4 Foi efetuada uma busca por frequência dos lemas verbais em todos os *corpora* escritos de Português Brasileiro e Europeu, disponíveis no AC/DC (Acesso a *Corpus*/Disponibilização de *Corpus*). A busca pode ser reproduzida através do endereço http://www.linguateca.pt/ACDC/.

Nesse sentido, esta pesquisa apresenta uma proposta de tipologia sintático-semântica de algumas construções em que o verbo *dar* ocorre, com base nas suas diferentes classificações. Além disso, apresenta a descrição formal dos predicados nominais construídos com o verbo-suporte (*Vsup*) *dar* e um nome predicativo (*Npred*) e uma proposta de processamento automático dessas construções a partir da integração dos dados linguísticos em um analisador sintático automático.

A noção de *exaustividade* utilizada neste trabalho é apenas relativa, e não absoluta. Consideramos que o recenseamento dos dados foi feito de maneira exaustiva porque considerou todas as instâncias do *corpus*, portanto é uma exaustividade relativa à extensão do *corpus* adotado, o que pode ser considerado uma amostra substancial em relação à língua como um todo. O conceito de exaustividade absoluta é utópico porque é impossível recensear todas as possíveis ocorrências de uma língua natural.

Os exemplos que aparecem no decorrer deste livro ou foram retirados da *web* ou são exemplos construídos. De forma geral, quando se pretende comprovar a existência de determinada construção ou fenômeno, recorre-se a frases reais da linguagem em uso (principalmente da *web*); quando se pretende demonstrar transformações sintáticas ou apenas exemplificar uma construção cuja aceitabilidade é certa, são utilizados exemplos construídos. Para distinguir entre os dois tipos, serão utilizadas as notações "[Ex.R]", para os exemplos reais, e "[Ex.C]" para os exemplos construídos/fabricados, ao final de cada exemplo.

Problemática relacionada ao objeto de estudo

Os sistemas computacionais que processam automaticamente as línguas naturais e, em particular, o Português analisam da mesma forma as construções verbais e as construções nominais, como se, em todas elas, o verbo fosse o predicador central da frase. Muitas vezes, até mesmo as análises linguísticas feitas por humanos incorrem nesse mesmo erro, de considerar a mesma estrutura sintática e a mesma análise para construções verbais e nominais.

Há, no entanto, diversos trabalhos na literatura que já provaram que existem diferenças sintáticas entre construções verbais, cujo predicador é o verbo, e construções nominais, cujo predicador pode ser um nome ou um adjetivo. Particularmente nas construções com verbo-suporte, o predicador central da frase é um nome, enquanto o verbo serve apenas como suporte desse predicador. Se existem diferenças linguísticas entre esses dois objetos, pergunta-se: por que não tratá-las e analisá-las, de formas diferentes, respeitando suas especificidades?

A partir dessa lacuna, que existe tanto nas descrições gramaticais e nos materiais didáticos quanto no processamento computacional da língua, este trabalho propõe uma abordagem teórica e metodológica para descrever as construções com o verbo-suporte *dar* e analisá-las automaticamente enquanto construções nominais. Para atingir esse objetivo, é necessário antes delimitar o escopo da pesquisa e do objeto a ser pesquisado, a saber, as construções com o verbo *dar*.

O verbo *dar*, de forma geral, é identificado em vários dicionários, gramáticas e materiais didáticos como verbo pleno que exige três argumentos: um doador, na posição sujeito, um objeto a ser dado, na posição de objeto direto, e um recebedor, na posição de objeto indireto. Em outras palavras, é um verbo distribucional com complemento dativo, cujo significado implica a transferência de determinado objeto de alguém a outrem.

(1) *E meu pai deu uma bicicleta para meu irmão.* [Ex.R]

(2) *A professora deu um livro para a gente.* [Ex.R]

Nesses dois casos, transfere-se a posse do objeto concreto (*bicicleta* ou *livro*) das mãos de um agente (*meu pai* ou *A professora*) para as mãos de um recipiente ou recebedor (*meu irmão* ou *a gente*). Além de possuir um sentido próprio (sentido dativo), o verbo *dar*, nos dois casos, impõe restrições de seleção ao sujeito doador, ao objeto a ser transferido e ao complemento dativo recebedor; dessa forma, é classificado como verbo pleno.

Por outro lado, a classificação de frases como *O professor deu um presente para o aluno* é discutível. Alguns autores, por exemplo, Coelho, Silva e Sousa (2013), classificam o verbo *dar* da construção *dar presente* como um "verbo pleno [que] apresenta comportamento lexical [...] *dar* é autônomo e está ligado à noção de transferência" (COELHO; SILVA; SOUSA, 2013, p.3211). Mais adiante, os mesmos autores distinguem o verbo pleno do verbo-suporte. "Essa categoria [verbo-suporte] coloca em questão o princípio da economia linguística, podemos encontrar na língua um verbo com o mesmo valor, como em: *dar amor*, *amar*; *dar um abraço*, *abraçar*" (COELHO; SILVA; SOUSA, 2013, p.3212). Pela definição apresentada pelos autores, a construção *dar presente* deveria ser classificada como uma construção com verbo-suporte, já que existe na língua "um verbo com o mesmo valor" (*presentear*).

Os casos mais típicos de construções com verbo-suporte, tais como, *dar um abraço*, *dar um beijo* e *dar um chute* são consensuais. O problema da classificação surge no momento em que se analisam os casos que não são característicos de uma nem de outra categoria.

O mesmo nome *presente* pode entrar em outros tipos de construções, cuja classificação também não é consensual, como em:

(3) *Uma vez ele deu de presente um ovo de dragão a Lorde Buterwells.* [Ex.R]

A expressão *dar de presente* se assemelha sintática e semanticamente às expressões *dar de brinde*, *dar de graça*, *dar de bandeja* e outras. Essas construções são classificadas por Fernandes (2007) como *expressões fixas*. As mesmas construções são classificadas por Palma (2009) como *advérbios compostos*. Na análise proposta por Noimann (2007, p.90), a expressão *dar de bandeja* é classificada como uma *locução verbal*.

É interessante notar que os complementos *de presente*, *de bandeja*, *de brinde etc.* não são fixos juntamente com o verbo. Assim como outros modificadores, podem mudar de posição sintática sem alterar o significado da frase como um todo.

(3) *Uma vez ele* ***deu de presente*** *um ovo de dragão a Lorde Buterwells.* [Ex.R]

(3a) *Uma vez ele* ***deu*** *um ovo de dragão* ***de presente*** *a Lorde Buterwells.* [Ex.C]

(3b) *Uma vez ele* ***deu*** *um ovo de dragão a Lorde Buterwells* ***de presente****.* [Ex.C]

Essas expressões entram tanto em construções verbais (*e.g. Uma vez ele deu de presente um ovo de dragão a Lorde Buterwells*) quanto em construções nominais com verbo-suporte, como em:

(4) *E olha que ele te deu de bandeja um conselho que custou caro a ele.* [Ex.R]

O exemplo (4) contém uma construção com verbo-suporte porque é o nome *conselho* quem impõe as restrições de seleção ao preenchimento lexical dos argumentos. O verbo *dar*, nesse caso, é classificado como suporte porque serve para "suportar" as marcas verbais número-pessoais e modo-temporais. Como o nome, devido à sua própria morfologia, não pode veicular esses valores, ele seleciona um verbo, que serve de suporte para atualizar essas marcas verbais.

O conceito de verbo-suporte, no entanto, não se restringe à sua propriedade sintática de carregar os valores gramaticais de número, pessoa, modo, tempo e aspecto. Respeitando-se as várias definições formalistas, gerativistas e funcionalistas para esse termo, consideramos que as construções com verbo-suporte (*CVS*) apresentam propriedades sintáticas e semânticas relativamente regulares; apresentam propriedades formais, distribucionais e transformacionais próprias; e podem possuir diferentes funções semânticas e pragmáticas. As propriedades que distinguem os verbos-suporte dos verbos plenos já foram longamente descritas por vários autores e serão retomadas neste trabalho como uma forma de sistematizar a classificação proposta.

Gross (1986, p.107-109) trata de um subtipo de verbos-suporte, chamados verbossuporte de ocorrência (*Vsup-oc*), porque eles instauram uma ocorrência, um acontecimento, ligando um evento a seu circunstancial. Os *Vsup-oc* apresentam especificidades combinatórias com os advérbios que eles "suportam". Essa categoria engloba verbos como: *acontecer*, *ter lugar*, *dar-se*, *ocorrer*, *decorrer*, *suceder*, *etc*. Em Português, essa construção pode ocorrer com o verbo *dar-se* em (5) ou apenas o verbo *dar*, exemplificado em (6).

(5) *Uma confusão se deu nas arquibancadas ainda na etapa inicial.* [Ex.R]

(6) *Deu uma confusão nas arquibancadas ainda na etapa inicial.* [Ex.C]

As duas construções expressam a ocorrência de um evento e, portanto, o nome predicativo é um nome eventivo, mas há diferenças sintáticas entre elas.

O problema dessa construção (6) é que aparentemente não admite sujeito, daí ter sido classificada por Smarsaro e Rocha (2011) como construção impessoal. Em Português, o sujeito é definido como aquele elemento com o qual o verbo concorda. Na linguagem em uso, é muito comum encontrar casos em que o verbo não concorda com o elemento posposto, como no exemplo seguinte:

(7) *Neste domingo só deu brigas e mais brigas em São Francisco do Oeste.* [Ex.R]

O fato de o verbo *dar* não concordar em número com o constituinte *brigas e mais brigas* poderia ser indicativo de que se trata de uma construção impessoal, já que não tem sujeito. Por outro lado, também é possível encontrar casos em que o verbo concorda em número com o nome predicativo, o que nos impediria de classificar construções como (6) como construções impessoais. Veja-se, por exemplo:

(8) *a gente foi pra ver filme mas* ***deram*** *confusões e a gente acabou não fazendo nada.* [Ex.R]

(9) *Infelizmente* ***deram*** *uns problemas no vídeo que seria para hoje e este só poderá ser enviado amanhã.* [Ex.R]

É preciso definir o estatuto dessas e de todas as outras construções sintáticas com o verbo *dar*, com base na aplicação de testes e na reprodutibilidade dos resultados.

Uma construção sintática semelhante ocorre com as construções que indicam fenômenos meteorológicos. Ruwet (1986) tratou dos verbos meteorológicos em diferentes línguas, considerando verbos como *chover*, *nevar*, *trovejar*, *esfriar*, dentre outros. Muitos verbos meteorológicos podem ser nominalizados – *chuva*, *neve*, *trovão*, *frio* – e constituir uma locução juntamente com o verbo *dar*.

(10) *Deu uma chuva forte no Rio de Janeiro.* [Ex.R]

(11) *Deu um trovão ontem à noite.* [Ex.R]

Além dos sentidos bastante recorrentes de *dar* como verbo pleno e como suporte, esse verbo também pode aparecer em construções causativas.

(12) *Comer sal dá sede em Ana.* [Ex.C]

(13) *Fazer exercício físico dá alegria ao Rui.* [Ex.C]

(14) *O escuro dá medo em Ivo.* [Ex.C]

Se adotarmos um conceito de verbo-suporte bastante difundido na literatura, de que uma construção com verbo-suporte e nome predicativo é aquela que tem equivalência com um verbo pleno (propriedade chamada de *nominalização*), deveríamos considerar, então, que *dar alegria* (13) e *dar medo* (14) são *CVS* porque admitem *alegrar* e *amedrontar*, respectivamente. A construção *dar sede* (12), no entanto, não poderia ser uma *CVS* porque não admite **sedentar*.

A propriedade da nominalização, apesar de ser muito usada, não dá conta da grande complexidade que envolve as construções com verbo-suporte, verbo pleno e verbo causativo. Nas construções (12), (13) e (14), o verbo *dar* deve ser classificado como um operador causativo (*VopC*) porque atende às propriedades gerais das construções causativas. Assim como existem propriedades formais que definem intencionalmente as *CVS*, também existem as propriedades formais que identificam as construções causativas.

Essas propriedades serão analisadas pormenorizadamente ao longo do livro, mas, para exemplificar, citam-se: (i) o sujeito de uma construção causativa é um nome não restrito (uma causa); (ii) o complemento das construções causativas é tipicamente um experienciador; (iii) as construções causativas não são construções de base, mas permitem uma construção de base formada pelo verbo *ter*; (iv) existe um rol de verbos também causativos que podem substituir o *VopC dar*; dentre outras propriedades. Em termos semânticos, a construção com *ter* expressa o significado dado pelo *Npred*, enquanto a construção com *dar* implica uma causa. Em termos sintáticos, a construção com *ter* serve como frase de base para a predicação, enquanto a construção com *dar* é uma transformação da frase de base:

(12a) *Ana tem sede (por causa do sal + porque comeu sal).* [Ex.C]

(13a) *Rui tem alegria (por causa do exercício físico + porque fez exercício).* [Ex.C]

(14a) *Ivo tem medo (por causa do escuro + porque está escuro).* [Ex.C]

Nesse sentido, as construções causativas veiculam, ao mesmo tempo, duas informações: (i) a informação expressa na frase de base e (ii) a causa da informação expressa na frase de base.

Há ainda outras construções que exigem um complemento preposicionado e são distribucionais, assim como as construções com verbo pleno, tais como:

(15) *Esse prato dá pra quatro.* [*N0 Vdar para Num*] [Ex.C]

(16) *A janela da casa dá para a praia.* [*N0 Vdar para Nloc*] [Ex.C]

(17) *Rui deu como acabado o relacionamento.* [*N0 Vdar N1* (*por + como*) *Modif*] [Ex.C]

Algumas delas apresentam semelhanças com as expressões cristalizadas, baseadas no grau de fixidez dos elementos, na não-composicionalidade de alguns dos constituintes e nas restrições dos argumentos a serem selecionados. Uma dessas construções que estão no limite das expressões cristalizadas é a estrutura [*N0 Vdar uma de N1*], que é realizada com *Nhum* agentivo na posição sujeito, mas na posição de complemento aceita um grande número de elementos lexicais, tais como:

(18) *Ivo deu uma de (bobo + vítima + mártir + santo).* [Ex.C]

A posição de complemento dessas construções deve ser preenchida por um nome que possa ser usado de forma predicativa, já que a interpretação desse argumento é de predicativo do sujeito. Os nomes que preenchem a posição de *N1* exprimem predicados em construções com o verbo *ser* e sujeito humano (*e.g. Ivo é bobo*, *Ivo é vítima etc.*).

Consideramos essas construções como expressões cristalizadas, mesmo reconhecendo que haja flexibilidade no preenchimento de *N1*, o que é diferente das expressões que possuem todos os elementos fixos e rígidos (com exceção de *N0*), não permitindo nenhuma ou quase nenhuma variação em sua composição estrutural, como em:

(19) *Zé deu com os burros na água.* [Ex.C]

(20) *Ana deu com a língua nos dentes.* [Ex.C]

Quando a construção contém vários constituintes fixos, como (19) e (20), é fácil identificá-la como expressão cristalizada, já que todos os elementos são fixos e a expressão como um todo tem sentido não-literal. No entanto, há também construções formadas apenas pelo verbo *dar* e um único constituinte, tais como, *dar fora*, *dar rata* e *dar zebra*. Essas construções devem ser classificadas como expressão cristalizada ou como construção com verbo-suporte? Existem argumentos que justificam ambos os pontos de vista.

Se considerarmos apenas os critérios sintáticos, a construção *dar uma pedalada*, por exemplo, seria classificada como *CVS*. Em contrapartida, se adotarmos o critério semântico da não-composicionalidade das expressões fixas, podemos considerar, pelo menos, xxx construções diferentes: (i) com o significado literal de "andar de bicicleta"; (ii) com o significado não-composicional de "bater na cabeça de alguém com a palma da mão aberta"; (iii) com o significado não-composicional de "driblar o adversário", ou seja, "no futebol, passar o pé sobre a bola por repetidas vezes, com o objetivo de enganar o marcador"; ou ainda (iv), no contexto político, o nome composto *pedalada fiscal*, referindo-se à "prática do Tesouro Nacional de atrasar de forma proposital o repasse de dinheiro para bancos".

Por todos esses exemplos, já é possível perceber a grande diversidade sintática, morfológica e semântica com que as construções com o verbo *dar* ocorrem em Língua Portuguesa. O objetivo desta seção é apresentar resumidamente alguns dos problemas relacionados à caracterização e classificação do objeto de estudo. Além das expressões e construções já mencionadas anteriormente, há ainda várias outras que também devem ser descritas e classificadas com base em suas propriedades sintático-semânticas, pois fazem parte do léxico da língua, tais como:

(21) *Deu duas horas.* [Ex.R]

(22) *Essa estrada dá em Porto Seguro.* [Ex.R]

(23) *Pe. Pio deu na cara de um fiel.* [Ex.R]

(24) *Hoje no meu show deu mil pessoas.* [Ex.R]

(25) *Minha vida daria um filme.* [Ex.R]

(26) *O antigo conflito Leste x Oeste deu lugar ao atual conflito Norte x Sul.* [Ex.R]

(27) *Ninguém se deu por vencido.* [Ex.R]

(28) *Deu bicho na goiaba.* [Ex.R]

(29) *Esse fim de semana vai dar praia.* [Ex.R]

(30) *Thor Batista e Paola Leça deram um tempo no relacionamento.* [Ex.R]

(31) *Em 2010, dei mais tempo para mim mesma.* [Ex.R]

(32) *Eu escrevi um livro que deu 200 páginas.* [Ex.R]

A partir dos exemplos já indicados, pretende-se neste *e-book* fazer uma análise refinada das construções com o verbo *dar*, propondo-se uma taxonomia e uma classificação dessas construções. O que as gramáticas e os materiais didáticos costumam fazer é indicar as categorias e dar exemplos característicos, que não geram discordância de interpretação, negligenciando a existência dos casos mais complexos. Nosso objetivo, por outro lado, não é apenas indicar a existência desses fenômenos, mas analisá-los sistematicamente, adotando procedimentos experimentais explícitos, que garantam o controle da metodologia e que permitam a escalabilidade dos dados para todos os casos. Todos esses critérios devem ser sistematicamente utilizados, já que pretendemos desenvolver uma descrição exaustiva e coerente dessas construções.

Objetivos e justificativa

A proposta inicial desta pesquisa era a de descrever as propriedades das construções nominais, formadas pelo *Vsup dar* e *Npred*. Para atingir esse objetivo, é necessário delimitar bem o objeto que se pretende descrever, pois algumas construções nominais com verbo-suporte podem apresentar características de construções livres e outras, de construções fixas.

Por vezes, a falta de rigor metodológico e taxonômico dos dicionários e gramáticas em descrever e classificar as construções não nos permite determinar com exatidão a qual categoria pertence cada construção. Assim, será necessário definir e explicar todas as categorias de construções em que o verbo *dar* aparece, para, posteriormente, restringir o estudo à análise das construções nominais de *Vsup* e *Npred*.

Tendo exposto alguns problemas que se colocam à descrição das construções com o verbo *dar* em Português Brasileiro, este trabalho objetiva: i) uma discussão teórica e prática sobre cada categoria que envolve o verbo *dar*; ii) a descrição formal das propriedades das construções com o verbo *dar*, com vistas a aplicações computacionais; iii) a elaboração de uma tipologia para o verbo *dar* em PB, com o objetivo de subsidiar futuras pesquisas e análises linguísticas; e (iv) a identificação de uma metodologia que permita o processamento automático (ou uma análise sintática automática) das *CVS*.

Esse tipo de descrição linguística se justifica por várias razões, dentre as quais citam-se:

- Para o desenvolvimento científico da teoria linguística e gramatical e avanço do estado da arte: o trabalho de investigação dos fenômenos linguísticos, tal como o que se propõe aqui, deve seguir um rigor teórico e metodológico, acumulando evidências

que permitam descrever adequadamente o fenômeno em causa e explicitando todos os procedimentos metodológicos adotados, bem como justificando os critérios e experimentos. O recenseamento exaustivo das construções em *corpora*, associado à análise sistemática dos casos, poderá permitir a escalabilidade da descrição e da classificação, já que propomos uma tipologia verbal que poderá ser replicada futuramente para outras construções verbais e nominais. A combinação desses fatores garante a cientificidade do trabalho – neste caso, trabalho taxonômico – contribuindo, assim, para um refinamento da teoria linguística e gramatical;

- Para o ensino-aprendizagem de Língua Portuguesa por parte de falantes nativos e estrangeiros aprendizes: muitas gramáticas normativas e materiais didáticos ainda desconhecem – ou pelo menos não abordam – a existência das diversas categorias verbais em que o verbo *dar* ocorre; por isso essas informações devem ser sistematizadas para que possam ser incorporadas aos recursos disponíveis para o ensino de língua;
- Para aplicações computacionais: para que as palavras, frases e textos possam ser lidos, interpretados, traduzidos e processados por programas computacionais – a exemplo do que fazem os programas de tradução automática, os sumarizadores automáticos, os sistemas de pergunta e resposta, dentre outros – os dados linguísticos precisam ser descritos em um formato legível por máquina. Um dos maiores problemas para o PLN é a escassez de recursos léxico-computacionais robustos para subsidiar a criação de ferramentas e aplicações computacionais; daí advém a necessidade de uma descrição e uma formalização bem estruturada, como a que se pretende fazer neste trabalho.

Estrutura do *e-book*

Este *e-book* está dividido em três partes. A primeira apresenta a classificação geral das construções (como verbo pleno, verbo-operador causativo e frase fixa). A segunda apresenta a descrição formal das construções com o verbo-suporte *dar* e nome predicativo. Além de uma discussão teórica sobre as características das construções com verbo-suporte, suas propriedades e suas subclassificações, será apresentada também a formalização dos dados por meio de uma matriz binária (Apêndice A). A terceira parte do *e-book* expõe o processo de integração das *CVS* em um analisador sintático automático, com vistas ao processamento automático dessas construções enquanto estruturas sintáticas complexas, porém analisáveis.

Antes da divisão das partes, no entanto, apresenta-se: (i) o arcabouço teórico referente às duas teorias que norteiam o trabalho, a saber a Gramática Transformacional de operadores (HARRIS, 1961, 1964) e o Modelo do Léxico-Gramática (GROSS, 1968, 1975, 1981); (ii) a metodologia utilizada, os dados do *corpus* e as ferramentas computacionais utilizadas em todas as etapas da metodologia; e (iii) uma revisão da literatura e dos trabalhos relacionados, na seção de Estado da arte.

A primeira parte do *e-book* consiste em uma proposta tipológica para as construções com o verbo *dar* e se organiza da seguinte forma: o Capítulo 4 discute a noção de verbo pleno e os casos em que o verbo *dar* se classifica como pleno; o Capítulo 5 apresenta a noção de verbo-operador causativo (GROSS, 1981) e discute suas estruturas e propriedades distribucionais; e o Capítulo 6 expõe e descreve as várias estruturas sintáticas das frases fixas com que o verbo *dar* se constrói, tanto em expressões cristalizadas quanto em provérbios.

A segunda parte do *e-book* está subdividida da seguinte forma: o Capítulo 7 apresenta a definição de *construção com verbo-suporte* (*CVS*), bem como as definições de *verbosuporte* (*Vsup*) e de *nome predicativo* (*Npred*); o Capítulo 8 discute os testes aplicáveis às *CVS* para a identificação de suas propriedades definitórias; o Capítulo 9 descreve as propriedades formais, distribucionais e transformacionais analisadas para as construções com o verbo-suporte *dar*; e, por fim, o Capítulo 10 propõe uma classificação das *CVS*, com base em suas propriedades sintáticas, e uma clusterização semântica dos *Npred*, com base em seus significados.

Já a terceira parte do livro, que trata do processamento automático das *CVS*, está organizada como: o Capítulo 11 define conceitos gerais sobre a área de Processamento Automático de Língua Natural (PLN) e apresenta a cadeia de processamento do Português STRING (MAMEDE *et al.*, 2012) e o módulo de *parser* utilizado nessa cadeia, onde as *CVS* com *dar* foram processadas; o Capítulo 12 descreve a produção de um *corpus* de referência anotado manualmente com *CVS*, que servirá como modelo para analisar a performance da STRING; o Capítulo 14 explica a metodologia de validação dos dados da matriz L-G, que verifica tanto a consistência formal quanto a consistência linguística dos dados, e também o conversor da matriz do L-G em regras que são usadas pelo *parser*; e, por fim, o Capítulo 13 apresenta a metodologia utilizada para integrar as *CVS* no *parser* usado pela STRING e os resultados da performance do sistema na tarefa de analisar *CVS* automaticamente.

Ao fim das três partes deste *e-book*, apresentam-se as considerações finais e os apêndices, que estão organizados como: (i) Apêndice A contém a matriz do Léxico-

Gramática para 1.489 *Npred* que selecionam o *Vsup dar* em construções de base; (ii) Apêndice B apresenta a lista dos nomes predicativos que ocorrem em construções com o verbo-operador causativo *dar*, porém suas construções de base se fazem com o *Vsup ter*; (iii) Apêndice C contém outra matriz do L-G com cerca de 170 expressões fixas, seguida de uma lista de provérbios em que o verbo *dar* é fixo; e (iv) Apêndice D com as *Guidelines* para a anotação de construções com verbos-suporte e construções com verbo operador causativo em Português.

1. Fundamentação teórica

A teoria de base que norteia este *e-book* é o Modelo do Léxico-Gramática (L-G) que, em linhas gerais, considera que cada unidade lexical da língua possui uma gramática própria, por ser regida por regras específicas de distribuição e organização sintática e semântica. Pelo modelo, os componentes lexical, sintático e semântico não podem ser dissociados quando se trata de descrição da língua.

O L-G estabelece a frase elementar como unidade mínima de análise e prevê a descrição da estrutura argumental dessas frases em matrizes binárias que relacionam os vários níveis linguísticos. Para proceder à análise e descrição dos itens da língua, o L-G toma como pressupostos os princípios da Gramática Transformacional (HARRIS, 1961), segundo a qual todas as frases de uma língua podem ser descritas a partir da descrição de suas frases simples e das operações de transformação que se estabelecem entre elas.

Outros princípios e pressupostos da gramática transformacional dão suporte a esta pesquisa e ao Modelo do L-G, bem como justificam a tomada de certas decisões que o linguista precisa fazer ao descrever uma língua. Assim, na seção 1.1, encontram-se as considerações sobre a gramática transformacional e a noção de transformação, e na seção 1.2 expõem-se as premissas e pressupostos do Modelo do Léxico-Gramática.

1.1 Harris e a gramática transformacional

A gramática transformacional foi proposta inicialmente por Harris (1951), a quem também são atribuídas as bases da Linguística de *Corpus* (HARRIS, 1952) e a proposta fundadora da Análise do Discurso (AD) de linha francesa. A análise transformacional das línguas foi introduzida no contexto da AD, visando descrever os processos de derivação de sentenças reais da língua a partir de sentenças de base (ou *kernel sentences*[5]).

Harris também foi pioneiro ao introduzir a representação algébrica e o tratamento matemático abstrato para a descrição de línguas, a partir dos anos 60. Como a Linguística é uma das únicas ciências que não possui uma metalinguagem própria, exterior à sua própria terminologia, a notação matemática seria mais apropriada do que a própria língua para descrever os fenômenos linguísticos. Assim, Harris (1968, 1982) propôs notações

5 A noção de *kernel sentence* foi bastante utilizada por Harris, durante toda a abordagem transformacional (dos anos 40 até o final dos anos 70). Harris se referia às sentenças de base (ou frases elementares) como *kernel sentences*. Essas frases do *kernel* da língua não são necessariamente aquelas frases atestadas na língua em uso, mas sim uma estrutura subjacente ao seu uso real, constituída por todos (e apenas) seus elementos necessários e suficientes. A partir de Harris (1982), o autor passou a adotar os termos *base sentences* (frases de base) ou *elementary sentences* (frases elementares) para designar essas frases.

algébricas e princípios matemáticos para descrever as estruturas linguísticas e identificar a informação veiculada pela língua.

Para a Gramática Transformacional, a informação (ou conteúdo informacional) é algo quantificável, baseado na organização das palavras entre si e na probabilidade de os elementos da língua se combinarem. Assim, a informação veiculada pela língua pode ser quantificada a partir de certas restrições combinatórias entre as unidades da língua; essas combinatórias não são definidas de forma aleatória nem equiprovável. A noção de informação perpassou toda a obra de Harris, mas foi mais explicitamente elaborada a partir da adaptação da teoria da informação das ciências exatas para a descrição de línguas (HARRIS, 1988, 1991).

1.1.1 A noção de operadores e argumentos e as restrições de seleção

Um dos conceitos-chave da gramática transformacional é a noção de *operadores* e *argumentos*, pois as frases de base só podem ser definidas pela combinatória de *operadores*, que exprimem tipicamente predicados, e as variáveis que os operadores selecionam, chamadas de *argumentos*. A presença de cada palavra numa frase pode estar relacionada ou não com as dependências entre outras palavras na frase. Assim, os operadores podem entrar numa frase independente da entrada de outras palavras anteriormente, enquanto os argumentos só entram se forem exigidos pelos operadores.

Os operadores são os elementos linguísticos (geralmente verbos ou adjetivos e nomes predicativos) que selecionam seus argumentos, seja para a posição de sujeito, seja para a posição de complemento. Alguns operadores (chamados operadores de primeira ordem) também podem funcionar como argumentos de outros operadores (chamados operadores de segunda ordem). Harris (1982, p.74-87) identifica, na gramática do Inglês, os seguintes operadores de primeira ordem: *On* (operadores com um argumento), *Onn* (operadores com dois argumentos) e *Onnn* (operadores com três ou mais argumentos). Os operadores de segunda ordem (*Oo*, *Oon* e *Ono*), que são aqueles que selecionam outros predicadores como argumentos, não serão analisados neste trabalho.

Já os argumentos são as demais palavras da língua, que dependem de outras palavras para serem selecionadas (HARRIS, 1982, p.70). "Nos termos da gramática tradicional, eles são, grosso modo, nomes de significado concreto, por exemplo: *livro*, *homem*, *lugar*, pronomes indefinidos, alguns nomes quantificadores e outros [...] mas não nomes predicativos, como *pai* em *Onn* e *ideia* em *Oo* ou operadores nominalizados, como

verdade[6]" (HARRIS, 1982, p.70). Apesar de Harris fazer menção somente aos operadores do Inglês, a relação de dependência entre os operadores e os argumentos é um princípio universal que regula a formação de sentenças de qualquer língua natural.

Essa noção de operadores e argumentos é fundamental para a proposta harrissiana e se opõe à visão das gramáticas tradicionais, que dividem as frases em sujeito e predicado. Para a gramática tradicional, as frases são compostas por dois constituintes básicos (o sujeito e o predicado) e a única relação entre esses constituintes é de ordem sintática, relativa à concordância do verbo com o sujeito. Em contrapartida, para a gramática transformacional, existem os operadores, que poderiam ser equiparados ao núcleo do predicado, e os argumentos, que equivalem às unidades lexicais que preenchem as posições sintáticas de núcleo do sujeito, do objeto direto, do objeto indireto *etc*.

A grande motivação para a gramática transformacional desconsiderar a divisão bipartida entre sujeito e predicado se deve ao fato de que o sujeito tem exatamente o mesmo estatuto que o objeto direto ou o objeto indireto (na nomenclatura tradicional), pois ambos são selecionados pelo verbo[7]. Em outras palavras, é o verbo quem impõe restrições sintáticas e semânticas para o preenchimento da posição de sujeito e de complementos. A abordagem proposta pela gramática transformacional (e que é utilizada pelo Modelo do Léxico-Gramática) também se baseia em uma divisão bipartida dos elementos linguísticos, mas considera a relação de dependência entre os elementos de forma diferente.

Em grande parte das abordagens linguísticas, convencionou-se considerar o verbo como o elemento central da predicação das frases, por exemplo, em *O Zé comeu o bolo*, é o verbo (*comer*) quem seleciona um nome humano (*Zé*) para a posição de sujeito e um nome de coisa comestível (*bolo*) para a posição de objeto. Atualmente, considera-se não apenas o verbo, mas também alguns adjetivos e nomes predicativos como os predicadores (ou operadores) das frases, pois são eles que selecionam seus argumentos, por exemplo, em *Ana é elegante*, é o adjetivo (*elegante*) quem seleciona o nome (*Ana*); já em *Ana é de uma elegância invejável*, é o nome (*elegância*) quem seleciona o nome (*Ana*).

6 Tradução minha. Do original: "In terms of traditional grammar they are roughly nouns of concrete meaning: for example *book*, *man*, *place*, indefinite pronouns, certain quantifier nouns [...] but not predicate nouns such as *father* in O_{nn} and *idea* in O_o or nominalized operators such as *truth*".

7 Referimo-nos ao verbo como predicador a título de exemplo, mas recorde-se que adjetivos e nomes predicativos também podem ser predicadores de frases.

Com base na combinatória entre os operadores e os argumentos, é possível chegar à informação. A informação, segundo a Gramática Transformacional, é calculada a partir das restrições de combinatórias estruturais, transformacionais e outras. Essas restrições são um princípio geral de todos os sistemas linguísticos e são elas que permitem que se façam generalizações sobre a relação de forma e significado. Harris (1988, p.10) elenca três restrições básicas comuns a todas as línguas:

- a ordenação parcial de palavras, ou seja, a ordem em termos de prioridade de entrada das palavras na frase;
- a desigualdade de probabilidade de palavras dentro da ordenação parcial;
- as reduções sentenciais.

Conforme aponta Batista (2008, p.63), a relação de dependência entre palavras, alicerçada na relação operador/argumento, está inclusa de forma essencial nessas restrições.

1.1.2 A frase elementar e a noção de *transformação* harrissiana

O conceito de *frase elementar* (ou *frase de base*) deve ser entendido não como uma estrutura linguística da *parole*[8], mas como um conceito subjacente à expressão linguística, que faz parte da *langue*. As frases que Harris considerou, desde os anos 40, como *kernel sentences* (ou frases de base) não são, necessariamente, frases diretamente observáveis em *corpus*. O conceito que utilizaremos de *frase de base*, no entanto, é apenas aproximativo do conceito teórico de *kernel sentence*, proposto por Harris (1946). As frases de base que construímos e analisamos representam a constituição básica de um predicado (contendo o predicador e seus argumentos obrigatórios).

A adoção desse ponto de vista teórico justifica a utilização de exemplos construídos na matriz (ver Apêndice A), em vez de exemplos retirados de *corpus*, apesar de todas as construções terem sido atestadas em *corpus*. As questões metodológicas relativas à construção de exemplos e aos testes de aceitabilidade serão apresentadas no Capítulo 2.

Às frases elementares podem ser aplicadas várias transformações para determinar a aceitabilidade ou não de alguma propriedade sintático-semântica. A noção de *transformação* aqui adotada foi proposta por Harris (1961), no sentido de explicar processos que nos permitem descrever as relações sintáticas entre as frases. Harris trata essas

8 Os termos *langue* e *parole* são atribuídos a Saussure (1916), que considerava que a *langue* é um sistema abstrato de regras que regem a organização de uma língua, enquanto a *parole* seria a realização das estruturas, a materialização concreta dessas regras na fala ou na expressão linguística.

transformações com relação às classes de equivalência, ou seja, as transformações são relações de equivalência parafrástica, desde que preservem as mesmas restrições de seleção lexical. Dito em outras palavras, existem diversas formas de veicular a mesma informação, mas elas não afetam o sentido da frase de base.

(1.1) *He is the father of three children.* [Ex.R]

≡ *He fathered three children.* [Ex.R]

Usando o mesmo exemplo de Harris (1964, p.414), o nome *father of* (*pai de*) é um operador que seleciona dois argumentos (*Onn*) e é do mesmo tipo do verbo *fathered*, que também é um operador sobre dois nomes (*Onn*). Em Português, não há um verbo correspondente a *to father*, mas Ranchhod (1990) usa uma adaptação do exemplo para:

(1.2) *Ele é padrinho de três crianças.* [Ex.R]

≡ *Ele apadrinhou três crianças.* [Ex.C]

Esse exemplo demonstra uma relação de equivalência parafrástica entre duas formas da mesma frase de base. A diferença entre elas está no operador, pois, enquanto o verbo *apadrinhar* pode conter, na sua própria morfologia, os afixos verbais (pessoa, número, tempo, modo e aspecto), o nome *padrinho* não pode, exigindo assim o verbo *ser*, para "carregar" essas marcas.

Ao propor a análise das construções a partir de frases simples, a Gramática Transformacional estabelece que é necessário transformar as sentenças que compõem o *corpus* em sentenças simples, sem alterar o significado com que elas apareceram nos dados reais. Como as construções são retiradas de *corpus*, a primeira transformação a ser aplicada em todas as sentenças é a simplificação das frases complexas. Entende-se que, em um *corpus* com textos reais, podem aparecer diferentes tipos de frases, desde as mais simples até os mais complexos períodos compostos, portanto, há a necessidade de transformar as frases complexas em frases simples constituídas de apenas um núcleo predicativo e seus argumentos obrigatórios. Para esclarecer o tipo de transformação, citemos como exemplo uma frase retirada do *corpus*:

(1.3) *David foi até ele e lhe deu um brinquedo de plástico com que as crianças estavam brincando.* [Ex.R]

A sentença sofreu algumas alterações: (i) apagamento da oração relativa restritiva (*com que as crianças estavam brincando*); (ii) apagamento da oração coordenada aditiva (*David foi até ele e*); e (iii) apagamento do adjunto adnominal (*de plástico*), pois julga-se esse tipo de informação desnecessária para a compreensão do sentido principal da frase, que passa a ser considerada em sua forma simples:

(1.3a) *David deu um brinquedo ao menino.* [Ex.C]

Pode-se notar também que foi aplicada outra transformação que se refere à substituição do clítico *lhe* em (1.3) pelo sintagma preposicionado *ao menino* em (1.3a). As transformações são, em geral, operações de redução ou apagamento, para que a frase elementar contenha todos – e, ao mesmo tempo, apenas – os seus elementos obrigatórios e necessários para a compreensão do predicado.

À frase elementar (1.3a) podem ser aplicados outros testes para verificar a aceitabilidade ou não de certas propriedades, como as regras transformacionais de pronominalização, de apassivação, de relativização, de modalização, de apagamento do sujeito, dentre outras. Essas transformações, além de não alterarem o significado global da frase, também não alteram as restrições distribucionais da frase de base.

As regras de transformações aplicáveis às frases também seguem determinados padrões, pois não é qualquer transformação que se pode aplicar às sentenças; aplicam-se apenas aquelas estritamente necessárias para o julgamento de aceitabilidade de uma ou outra propriedade sintática.

A proposta de aplicar regras de transformação às frases para testar e descrever suas propriedades é o que garante o caráter experimental da pesquisa. Segundo Gross (1975, p.19), "a construção de exemplos e contraexemplos constitui a atividade experimental do linguista que verifica a teoria de certos fenômenos"[9].

As relações que se formam entre uma sentença e outra são relações de equivalência, ou seja, elas são alteradas sintaticamente, mas mantêm uma equivalência semântica, como se pode observar pelas frases:

(1.3a) *David deu um brinquedo ao menino.* [Ex.C]

≡ *Um brinquedo foi dado ao menino por David.* (apassivação) [Ex.C]

≡ *David deu-lhe um brinquedo.* (pronominalização do SP) [Ex.C]

≡ *David deu-o ao menino.* (pronominalização do SN) [Ex.C]

As duas primeiras frases exemplificam a mudança da frase na voz ativa para a voz passiva, ou o inverso. Essas transformações são chamadas, respectivamente, de apassivação da voz ativa e ativização da voz passiva, e podem ser assim formalizadas:

9 Tradução minha. Do original: "La construction d'exemples et de contre-exemples constitue l'activité expérimentale du linguiste qui vérifie la théorie de certains phénomènes" (GROSS, 1975, p.19).

*N*0 *V N*1 *prep N*2: *David deu um brinquedo ao menino.*

*N*1 *Vser V part prep N*2 *prep N*0: *Um brinquedo foi dado ao menino por David.*

Na transformação de ativa para passiva, os elementos *N*0 e *N*1 mudaram de posição sintática, mas permaneceram com os mesmos papéis temáticos: AGENT-GIVER (*David*), OBJECT-GENERIC (*brinquedo*) e RECIPIENT (*menino*). Foram ainda acrescentados: o verbo *ser*, a terminação de particípio ao verbo principal e a preposição *por* antecedendo o argumento AGENT-GIVER, que funciona como agente da passiva.

As transformações foram propostas inicialmente por Harris (1961) como relações orientadas, por exemplo, da ativa para a passiva ou, no caso das nominalizações, da construção verbal para a construção nominal.

(1.4) *He studies eclipses.* → *He makes studies of eclipsis.* [Ex.R] *Ele estuda eclipse.* → *Ele faz estudos sobre eclipse.* [Ex.C]

(1.5) *He loves Italy.* → *He is in love with Italy.* [Ex.R]

Ele adora a Itália. → *Ele tem uma adoração enorme pela Itália.* [Ex.C]

O que indica a orientação das relações é a orientação da seta (→). Essa relação unidirecional foi proposta por Harris, em 1968, e foi retomada em 1976, mas as transformações passaram a ser vistas como relações não-orientadas, ou seja, não é necessariamente a ativa que dá origem à passiva ou vice-versa; as duas frases são equivalentes. Conforme aponta Ranchhod (1990), as frases transformadas passaram a ser ligadas por relações de equivalência (de tipo algébrico), não-orientadas, que são representadas pela seta bidirecional (⟷).

1.1.3 As transformações unárias

Apesar de muitas descrições da gramática transformacional lidarem com frases complexas e com as transformações binárias, neste trabalho seguiremos a abordagem do Léxico-Gramática, que optou por descrever somente frases simples (ou frases elementares) e suas transformações unárias.

São consideradas transformações unárias aquelas que incidem sobre uma frase elementar originando sempre uma outra frase elementar, ou seja, que alteram a frase de base, tais como, a apassivação, a reestruturação dativa (LECLÈRE, 1995, p.179-198), a permuta de comprimento (HARRIS, 1976, p.148-149), a pronominalização, a desfinitização (CASTELEIRO, 1981, p.245-265), a extração (GROSS, 1981, p.19) e o apagamento de elementos. Por outro lado, são consideradas transformações binárias

aquelas que combinam duas estruturas elementares para dar origem a uma estrutura complexa, que é uma outra frase resultante da combinação das duas formas frásicas simples iniciais, tais como, as coordenações e as subordinações.

Para as construções que pretendemos descrever neste trabalho, considera-se que não há diferenças entre as frases simples e as frases complexas. Na verdade, a adoção de frases complexas poderia acrescentar "ruídos" ou outros elementos indesejáveis para a análise, tirando o foco das estruturas realmente necessárias. Outra razão para a escolha de frases simples como objeto de análise se justifica pelo princípio da recursividade linguística. Gross (1975, p.17-18) realizou cálculos sobre a quantidade de possibilidades de construções frásicas em dada língua, considerando um vocabulário ínfimo de 20.000 palavras e chegou à conclusão de que um falante é capaz de produzir 1.050 frases com 20 palavras cada. Se a possibilidade de produção de frases simples se aproxima de 1.050, esse número chegaria próximo do infinito se fossem levadas em conta as frases complexas (períodos compostos).

Seguindo, portanto, a mesma opção teórica e metodológica do Léxico-Gramática, consideraremos como objeto de análise linguística a frase de base (ou frase elementar), constituída de sujeito, verbo e complementos essenciais (quando houver). A estrutura sintática das frases é assim determinada especialmente pelas propriedades combinatórias do léxico que as constitui, ou seja, existe uma estreita relação entre os níveis lexical, sintático e semântico.

Todas as transformações unárias seguem um princípio geral, identificado por Gross (1975, p.27) como princípio da invariância morfêmica, segundo o qual os morfemas da frase que sofreu transformação devem se assemelhar aos morfemas da frase-base que foi transformada. Em outras palavras, quando duas frases que possuem o mesmo sentido estão relacionadas por alguma transformação, os morfemas semanticamente plenos (verbos, substantivos e adjetivos, por exemplo) devem ser os mesmos nas duas frases. Os únicos morfemas que aceitam variação são os morfemas presos e dependentes, tais como, os afixos, as desinências, as preposições etc. As frases podem variar quanto a aspectos formais, estruturais ou até morfológicos, mas os morfemas livres, que constituem o elemento principal da significação, não variam. O princípio da invariância morfêmica será, em larga escala, adotado neste trabalho.

1.2 O modelo do Léxico-Gramática

As gramáticas tradicionais do Português costumam adotar uma abordagem gramatical bastante antiga, baseada na classificação dos elementos linguísticos em categorias

independentes e autônomas. A proposta do Léxico-Gramática vai de encontro a essa visão, por considerar que todos os níveis estão integrados e devem ser analisados de forma conjunta, numa perspectiva em que o léxico, a morfologia, a sintaxe e a semântica não são concebidos como componentes independentes e autônomos, mas como uma rede em que tais elementos se instauram mutuamente.

O modelo do Léxico-Gramática foi proposto por Gross (1968, 1975) e tem suas bases no Distribucionalismo e no modelo Transformacional de Harris (1955, 1961), abordado na seção anterior. O nível de representação proposto pela gramática transformacional foi submetido a trabalho experimental de linguistas e cientistas da Computação por vários anos e para várias línguas. A validação dos trabalhos experimentais para uma grande quantidade de dados em várias línguas nos permite afirmar que o Modelo do L-G não é apenas uma metodologia, mas um Modelo Teórico-Metodológico.

A despeito de outras designações para o modelo do Léxico-Gramática, por exemplo, teoria, abordagem, gramática ou método, adotaremos a designação "Modelo do Léxico-Gramática", por considerarmos que o L-G seja, ao mesmo tempo, uma proposta teórica e metodológica, ou seja, uma representação capaz de relacionar noções teóricas mais gerais com procedimentos experimentais de cada língua.

A ideia proposta pelo modelo é de criar uma "gramática do léxico". Assim, parte-se dos verbos (ou outros elementos predicativos, tais como, adjetivos e substantivos) e descrevem-se as propriedades sintáticas e semânticas das frases de base de uma língua. Proposto inicialmente para o Francês, o Léxico-Gramática[10] vem ganhando espaço no quadro teórico de descrição de várias outras línguas, tais como, Italiano, Espanhol, Português, Romeno, dentre outras.

A ideia introduzida por Gross (1968, 1975, 1981) de conjugar léxico e gramática na análise e descrição das línguas parte do princípio de que os elementos lexicais podem se combinar gramaticalmente de formas bastante específicas, ou seja, nem todo item lexical pode ocupar qualquer posição sintática, devendo adequar-se a determinadas restrições semânticas que a sintaxe lhe impõe, e a restrições sintáticas que a semântica lhe impõe. Essa ideia decorre da segunda restrição básica de Harris (1988, p.10), em relação à desigualdade de probabilidade de palavras dentro da ordenação parcial, segundo a qual as combinatórias entre as palavras não são definidas de forma aleatória nem equiprovável.

10 A proposta foi um empreendimento de laboratórios e grupos de pesquisas que constituem a rede RELEX. Grande parte do trabalho prático desse modelo foi desenvolvido pelos integrantes do LADL (*Laboratoire d'Automatique Documentaire et Linguistique*), que fizeram a coleta, análise e classificação de verbos, descrevendo-os nos moldes de tábuas do L-G, inicialmente em Francês.

Tome-se a palavra *corrida* como exemplo. Assim como vários outros itens lexicais de qualquer língua, a palavra *corrida* pode apresentar diferentes significados dependendo do contexto linguístico no qual está inserida. Observe-se, portanto:

(1.6) *Diógines Alves ainda pagou a corrida do taxista.* [Ex.R]

(1.7) *Dia 31 de dezembro acontecerá a corrida de São Silvestre.* [Ex.R]

(1.8) *Tenho uma vida muito corrida.* [Ex.R]

(1.9) *Dica de como aplicar massa corrida na parede.* [Ex.R]

(1.10) *O jogador do São Paulo deu uma corrida para manter a forma.* [Ex.R]

A mesma forma lexical *corrida* deve ser classificada morfossintaticamente de maneiras diferentes porquanto seus significados também se diferenciem. Conforme atesta Ranchhod (1993, p.110), "as palavras isoladas contêm em si mesmas certas potencialidades semânticas, mas só adquirem uma significação precisa se combinadas com outras numa forma sintaticamente adequada".

Apesar de o Léxico-Gramática não ter surgido no escopo de uma teoria, identificamos (RASSI, 2008) alguns princípios teóricos básicos que regem todos os procedimentos experimentais de descrição de língua nos moldes do L-G. Nas subseções seguintes, são apresentadas três premissas em que se fundamenta o Modelo do L-G.

1.2.1 Premissa 1: Cada unidade lexical tem uma gramática própria

O modelo do L-G defende que, dentro da gramática da língua, existe uma gramática própria para cada item lexical, e o trabalho do linguista, nesse sentido, é descrever cada uma dessas gramáticas. Considere-se, por exemplo, a diferença entre os verbos *morrer* e *falecer*. Ambos possuem significados semelhantes, mas cada um tem uma gramática própria:

(1.11) *(A planta + O Pedro) morreu.* [Ex.C]

(1.12) *(*A planta + O Pedro) faleceu.* [Ex.C]

Considera-se que cada unidade lexical possui uma gramática própria porque impõe restrições de seleção específicas ao preenchimento lexical de seus argumentos. O verbo *morrer* admite como sujeito qualquer entidade viva, incluindo pessoas, animais e plantas, ou ainda metaforicamente admite entidades abstratas (*e.g. A esperança morreu*), ao passo que o verbo *falecer* somente admite nomes humanos ou humanizados na mesma posição.

Cada item lexical da língua age e funciona de forma específica, seguindo determinados comportamentos. A aceitabilidade ou inaceitabilidade das construções linguísticas é determinada pelo funcionamento de cada item lexical ou expressão da língua.

As primeiras investigações sistemáticas da gramática do léxico das línguas naturais permitiram determinar a complexidade da distribuição de suas propriedades formais; e essa complexidade cresce progressivamente, na medida em que aumenta o número de estudos sobre os elementos do léxico (GROSS, 1975). Gross (1981, p.10), referindo-se ao L-G dos predicados verbais em Francês, ressalta que "uma comparação entre as linhas dessa matriz revela que elas são todas diferentes par a par, ou seja, que não existem dois verbos com o mesmo conjunto de propriedades sintáticas"[11].

Ressalte-se ainda que essa premissa não impede que as unidades lexicais ou as construções linguísticas sejam agrupadas em classes. A classificação que se faz é apenas um agrupamento aproximativo dos elementos, com base em suas características ou propriedades mais proeminentes. Os verbos *correr* e *disparar*, por exemplo, parecem muito semelhantes tanto do ponto de vista sintático quanto semântico e, por isso, certamente seriam classificados num mesmo grupo, porém suas construções não são idênticas. Uma prova de que cada uma dessas lexias possui uma gramática própria é que é possível construir *Rui correu os olhos pelo documento*, mas não **Rui disparou os olhos pelo documento*. A premissa de que cada unidade lexical tem uma gramática própria, segundo Vale (2001), acarreta uma diferença metodológica fundamental: ao invés do método hipotéticodedutivo preconizado pela Gramática Gerativa, o Léxico-Gramática assume claramente uma postura taxonômica. Assim, a taxonomia torna-se um meio para encontrar as regularidades da língua.

1.2.2 Premissa 2: A unidade mínima de análise é a frase elementar

Gross (1975) considera, no modelo do L-G, que a unidade mínima de análise gramatical deve ser a frase elementar, e não uma palavra ou um morfema. Essa opção teórica é justificada por dois argumentos: i) o estudo de uma palavra isolada priva o descritor da possibilidade de avaliar aceitabilidades, portanto, a palavra deve estar inserida numa frase para que possa ser analisada; ii) numa frase elementar, o contexto remove muitas vezes a ambiguidade da palavra isolada.

A esse respeito, Lamiroy (1998, p.10) explica que "Uma das ideias fundamentais do Léxico-Gramática é ligar a frase elementar e o léxico, pela simples razão de que as

11 Tradução minha. Do original: "Une comparaison des lignes de cette matrice révèle qu'elles sont toutes différentes deux à deux, c'est-à-dire qu'il n'exist pas deux verbes ayant le même ensemble des propriétés syntaxiques" (GROSS, 1981, p.10).

entradas lexicais instauram seus satélites de modo mais evidente nos contornos de uma frase simples"[12].

Uma palavra isolada nem sempre é passível de interpretação unívoca, pois o significado de uma palavra só pode ser determinado pela vizinhança em que ela se encontra. Um exemplo simples é o do verbo *bater*, cujos empregos são determinados pelo tipo de argumentos que o acompanha, bem como pelas restrições distribucionais ao preenchimento lexical dessas posições sintáticas:

(1.13) *O Rui bateu na porta.* [Ex.C]

(1.14) *O Rui bateu na Ana.* [Ex.C]

Em (1.13), o verbo *bater* seleciona um nome humano (*O Rui*) para a posição de sujeito e um nome não-humano (*na porta*) para a posição de complemento, que é introduzido pela preposição *em*, ao passo que, em (1.14), o verbo *bater* seleciona um nome humano (*O Rui*) para a posição de sujeito e outro nome humano (*na Ana*) para a posição sintática de complemento, também introduzido pela preposição *em*. O que distingue os sentidos do verbo *bater*, como '*knock*' em (1.13) e como '*hit*' em (1.14), são as restrições de seleção dos argumentos. Gross (1981, p.21) acrescenta ainda que:

> Por convenção, as unidades de base do significado seriam realizadas pelas frases simples. A descrição semântica consistiria, portanto, em decompor as frases complexas segundo as frases simples de base, por isso não é diferente da descrição sintática (Gross, 1978). Essa posição é a de Harris (1968, 1976, 1978), que ela dificilmente poderia ser posta em causa se os métodos sintáticos se revelassem importantes para reduzir frases complexas em frases simples[13].

1.2.3 Premissa 3: Os testes de aceitabilidade são feitos com recurso massivo à introspecção

No que se refere à obtenção dos dados linguísticos a serem descritos, há duas formas possíveis na literatura: (i) recorrendo-se a *corpus*; e (ii) por meio da introspecção, ou seja, o conhecimento linguístico do falante. Segundo Laporte (2015), ambas são úteis já

12 Tradução minha. Do original: "Une des idées fondamentales du lexique-grammaire est en effet de lier phrase élémentaire et lexique, pour la simple raison que les entrées lexicales étalent leurs satellites de façon la plus évidente dans contourns de la phrase simple" (LAMIROY, 1998, p.10).

13 Tradução minha. Do original: "Par convention, les unités de base du sens seraient portées par ces phrases simples. La description sémantique consisterait donc à décomposer les phrases complexes selon les phrases simples de base, elle ne différerait donc guère de la description syntaxique (Gross, 1978). Cette position est celle de Harris (1968, 1976, 1978), elle ne pourrait guère être remise en cause que si les méthodes syntaxiques s'avéraient impuissantes à réduire les phrases complexes à des phrases simples" (GROSS, 1981, p.21).

que elas abarcam dois aspectos da realidade e uso linguístico. A utilização de *corpora* é importante para as formas que poderiam passar despercebidas pelo conhecimento do linguista, enquanto a introspecção é necessária para distinguir entre formas raras e aquelas que não estão em uso.

As duas formas (introspecção e observação em *corpus*) serão usadas neste trabalho para recensear os dados. Por outro lado, no que se refere ao julgamento de aceitabilidade das frases, o L-G requer um recurso massivo à introspecção.

Gross (1975, p.27) considera que a utilização apenas de *corpus* não permite uma análise fina da aceitabilidade das construções, pois pode não aparecer no *corpus* algum tipo frásico que existe na língua. Nesse sentido, é recomendável o uso da introspecção do linguista nativo. "A aceitabilidade é então uma noção muito complexa, que abrange as intuições de forma e de sentido, e que depende de inúmeros fatores culturais"[14] (GROSS, 1975, p.23).

O julgamento daquilo que se considera aceitável ou inaceitável na língua depende de dois critérios básicos: (i) a frase deve ser pronunciável e (ii) a frase deve ser imediatamente interpretável. Esse segundo critério permite que façamos paráfrases com elementos linguísticos semelhantes, a fim de comparar as estruturas. O fato de ser imediatamente interpretável não exclui os casos de ambiguidade de interpretação; pelo contrário, uma mesma frase pode ter diversos significados, e todas essas construções serão aceitáveis, desde que sejam pronunciáveis e imediatamente interpretáveis.

Laporte (2015) reformula o conceito de *aceitabilidade*, dizendo que: "Para ser aceitável, uma forma tem que ser significativa. Quando linguistas julgam a aceitabilidade de uma forma, eles julgam a probabilidade de aquela forma ser usada em algum contexto para veicular informação"[15]. O autor ainda complementa que:

> A aceitabilidade é uma forma simplificada de probabilidade: uma sequência inaceitável é improvável de ocorrer, seja no discurso ou em qualquer outro lugar. Considerando-se que as probabilidades pertencem a uma escala contínua, a realidade linguística é mais complexa do que qualquer coisa que uma visão binária de aceitabilidade pudesse sugerir. Na prática, não há nenhuma maneira de medir a probabilidade de qualquer sequência em uma língua[16]. (LAPORTE, 2015).

14 Tradução minha. Do original: "L'acceptabilité est une notion très complexe qui comporte des intuitions de forme et de sens, et qui dépend de nombreux facteurs culturels" (GROSS, 1975, p.23).

15 Tradução minha. Do original: "To be acceptable, a form must be meaningful. When linguists assess the acceptability of a form, they assess the probability that it might be used in some context to convey information" (LAPORTE, 2015).

16 Tradução minha. Do original: "Acceptability is a simplified form of probability: an unacceptable sequence is unlikely to occur, whether in discourse or anywhere else. Since probabilities belong to a continuous scale, linguistic reality is more complex than anything a binary view of acceptability might suggest. In practice there is no way to measure the probability of any sequence in a language" (LAPORTE, 2015).

Apesar dos problemas conceituais relacionados ao termo *aceitabilidade*, este ainda é o critério científico mais utilizado pelo Modelo do L-G. Ademais, o julgamento das aceitabilidades deve ser feito com recurso massivo à introspecção, sem descartar o uso de *corpus*. A noção de introspecção também é um conceito importante para a linguística teórica e não se confunde com a noção de *intuição* ou "*achismo*". Conforme aponta Laporte (2015), a tarefa empírica de coletar dados introspectivamente segue sistematicamente procedimentos formais controlados.

Não se pode confundir a noção de *aceitabilidade* com a concepção de *gramaticalidade* porque essa segunda lida com o que é considerado certo e errado pelas gramáticas normativas, enquanto a aceitabilidade diz respeito às leis que regem a competência do falante. Dessa forma, um falante nativo seria capaz de dizer *Eu lavei as vasilhas sujas*, mas certamente não diria, em condições normais, **Vasilhas as eu lavou suja*, que é considerada uma frase inaceitável porque, apesar de pronunciável, não é imediatamente interpretável.

O conceito de (in)aceitabilidade é fundamental para o L-G porque é esse conceito que delimita o objeto de estudo: só podem ser analisadas e descritas as construções que, de fato, existem na língua.

Há na literatura trabalhos que se basearam no Modelo teórico-metodológico do Léxico-Gramática, porém optaram por utilizar apenas *corpus* para validar os exemplos (RADIMSKY, 2011, p.181): "o 'teste de aceitabilidade', considerado como o principal meio de verificação por Gross (1975, p.19-22), foi substituído por um recurso sistemático a grandes *corpora*, o *Corpus* Nacional Tcheco"[17]. Apesar dos resultados satisfatórios apresentados pelo autor, ainda assim consideramos que nem todas as possibilidades de construções linguísticas estão no *corpus*, por maior que ele seja.

Seguindo a abordagem geral proposta pelo Léxico-Gramática, os dados lexicais a serem analisados nesta pesquisa serão retirados de *corpus*, porém a verificação de aceitabilidade passará também pelo crivo introspectivo da pesquisadora. Em suma: (i) recorre-se a *corpus* para recensear os dados linguísticos ou o fenômeno a ser descrito; (ii) propõe-se a análise e descrição das entradas lexicais a partir da aplicação de regras de transformação sintática às frases simples; e, por fim, (iii) verifica-se a aceitabilidade ou inaceitabilidade das frases.

17 Tradução minha. Do original: "'test de l'acceptabilité', considéré comme le moyen principal de vérification par M. Gross (1975: 19-22), a été remplacé par un recours systématique aux grands corpus, le Corpus national tchèque en l'occurrence" (RADIMSKY, 2011, p.181).

Neste capítulo, apresentamos as duas principais teorias que norteiam esta pesquisa: a teoria transformacional e o Modelo teórico-metodológico do Léxico-Gramática. A teoria transformacional dá as bases para o surgimento do Léxico-Gramática, a partir dos conceitos de operadores e argumentos, das restrições de seleção que o operador impõe aos seus argumentos e das noções de frase elementar e de transformação. A partir dessa teoria, surge então o Modelo do L-G, que é ao mesmo tempo uma teoria e uma metodologia de descrição de línguas, que parte de três premissas fundamentais. São elas: (i) cada unidade lexical de uma língua possui sua própria gramática, ou seja, funciona de um modo particular na combinação com os outros elementos da língua; (ii) a unidade mínima de análise linguística deve ser a frase simples, que é constituída por um predicador e os argumentos essenciais que ele seleciona; e (iii) os testes para verificar a aceitabilidade de determinada construção ou fenômeno devem ser feitos usando a introspecção do linguista falante nativo da língua a ser descrita.

A parte teórica do Modelo do L-G foi descrita neste capítulo. A parte que concerne à metodologia do L-G, bem como toda a metodologia utilizada neste livro, será apresentada no capítulo seguinte.

2. Princípios e procedimentos metodológicos

Esta seção apresenta: (i) o *corpus* utilizado para a obtenção dos dados (seção 2.1); (ii) as ferramentas computacionais de auxílio ao recenseamento, processamento e análise das ocorrências (seção 2.2); (iii) os procedimentos experimentais utilizados na análise (seção 2.3); (iv) o modelo de formalização dos dados em matrizes (seção 2.4); e (v) a proposta metodológica para a integração dos predicados nominais com o verbo *dar* em um *parser* (seção 2.5).

As construções foram recenseadas a partir de três fontes diferentes (seção 2.1): um *corpus* com textos reais (subseção 2.1.1), uma lista de colocados em vários *corpora* (subseção 2.1.2) e uma lista pré-construída de *Npred* (subseção 2.1.3). Para a extração dos dados do *corpus* em textos reais, utilizamos a ferramenta Unitex (seção 2.2.1). Para o recenseamento dos colocados do verbo *dar* em vários *corpora*, utilizamos o AC/DC[18]. Para a análise das propriedades de algumas construções, utilizamos, além da introspecção, o concordanciador WebCorp, que usa a *web* como *corpus* (seção 2.2.2).

Após o recenseamento dos colocados de *dar*, construímos frases simples para cada uma das construções encontradas, para que se pudesse proceder à análise de suas propriedades. Em seguida, separamos as construções em diferentes classes (ver Parte I), com base em suas propriedades sintáticas e semânticas. A análise propriamente dita fundamenta-se nos princípios teóricos e metodológicos do Léxico-Gramática (GROSS, 1975, 1981), que tem como base a frase simples.

A análise foi feita com base nas propriedades formais, distribucionais e transformacionais das construções (seção 2.3), e a descrição dessas propriedades foi feita em matrizes binárias, que é a formalização proposta pelo Modelo do L-G (seção 2.4). Maurice Gross (1975) considerava "falhas" as tentativas de formalização das regras gramaticais anteriores justamente porque aquelas propostas não previam a relação intrínseca que existe entre o léxico e a gramática que o descreve, por isso, propôs a metodologia do L-G, que tem como pano de fundo os procedimentos experimentais da gramática transformacional de Zellig Harris (1961, 1964, 1968).

Vale ressaltar que, mesmo que se tenha adotado *corpora* para a obtenção dos dados, é possível mesclar essa metodologia com o método introspectivo de análise. Algumas vezes, criam-se hipóteses que podem ser comprovadas por meio de exemplos construídos, ou seja, que não foram retirados do *corpus*, porém já têm seu uso consagrado na língua. As duas abordagens serão usadas simultaneamente para garantir uma descrição mais completa e abrangente do objeto de estudo.

18 Disponível para consulta em: http://www.linguateca.pt/ACDC/.

Por fim, apresentaremos também uma proposta de integração dos predicados nominais formados pelo verbo-suporte *dar* e um nome predicativo no *parser* XIP (MOKHTAR; CHANOD; ROUX, 2002), que é o analisador sintático usado na cadeia de processamento do Português STRING (MAMEDE *et al.*, 2012). Na seção 2.5, apresentaremos brevemente a metodologia utilizada para a integração desses dados no XIP, porém os resultados da tarefa, bem como a avaliação da performance do sistema, serão apresentados na parte III.

2.1 Obtenção dos dados

Os dados a serem analisados neste trabalho são provenientes de três fontes distintas, que constituem nosso *corpus*: (i) o *corpus* PLN.Br Full; (ii) uma busca por frequência dos colocados do verbo *dar* nos *corpora* disponíveis no AC/DC para o Português Brasileiro; e (iii) uma lista de nomes predicativos do Português Europeu (VAZA, 1988). Cada uma das fontes será descrita nas próximas subseções.

2.1.1 Corpus PLN.Br Full

As ocorrências do verbo *dar* foram buscadas inicialmente no *corpus* PLN.Br Full, um resultado da criação do Portal de *Corpus* (MUNIZ *et al.*, 2007), com cooperação multi-institucional (USP, UFSCar, Unisinos, PUC-RS, PUC-Rio, Mackenzie University, UNESP). O PLN.Br Full é um grande *corpus* do gênero informativo, subgênero jornalístico, que já gerou dois *subcorpus*: o PLN.Br Categ e o PLN.Br Gold, e foi anotado por um grupo de pesquisadores do Núcleo Interinstitucional de Linguística Computacional (BRUCKSCHEN *et al.*, 2008).

Atualmente esse *corpus* conta com 103.080 mil textos da *Folha de São Paulo*, um dos principais jornais brasileiros, e 29.014.089 *tokens*, dos anos de 1994 a 2005. Os doze anos do jornal (1994-2005) foram agrupados em um único arquivo para facilitar o processamento. Utilizando-se a ferramenta Unitex (ver seção 2.2.1), constatou-se que o *corpus* apresenta 66.798 ocorrências (*tokens*) de *dar*, conjugado ou não, em todas as formas do lema. Todas essas ocorrências foram conferidas manualmente, uma a uma, a fim de proceder a uma primeira etapa de classificação.

Para essa primeira etapa, foram consideradas 3.560 construções diferentes (*types*) com o verbo *dar*, as quais foram, posteriormente, distribuídas em diferentes classes (classe dos nomes predicativos associados ao verbo-suporte, classe dos nomes predicativos associados ao verbo-operador causativo, classe das expressões fixas, dentre outras).

2.1.3 Lista de colocados no AC/DC

Além do PLN.Br Full, utilizamos também o projeto Acesso a *Corpus*/Disponibilização de *Corpus* (AC/DC), da Linguateca, que permite buscas rápidas pela distribuição dos lemas de nomes associados ao lema do verbo *dar*. Uma das vantagens em consultar as ocorrências pelo AC/DC consiste no fato de que é possível buscar a expressão ou padrão que se deseja em vários *corpora* ao mesmo tempo, ou especificar apenas *corpus* de uma das variantes do Português.

Para proceder à consulta, testamos dois tipos de buscas, utilizando duas expressões regulares diferentes: (i) [lema="dar"]@[pos="N.*"], que busca o substantivo imediatamente subsequente ao verbo *dar*, flexionado ou não, o que resultou em 1.717 lemas diferentes (*types*); e (ii) [lema="dar"] [ ]{0,3} @[pos="N.*"], que também busca o substantivo que ocorre após o verbo *dar*, mas admitindo uma janela de 0 até 3 palavras entre eles. Como resultado da busca da segunda expressão regular, obtivemos 2.452 lemas de substantivos[19], utilizando os mesmos *corpora*. Decidimos por adotar a segunda listagem, por prover mais dados da língua.

As buscas das duas expressões regulares foram feitas em todos os *corpora* exclusivamente do Português Brasileiro, disponíveis no AC/DC, a saber: AmostRA-NILC (98.633 *tokens*), ANCIB (1.257.109 *tokens*), *Corpus* Brasileiro (977.944.871 *tokens*), C-Oral-Brasil (263.396 *tokens*), ECI-EBR (723.995 *tokens*), FrasesPB (19.162 *tokens*), OBras (1.201.787 *tokens*), ReLi (134.027 *tokens*) e NILC/São Carlos (32.461.799 *tokens*).

Listados por frequência, os 2.452 nomes foram analisados e, posteriormente, classificados (ver Parte I).

2.1.3 Lista de *Npred* em PE

Para a parte II, que trata especificamente das construções nominais com o verbo-suporte *dar* e um nome predicativo, utilizamos também uma lista com 1.466 nomes predicativos que ocorrem com o verbo-suporte *dar* em Português Europeu (VAZA, 1988). Nem todos aqueles, no entanto, existem em Português Brasileiro; citem-se como exemplos os nomes *marrada*, *rapadela*, *tareia*, *turra*, *tau-tau*, *zaragata*, dentre vários outros.

Da intersecção entre as listas de *Npred* em PB (retiradas manualmente do *corpus* PLN.Br Full e do AC/DC - ver subseção 2.3.2) e PE (retiradas da lista de Vaza (1988)), cerca de 520 pertenciam às duas variantes. Alguns *Npred* que constavam da lista de

19 Vale ressaltar que o AC/DC restringe a busca por distribuição de lemas em, no máximo, mil formas diferentes do lema por *corpus*. Isso significa que, para os dois *corpora* mais extensos da coleção - *Corpus Brasileiro* e *NILC/São Carlos* - o AC/DC não retorna todos os resultados, mas apenas os 1.000 mais frequentes, o que pode ter nos privado de encontrar alguns nomes com baixa frequência.

Vaza (1988) foram desconsiderados por não se tratarem de nomes predicativos, por exemplo, *dar à-vontade*, ou nomes que, apesar de serem predicativos, ocorrem com o verbo *dar* em construções causativas, como *dar alegria, alergia, medo, etc.*, os quais serão analisados no Capítulo 5, mas não como nomes predicativos de frases de base. Outros *Npred* que, sob o julgamento da pesquisadora, não pertenciam à variante brasileira, foram consultados na *web*, em páginas com extensão .br, utilizando-se os motores de busca Yahoo e Google. Aqueles que pertenciam ao PB, foram inseridos na base de dados.

Recorremos ao ViPEr (BAPTISTA, 2012), que é o Léxico-Gramática das construções verbais do Português Europeu, para encontrar candidatos de verbos que poderiam ser nominalizados em *-da*, tais como, *andada*, *corrida*, *passeada etc*. A versão consultada do ViPEr (versão 1.3.6) conta com cerca de 6.300 verbos, dos quais aproximadamente 500 admitem nominalização em *-da*. Dessa forma, alguns *Npred* que não haviam sido encontrados no *corpus* PLN.Br Full, associados ao verbo *dar*, tais como, *programada*, *progredida* e *proliferada*, foram acrescentados à lista de nomes predicativos a serem descritos.

2.2 Ferramentas de busca e processamento

2.2.1 A ferramenta Unitex

Para recensear todas as ocorrências do verbo *dar*, foi utilizada a ferramenta Unitex 3.1 Beta, que permite a busca e o processamento de qualquer lexia em grandes *corpora* em tempo real. As informações linguísticas estão disponíveis no Unitex por meio de dicionários eletrônicos e gramáticas, representados por autômatos de estados finitos.

O Unitex dispõe de dicionários eletrônicos de todas as línguas em que ele funciona (Inglês, Finlandês, Francês, Alemão, Grego, Coreano, Italiano, Norueguês, Polonês, Português Brasileiro, Português Europeu, Russo, Espanhol, Sérvio e Tailandês). A ferramenta faz distinção entre palavras simples e compostas, e entre formas canônicas (lematizadas) e as formas flexionadas. Além disso, os dicionários também contêm informações morfológicas e sintáticas, tais como, categorias gramaticais (*part-of-speech - POS*) e etiquetas morfológicas de lema, gênero, número, grau, pessoa, tempo e modo.

A partir do comando <dar>, o programa recenseia todas as ocorrências das formas flexionadas do verbo *dar*. Um dos problemas inerentes a essa busca é que o Unitex não faz análise sintática para etiquetar as *POS*, o que pode gerar uma grande quantidade de

informações desnecessárias para esta pesquisa. Das 66.798 ocorrências de *dar* no *corpus*, cerca de 30.000 não deveriam ser classificadas como verbos, e sim como substantivo, preposição ou pronome, tais como nos exemplos seguintes:

(2.1) *A história **deste** país provou que a indexação salarial não protege o poder de compra.* [Ex.R]

(2.2) *Apesar de não haver **dados** oficiais até o início da noite, [...]* [Ex.R]

(2.3) *Nada havia que descortinasse a vida **desse** grupo social.* [Ex.R]

(2.4) *Três garotas, que irão em breve ao ar, batizadas de **Dá**, **Dá** e **Dá**.* [Ex.R]

(2.5) *Carlos Alberto Spinelli, morador da Barra e dono da **Der**Tee Laden.* [Ex.R]

Em (2.1), a forma *deste*, que é uma contração da preposição *de* com o pronome demonstrativo *este*, confunde-se com a forma verbal *deste* na segunda pessoa do singular (*tu*) do pretérito perfeito do indicativo. O substantivo *dados*, em (2.2), tem a mesma forma do verbo *dar* no particípio plural. Já a forma *desse*, em (2.3), que é a junção da preposição *de* com o pronome demonstrativo *esse*, confunde-se com a forma verbal *desse*, na primeira e terceira pessoas do singular (*eu* e *ele*) no pretérito imperfeito do subjuntivo. A forma *Dá*, usada em (2.4) como nome próprio, confunde-se com a forma verbal *dá*, na terceira pessoa do singular do presente do indicativo. Por fim, a forma *Der*, que também é usada como nome próprio de empresa, em (2.5), é homônima da forma *der* da primeira e terceira pessoas do singular (*eu* e *ele*) no futuro do subjuntivo.

A análise para a seleção do que é verbo e descarte daquilo que não o é foi feita manualmente, conferindo caso a caso. Dentre as ocorrências que o Unitex retornou, encontram-se também formas verbais compostas – ou locuções verbais – tais como *vem dando*, *foram dados*, *tinha dado*, *se fossem dados, etc.*, o que garante a eficácia da ferramenta na tarefa de reconhecimento das formas flexionadas de *dar*, pois o Unitex é capaz de recuperar todas essas ocorrências.

O Unitex permite também a construção de grafos, que podem ser usados para fazer buscas por padrões sintáticos ou por combinações de padrões lexicais. Os grafos são um tipo de representação formal utilizado em abordagens estruturais para a descrição de línguas e é também um recurso disponível no Unitex. A seguir, apresenta-se um exemplo de grafo do Unitex:

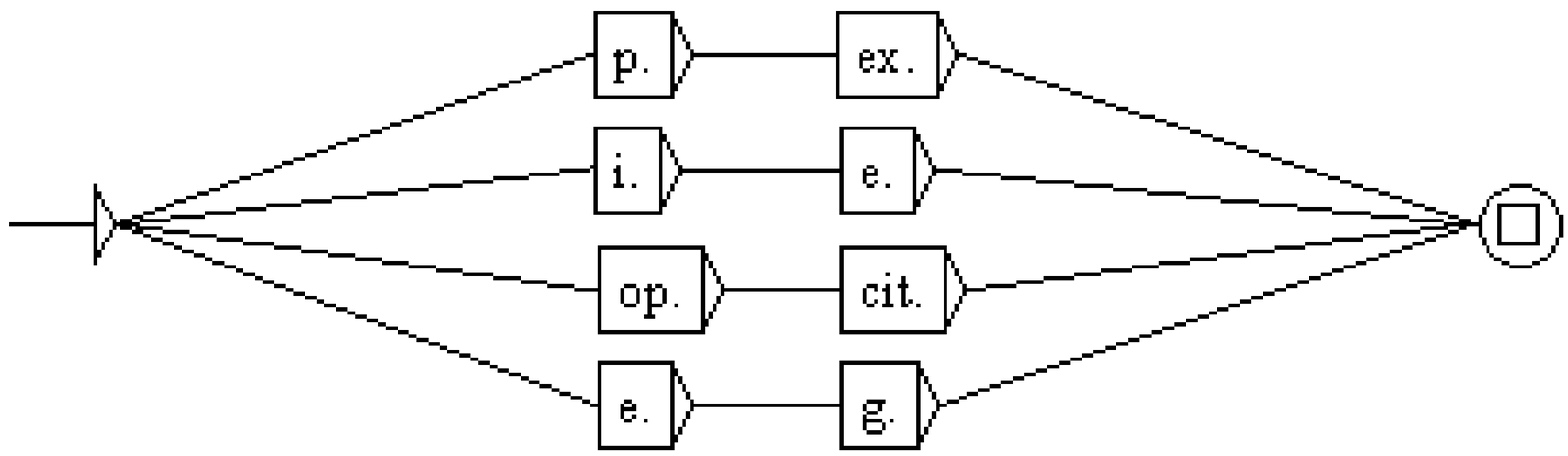

Figura 1 – Exemplo de grafo produzido no Unitex
Fonte: Elaborado pela autora

Os grafos são autômatos de estados finitos e são lidos da esquerda para a direita: a seta mais à esquerda indica o primeiro estado do grafo e o quadrado dentro de um círculo (mais à direita) indica o estado final. Entre os estados inicial e final, existem vários estados intermediários, que são representados pelas caixas (retângulos horizontais). Os caminhos entre um estado e outro são indicados por meio de linhas e setas.

Esse grafo em específico descreve as regras de abreviação de marcadores discursivos, tais como, *ou seja*, *isto é* e *por exemplo*. O grafo possui quatro caminhos: o primeiro deles busca pela letra "*p*", seguida de ponto final (*p.*), que deve ser imediatamente seguido da sequência "*ex.*". Depois o grafo segue para o estado final, indicado pelo quadrado dentro do círculo. O segundo caminho também possui dois estados intermediários, a fim de cobrir expressões como "*i.e.*"; e assim ocorre com os outros caminhos do grafo.

Pelo Unitex, também se pode intersectar os grafos de referência com tabelas de dados, por exemplo, as tabelas do L-G, e gerar automaticamente grafos de resultados, que instanciam os dados das tabelas nos grafos. A forma como esse recurso foi utilizado será explicada, pormenorizadamente, no Capítulo 12.

2.2.2 O concordanciador WebCorp Live

Após a seleção de todo o material linguístico a ser analisado, foi necessário consultar as propriedades formais e distribucionais das construções, bem como verificar as transformações que as frases admitem. Como algumas dessas propriedades ou transformações não se verificavam no *corpus* PLN.Br Full, recorremos também à *web*, por meio do concordanciador WebCorp Live[20], para atestar os usos em uma quantidade maior de dados.

20 Disponível livremente em: www.webcorp.org.uk.

O WebCorp (*The Web as a corpus*) é um conjunto de ferramentas que permite o acesso à *World Wide Web* como um *corpus* voltado à busca de informação linguística. O WebCorp oferece basicamente dois tipos de busca: simples e avançada. A busca simples inclui os campos: *keyword* (palavra-chave ou palavra de busca), *seleção do buscador*, *sensitive case* (seleção de palavras iniciadas por maiúsculas ou minúsculas), *span* (extensão do contexto da palavra-chave, isto é, se queremos 4, 5, 6 ou mais palavras à esquerda e à direita da palavra de busca), e a língua sobre a qual queremos obter informações.

A busca avançada (*advanced options*), por sua vez, oferece outros filtros, além dos incluídos na busca simples: domínio na *web* (.br, .es; .uk, .it, *etc.*), *sites* populares e área ou tema (*word filter*) relacionado à busca. Por exemplo, para buscar informações sobre o uso do nome *cadeira* somente no Português Brasileiro, podemos inserir o domínio .br, excluindo, assim, a busca de textos escritos nos demais domínios de topo associados a países de Língua Portuguesa.

A proposta do WebCorp é basicamente a mesma dos outros motores de busca, tais como, o *Google* ou o *Yahoo search*: utilizar a *web* como *corpus*. Neste trabalho, demos preferência para o uso da ferramenta WebCorp por ser mais linguisticamente motivada do que outros motores de busca.

2.2.3 A ferramenta CorpusAnnotator

Foi utilizada a ferramenta CorpusAnnotator (SUISSAS, 2014) para auxiliar no processo de anotação do *corpus*. Como parte da avaliação da tarefa, procedemos à anotação de uma amostra do *corpus* PLN.Br Full, que servirá como *corpus* de referência para avaliar a performance do sistema STRING.

Essa ferramenta foi desenvolvida em Java e precisa de dois arquivos com extensão .txt para funcionar: (i) um arquivo com todas as frases a serem anotadas (uma frase por linha); e (ii) um arquivo de parametrização com todas as formas de singular e plural dos nomes predicativos, com o objetivo de assinalar em cada frase a palavra-alvo da anotação, neste caso, o *Npred*. A Figura 2 apresenta um exemplo de frase a ser anotada.

Figura 2 – Exemplo de frase a ser anotada no CorpusAnnotator
Fonte: Tela da ferramenta CorpusAnnotator (SUISSAS, 2014)

Para cada frase a ser anotada, há apenas um par candidato a *CVS*, o qual está indicado entre parênteses no início da frase, conforme se verifica pela Figura 2. No canto superior esquerdo, um contador indica o número da frase corrente. Na Figura 2, a frase a ser anotada é a 48, de um total de 2.646.

Nesta seção, explicamos o funcionamento da ferramenta de anotação de *corpus*, em linhas gerais. As questões específicas sobre a anotação em si, as etiquetas (SVCSTANDARD, SVC-CONVERSE, VOPC ou OTHER) a serem atribuídas e os resultados dessa anotação serão apresentados na seção 12.3, na Parte III.

2.2.4 A ferramenta ReCal 0.1 Alpha for3+ Coders

A ferramenta ReCal 0.1 Alpha for 3+ Coders foi utilizada para calcular a concordância entre os anotadores[21]. Essa versão da ferramenta difere da versão anterior por permitir o cálculo da concordância entre três ou mais avaliadores. A versão anterior da ferramenta ReCal permitia a comparação apenas entre 2 avaliadores.

Para proceder ao cálculo da concordância entre anotadores, é preciso que os dados da anotação sejam numéricos e tabulados em colunas: em cada coluna devem constar os dados de cada anotador. O arquivo a ser analisado deve ter extensão .csv. A Figura 3 mostra um exemplo da saída da ferramenta ReCal 0.1 Alpha for 3+ Coders.

21 A ferramenta está disponível *online*, através do *link* http://dfreelon.org/recal/recal3.php#result1.

ReCal 0.1 Alpha for 3+ Coders results for file "teste13.csv"

File size:	3905 bytes
N coders:	3
N cases:	640
N decisions:	1920

Average Pairwise Percent Agreement

Average pairwise percent agr.	Pairwise pct. agr. cols 1 & 3	Pairwise pct. agr. cols 1 & 2	Pairwise pct. agr. cols 2 & 3
81.667%	81.094%	84.531%	79.375%

Fleiss' Kappa

Fleiss' Kappa	Observed Agreement	Expected Agreement
0.614	0.817	0.525

Average Pairwise Cohen's Kappa

Average pairwise CK	Pairwise CK cols 1 & 3	Pairwise CK cols 2 & 3	Pairwise CK cols 2 & 3
0.616	0.611	0.667	0.57

Krippendorff's Alpha (nominal)

Krippendorff's Alpha	N Decisions	$\Sigma_c o_{cc}$***	$\Sigma_c n_c(n_c - 1)$***
0.614	1920	1568	1934486

***These figures are drawn from Krippendorff (2007, case C.)

Figura 3 – Exemplo da saída da ferramenta ReCal 0.1 Alpha for 3+ Coders
Fonte: *Output* da ferramenta ReCal 0.1 Alpha for 3+ Coders

Nesse exemplo, foram considerados três anotadores (*N coders 3*), 640 frases a serem anotadas (*N cases 640*) e 1.920 sequências de anotações (*N decisions 1920*), considerando-se os três anotadores.

A saída da ferramenta apresenta quatro medidas para a avaliação da concordância, as quais são indicadas nas quatro tabelas: (i) a primeira tabela (*Average Pairwise Percent Agreement*) corresponde à porcentagem de concordância entre cada par de anotadores, considerando-se os três anotadores; (ii) a segunda tabela (*Fleiss' Kappa*) indica o coeficiente Kappa de Fleiss (1971), que é uma medida estatística que compara o grau de concordância efetivamente observado entre um número fixo de avaliadores e a concordância que seria esperada se as anotações fossem aleatórias; (iii) a terceira tabela (*Average Pairwise Cohen's Kappa*) indica o coeficiente Kappa de Cohen (1960), que também é uma medida estatística que avalia a concordância entre pares de avaliadores, também chamada de *inter-annotator agreement* ou *inter-rater agreement*; e (iv) a última

tabela (*Krippendorff's Alpha (nominal)*) apresenta o coeficiente Alpha de Krippendorff (1970), que é uma medida estatística que calcula a concordância obtida quando se codifica um conjunto de unidades de análise em termos dos valores de uma variável.

Nesta seção, limitamo-nos a apresentar a ferramenta para o cálculo da concordância entre anotadores. Os resultados da concordância para a anotação feita neste trabalho serão apresentados na seção 12.4.

2.3 Princípio de classificação

2.3.1 Da classificação geral

Conforme indicado na seção anterior, a busca pelo verbo *dar* no Unitex retornou 66.798 ocorrências (*tokens*). Todas elas foram analisadas manualmente para que pudessem ser classificadas. Inicialmente, excluíram-se os casos em que as formas *desse(s)*, *deste(s)* e *dado(s)* pertenciam a outra categoria gramatical diferente de verbo, restando 36.686 *tokens* a serem analisados.

Dos 36.686 *tokens* do verbo *dar*, foram identificados 3.560 *types* de colocados desse verbo. Essa lista foi intersectada com a lista de colocados do AC/DC, que contava com 2.452 lemas de substantivos associados ao lema do verbo *dar*. Da intersecção, excluíram-se os *types* duplicados e as diferentes formas do mesmo lema, restando um total de 3.602 construções a serem analisadas. Essas construções foram, posteriormente, agrupadas nas seguintes classes: verbo pleno, verbo-operador causativo, expressão cristalizada, provérbio e candidatos a verbo-suporte. As quatro primeiras classes admitiam todos os casos prototípicos de cada uma delas, ou seja, as construções que possuem as propriedades específicas de cada categoria[22] foram agrupadas nas suas respectivas classes.

Em seguida, os candidatos a construções com verbo-suporte foram testados, um a um, com base na aceitabilidade das propriedades definitórias das *CVS* (ver Capítulo 8). Os candidatos que admitiam, pelo menos, a propriedade da relação particular entre o *Npred* e um dos argumentos da construção foram inseridas na classe das *CVS*. As que não admitiam nenhuma das propriedades definitórias foram separadas para serem analisadas em uma segunda etapa.

22 As propriedades específicas de cada categoria verbal estão descritas nos respectivos capítulos de verbo pleno (Cap. 4), verbo-operador causativo (Cap. 5) *etc.*

2.3.2 Das construções com verbo-suporte (*CVS*)

Na classe de "candidatos a verbo-suporte" entraram todos os casos típicos de construções com verbo-suporte e também os casos duvidosos quanto à classificação. Dessa forma, prevê-se que todos os casos fronteiriços entre duas ou mais categorias constem no rol de "candidatos a verbo-suporte". Nessa classe, havia inicialmente 2.625 entradas (*types*) potenciais formadas pelo verbo *dar* e um nome, incluindo os resultados da busca no *corpus* PLN.Br Full e no AC/CD.

A lista de 2.625 colocados foi intersectada com a lista dos *Npred* para o Português Europeu (VAZA, 1988); foram excluídos os *types* duplicados e foram acrescentados apenas os nomes do PE que também ocorrem em PB e que, de fato, consistem em nomes predicativos. Após essa etapa, restaram cerca de 3.100 entradas lexicais (*types*), cada uma em uma linha da matriz.

A partir da aplicação de testes transformacionais, que serão explicados no Capítulo 8, às potenciais construções com verbo-suporte, pôde-se comprovar sistematicamente os nomes predicativos que, de fato, constituem construções com verbo-suporte (*CVS*), e foram excluídos os casos que não correspondiam a *CVS*. Após essa etapa, restaram 908 entradas lexicais confirmadas como *CVS*.

Esse número aumentou para 1.489 entradas lexicais depois de acrescentar cerca de 500 *Npred* com sufixo *-da* que se formam a partir dos lemas dos verbos do ViPER, descritos na subseção 2.1.3.

2.4 Método de descrição do Léxico-Gramática

A Linguística Estrutural, conforme aponta Borba (1979), prevê diferentes maneiras de formalização dos dados linguísticos. Alguns dos tipos mais comuns de representação gráfica são os diagramas, os grafos, os parênteses ou colchetes, as caixas, os grafos simplificados ou alterados, as árvores, dentre outras (BORBA, 1979). O modelo do Léxico-Gramática, para descrever formalmente a língua, utiliza a representação em tábuas, ou matrizes binárias de dupla análise[23]: (i) as linhas, que representam os verbos ou quaisquer outras entradas lexicais - no nosso caso, representam os nomes predicativos; e (ii) as colunas, que representam as propriedades sintáticas, semânticas, estruturais, distribucionais e transformacionais das construções.

23 O desenvolvimento dos estudos descritivos em tábuas iniciou-se por volta dos anos 1970, por Gross (1975), no *Laboratoire Automatique Documentaire et Linguistique* (LADL) da Université Paris-Est

As tábuas são construídas no formato de matrizes binárias, ou seja, tabelas retangulares preenchidas por sinais de "+" e "–", de acordo com a aceitabilidade ou inaceitabilidade de tal propriedade sintático-semântica. A Tabela 1 demonstra um exemplo de representação em matriz binária.

Npred	N0=:Hum	N0=:N-Hum	DET=:E	DET=:Def	DET=:Indef	PrepPB	Npred=V-n	N1=Hum	N1=:N-Hum	N1=:Npc	Verboplenocorrespondente	Exemplo
açoite	+	–	+	–	+	em	+	+	–	–	açoitar	O Zé deu um açoite na Ana.
agulhada	+	–	–	–	+	em	+	+	+	+	agulhar	O Zé deu uma agulhada no dedo.
alfinetada1	+	–	–	–	+	em	+	+	+	+	alfinetar	O Zé deu uma alfinetada no dedo.
alfinetada2	+	–	–	–	+	em	+	+	–	–	alfinetar	O Zé deu uma alfinetada na Ana.
almofadada	+	–	–	–	+	em	–	+	+	+	–	O Zé deu uma almofadada na Ana.
aperto	+	–	–	–	+	em	+	+	+	+	apertar	O Zé deu um aperto no parafuso.
balão	+	–	+	+	+	em	–	+	–	–	–	O Zé deu um balão no Neymar.
banho	+	+	–	–	+	em	–	+	+	+	–	O Zé deu um banho na Ana.
bastonada	+	–	–	–	+	em	–	+	+	+	–	O Zé deu uma bastonada na cabeça da Ana.
beijo	+	–	+	+	+	em	+	+	–	–	beijar	O Zé deu beijo na Ana.
bengalada	+	–	–	–	+	em	–	+	+	+	–	O Zé deu uma bengalada na cabeça da Ana.
bicada	+	+	–	–	+	em	+	+	+	+	bicar	O pássaro deu uma bicada na perna da Ana.
bicuda	+	–	+	+	+	em	–	+	+	+	–	O Zé deu uma bicuda na bola.

bofetada	+	–	–	–	+	em	+	+	+	+	bofetear	O Zé deu uma bofetada na cara da Ana.
bordoada	+	–	+	+	+	em	+	+	+	+	bordoar	O Zé deu uma bordoada no filho.
bote	+	–	+	+	+	em	–	+	–	–	–	A cobra deu um bote em Rui.
botinada1	+	–	–	–	+	em	–	+	+	+	–	O Zé deu uma botinada na perna da Ana.
botinada2	+	–	+	–	+	em	–	+	–	–	–	O Zé deu botinada no Neymar.
breque	+	–	+	–	+	em	+	–	+	–	brecar	O Zé deu um breque no carro.
cabeçada	+	+	+	+	+	em	–	+	+	+	–	O Zé deu uma cabeçada no nariz da Ana.
cacetada	+	+	+	+	+	em	+	+	+	+	cacetear	O Zé deu uma cacetada na cabeça da Ana.

Tabela 1 – Exemplo de matriz do Léxico-Gramática
Fonte: Fragmento de tábua da classe DL (RASSI *et al.*, 2015)

Na coluna mais à esquerda encontram-se as entradas lexicais, que, neste caso, são os nomes predicativos que selecionam o verbo-suporte *dar* e fazem a conversão com *levar*, mas poderiam ser verbos, adjetivos, expressões cristalizadas ou outras. As duas colunas subsequentes dizem respeito à natureza do sujeito (se humano ou não-humano), seguidas das especificidades dos determinantes (sem determinante ou artigo definido ou artigo indefinido), seguido ainda do tipo de preposição que introduz o complemento. Na oitava coluna (*Npred=V-n*), descreve-se a propriedade da nominalização, ou seja, se o nome predicativo é ou não uma nominalização (*e.g. beijo* é uma nominalização de *beijar*, mas *bote* não é nominalização de **botar* ou **botear*). As colunas subsequentes especificam a natureza do complemento *N*1 (se nome humano, não-humano ou parte-do-corpo). A penúltima coluna traz a informação do verbo pleno correspondente, no caso de nominalizações, e a última coluna mostra um exemplo fabricado de cada construção (BOONS; GUILLET; LECLÈRE, 1976; GUILLET; LECLÈRE, 1992).

Caso a construção admita a propriedade sintática que está representada na décima coluna, por exemplo, marca-se o sinal "+" na intersecção entre dada linha e dada coluna. Se não admite, marca-se "–". Além dos sinais de "+" e "–" para indicar aceitabilidade ou não da propriedade, também é possível preencher células com valores lexicais, por exemplo, no caso das preposições e dos verbos plenos correspondentes.

Esse é o formato geral das descrições do Léxico-Gramática, porém o tipo de propriedade a ser analisada nas colunas varia conforme o fenômeno estudado. Cada tábua corresponde a uma classe que reagrupa os elementos de uma categoria lexical ou gramatical de acordo com suas propriedades sintático-semânticas, chamadas de propriedades definitórias. Conforme atestam Tolone, Sagot e Clergerie (2012), as tábuas do L-G constituem hoje um dos principais recursos de informação léxico-sintática para o Francês.

Além disso, ressalta-se também o número significativo de entradas lexicais nas matrizes. O Léxico-Gramática do Francês, por exemplo, possui cerca de 31.000 entradas verbais, além das entradas de nomes e adjetivos predicativos. Em relação ao Léxico-Gramática do Português Europeu, há descrições já formalizadas para cerca de 2.000 adjetivos (CASTELEIRO, 1981), cerca de 4.000 nomes predicativos (RANCHHOD, 1990; BAPTISTA, 2005b), e cerca de 3.500 expressões fixas (BAPTISTA; CORREIA; FERNANDES, 2004). Cita-se também o ViPEr, que é o L-G das construções verbais do PE, que está em fase de desenvolvimento, mas já conta com cerca de 6.300 entradas verbais.

Também vêm sendo desenvolvidos Léxico-Gramáticas para outras línguas, como Italiano, Espanhol, Grego moderno, Coreano, dentre outras. As representações das matrizes contam com um número significativo de entradas, na casa dos milhares, no mínimo, para cada idioma. Eles são o resultado da aplicação de um conjunto de princípios linguísticos comuns:

- os significados dos predicados têm que ser rigorosamente determinados e distinguidos, quando necessário; e eles são assim determinados pelas restrições de seleção dos seus argumentos. Nesse sentido, os tipos de argumentos também devem ser cuidadosamente definidos (por exemplo, nomes do tipo humano, não-humano, parte-do-corpo, locativo etc.);
- a forma das frases tem que ser restrita a um conjunto característico de argumentos, ou seja, as frases-exemplo são constituídas apenas com argumentos essenciais (por exemplo, sujeito e complementos), excluindo-se os não essenciais (por exemplo, expressões adverbiais e circunstanciais);
- um conjunto de transformações tem de ser determinado experimentalmente e testado sistematicamente para uma grande quantidade de dados. Essas transformações devem ser operativas e reprodutíveis para outros dados.

Esses princípios gerais foram adotados para as línguas já descritas nos moldes do L-G e a experiência, ao longo dos anos, tem mostrado que várias descrições chegaram

a resultados muito semelhantes: (i) não existem dois itens lexicais que contenham exatamente as mesmas propriedades formais e distribucionais; (ii) o número de predicados semânticos é semelhante em todas as línguas; (iii) há elementos não-verbais, tais como adjetivos e substantivos, que também podem predicar, assim como também pode haver verbos que não predicam, tais como, os *Vsup*; (iv) as classes morfossintáticas e distribucionais mantêm proporções quantitativas semelhantes em línguas aparentadas, tais como, o Espanhol, o Português e o Francês (isso pode ser dito para os verbos psicológicos e para várias classes de construções fixas, por exemplo); (v) o Léxico-Gramática pode ser um recurso léxico-sintático bastante útil para aplicações computacionais.

2.5 Integração dos dados em *parser*

O conjunto de dados referentes a *CVS* com o *Vsup dar* foram integrados à gramática de uma analisador sintático automático (*paser*) chamado XIP (MOKHTAR; CHANOD; ROUX, 2002), de forma a possibilitar a correta análise sintática das *CVS*. Em trabalho anterior (RASSI *et al*., 2014), delineamos as estratégias para implementação dos dados com os *Vsup dar* e *ter* na cadeia de processamento do Português STRING (MAMEDE *et al*., 2012).

Em um segundo momento, procedemos à anotação de uma amostra do *corpus* PLN. Br Full, para a construção de um *subcorpus* de referência para o tratamento de *CVS*. Esse *subcorpus* contém 2.646 frases anotadas e revisadas por cinco anotadores humanos, considerando-se 45 variantes dos *Vsup dar*, *ter* e *fazer* e cerca de 3.200 nomes predicativos diferentes.

Em um terceiro momento, integramos os dados referentes ao *Vsup dar* no *parser* XIP, que é utilizado na STRING e avaliamos a performance do sistema com base na comparação com o *corpus* de referência. A integração dos dados no *parser* foi feita com base na criação de programas que: (i) validam as informações constantes na matriz do L-G; e (ii) geram automaticamente regras para a extração de dependências. Todos esses procedimentos serão explicados detalhadamente na Parte III.

Neste capítulo, apresentamos a metodologia utilizada ao longo trabalho: para a obtenção dos dados, para a busca e processamento dos dados recolhidos, para a classificação das construções, para a descrição das propriedades linguísticas das *CVS* e para o processamento automático dessas *CVS*. As construções com o verbo *dar*, em Português Brasileiro, foram recenseadas em *corpora* e em listas de construções previamente elaboradas por outros autores.

Foram utilizadas diferentes ferramentas computacionais: (i) o *software* Unitex foi usado para recensear as construções no *corpus* PLN.Br Full, utilizando-se gramáticas locais e grafos; (ii) o concordanciador WebCorp Live foi usado para atestar a aceitabilidade de certas construções do Português, utilizando a *web* como *corpus*; (iii) o CorpusAnnotator foi utilizado como uma ferramenta de auxílio para a anotação das *CVS* do *corpus* de referência; e (iv) o Recall 3+Coders foi usado para calcular a concordância entre os anotadores no processo de anotação do *corpus* de referência.

Este capítulo apresentou os critérios para a classificação das construções, numa proposta taxonômica. Apresentou também a metodologia de descrição do Léxico-Gramática, que é feita em matrizes binárias em que se assinalam as propriedades formais, distribucionais e transformacionais das construções com verbo-suporte. Por fim, ainda neste capítulo, foi introduzida, resumidamente, a proposta metodológica de integração das *CVS* em um analisador sintático automático. A descrição completa dessa metodologia, porém, será apresentada na Parte III.

No próximo capítulo, indicaremos os principais trabalhos relacionados à classificação de verbos e à descrição de *CVS*, além de apresentar o estado da arte das pesquisas em Linguística e em PLN que fazem o processamento sintático automático dessas construções.

3. Estado da arte

Há, na literatura, diversos trabalhos descritivos sobre as construções com o verbo *dar* ou trabalhos que propõem a classificação geral dos verbos, seja adotando critérios sintáticos seja baseando-se em critérios semânticos. Destacam-se também os trabalhos que tratam especificamente das construções nominais com verbo-suporte, tanto em Português quanto em outras línguas. Por fim, a literatura oferece também abordagens voltadas para o processamento computacional dessas construções nominais. Os trabalhos de maior relevância para o desenvolvimento do estado da arte serão referidos e discutidos a seguir.

Este capítulo está dividido em três seções, que correspondem às três partes deste *e-book*: a primeira seção apresenta o estado da arte da classificação geral das construções verbais em Português, indicando e discutindo as referências que tenham tratado os fenômenos que envolvem o verbo *dar*; a segunda seção apresenta os trabalhos de descrição linguística das construções nominais com verbo-suporte e nome predicativo, comparando o estado da arte das descrições para o Português com o de outras línguas; a terceira seção apresenta o estado da arte do processamento automático de construções nominais com verbo-suporte e nome predicativo.

3.1 Classificação geral das construções verbais

Quando recorremos aos dicionários para tentar caracterizar os fenômenos que envolvem o verbo *dar*, encontramos listados alguns desses fenômenos (mas não todos), pouca (ou nenhuma) explicação sintática sobre eles e, muitas vezes, uma classificação falha ou inconsistente.

O *Dicionário Gramatical de Verbos* (BORBA, 1991), por exemplo, considerado um dicionário de referência para o Português, apresenta algumas inconsistências. Uma delas está na análise da frase *aquilo me dava felicidade* como exemplo de construção com "verbo de ação-processo com sujeito agente e dois complementos (um expresso por nome abstrato, e outro, destinatário)" (BORBA, 1991, p.364). Na classificação que propomos, a construção *aquilo me dava felicidade* deve ser tratada como uma causativa, cuja posição sujeito é ocupada por um argumento que exprime uma causa, e não por um sujeito exclusivamente agente, assim como também não consideramos o complemento *me* como um destinatário, e sim como o experienciador do estado psicológico denotado pelo nome predicativo *felicidade*.

O dicionário (BORBA, 1991) também apresenta uma seção, na entrada do verbo *dar*, para os verbalizadores[24], que são, grosso modo, o que chamamos *construções com verbo-suporte*. Apesar de haver uma seção específica para os verbalizadores, os autores citam *Nunca lhe dera um beijo* como exemplo de construção com "verbo de ação-processo com sujeito agente e dois complementos (um expresso por nome não-animado, e outro da forma *em* + nome animado)" (BORBA, 1991, p.364). Na abordagem aqui proposta, a frase *Nunca lhe dera um beijo* é tipicamente uma construção com verbo-suporte e deveria, portanto, ter sido inserida na classe a que eles chamam *verbalizadores*.

Quando o dicionário indica exemplos de modalizador factivo, cita a frase "*Tudo coincidiu com umas frases de Leopoldo que* ***davam a entender*** *sua vontade de empregar-se*" (p.366) e depois cita novamente a construção *dar a entender* em "*Alaíde* ***dava a entender*** *ao companheiro que fora ali se ocultar*" (*ibidem*) como uma expressão fixa. Esse é um caso típico de inconsistência terminológica, o que ocorre com frequência em alguns trabalhos lexicográficos, sobretudo nos que são feitos em equipe.

Além dessas e várias outras inconsistências de classificação, há, muitas vezes, uma superespecificação de algumas categorias, por exemplo, a construção "com sujeito paciente expresso por nome indicativo de relógio e com complemento expresso por nome indicativo de hora, significa *soar, bater*: *O relógio dava duas horas da tarde*" (p.365) ou então a construção "com sujeito inativo expresso por nome de bicho da lista do jogo de bicho, significa *ser sorteado (no jogo)*: *Hoje deu cobra*" (p.366).

No *Dicionário Sintáctico de Verbos Portugueses*, Busse (1994) elenca sete entradas sintaticamente diferentes para o verbo *dar*, sendo cada entrada ainda subdividida em outras subclasses, dependendo do tipo de nome que preenche as diferentes posições argumentais (*N*), ou o tipo de preposição (*prep*), etc. Como a classificação é estritamente sintática, o autor agrupa, numa mesma entrada, sentenças como: (i) *Ela deu um livro ao Paulo*, (ii) *Deu-lhe uma bofetada*, (iii) *Aquele médico deu ao doente a possibilidade de ouvir novamente*, (iv) *dar pérolas aos porcos*, (v) *dar parabéns* e (vi) *dar-se conta de N*. Ao nosso ver, todas essas construções possuem estatutos diferentes e devem ser classificadas sistematicamente com base em propriedades sintático-semânticas específicas, como será justificado no decorrer deste *e-book*.

Apesar de algumas inconsistências, que são próprias de trabalhos lexicográficos exaustivos, reconhece-se o mérito desse tipo de trabalho para a descrição linguística,

24 A definição de verbalizadores se encontra no glossário do dicionário (BORBA, 1991, p.XXI): "VERBALIZADOR OU VERBO SUPORTE - verbo que, numa construção complexa, é mero suporte de categorias e indica que o verdadeiro verbo (= núcleo do predicado) está no radical de seu complemento. Ex.: *ter* medo=*temer*; *causar* dano=*danificar*; *abrir* falência=*falir*; *ganhar* distância=*distanciar-se*".

dado que eles abarcam um vasto número de construções. Borba (1991) apresenta cerca de 50 sentidos para a entrada do verbo pleno *dar*; Busse (1994) dá sete entradas diferentes para o verbo *dar*, sendo que, na primeira entrada, constam 38 sentidos da construção com *dar* como verbo pleno. As classificações e tipologias apresentadas por esses dicionários serão levadas em consideração nesta pesquisa, principalmente em relação às construções com verbo pleno.

Outros trabalhos de análise linguística também tentaram traçar uma taxonomia para o verbo *dar*, como é o caso de Coelho, Silva e Sousa (2013), que fazem uma análise do processo de gramaticalização do verbo *dar*. Na análise proposta por Coelho, Silva e Sousa (2013), os autores identificam o significado do verbo pleno *dar* como *presentear* (*mas você* ***dá o presente*** *também, não dá?*). Em seguida, os autores constatam "uma outra categoria semelhante ao verbo pleno, mas que se distancia um pouco por possuir outra extensão de sentido. Nesse caso, o verbo *dar* apresenta o valor de *fornecer*" (COELHO; SILVA; SOUSA, 2013, p.3211), como no exemplo: *uma pessoa* ***dava informação*** *pra agente* [*sic.*].

Os autores ainda recorrem à classificação de Esteves (2008), que nomeia essa categoria como *verbo predicador não pleno*. Os conceitos que esses autores usam de *verbo pleno* ou de *verbo predicador* ou ainda de *verbo predicador não pleno* não são mencionados. Em seguida, Coelho, Silva e Sousa (2013, p.3212) definem o conceito de verbo-suporte: "Outra categoria analisado [*sic.*] no *corpus* se distingue um pouco da anterior, o verbo suporte [...] podemos encontrar na língua um verbo com o mesmo valor, como em *dar amor, amar*; *dar um abraço, abraçar*".

Pela definição adotada pelos autores, as duas construções anteriores (*dar presente* e *dar informação*) deveriam ser classificadas como construções com verbo-suporte, já que existem, na língua, outros verbos "com o mesmo valor", tais como *presentear* e *informar*, respectivamente.

Mais adiante, Coelho, Silva e Sousa (2013, p.3212) identificam "a existência do verbo *dar* como causa, significando condição física ou psicológica" e citam como exemplos: *dar febre*, *dar nervoso* e *dar agonia*. Na mesma página, em parágrafo anterior, os mesmos autores classificaram a construção *dar conforto* (em *graças a Deus, ele sempre me* ***dá conforto***) como uma construção com verbo-suporte, já que admite um verbo pleno equivalente (*confortar*). Por um lado, o verbo *dar*, em *dar conforto*, expressa uma causa, significando condição física ou psicológica, então deveria ser classificado como verbo causativo. Por outro lado, utilizando-se os mesmos critérios que os autores propõem, a construção *dar agonia* tem um verbo equivalente (*agoniar*), então deveria ser classificada como construção com verbo-suporte, assim como *dar presente* e *dar informação*.

Essa breve explicação dá uma ideia da grande complexidade que envolve as construções com o verbo *dar* e também da falta de sistematicidade nas descrições e análises encontradas na literatura.

Além dos trabalhos lexicográficos e de análise descritiva, há ainda diferentes trabalhos linguísticos de descrição das construções verbais, que podem ser adaptados à análise das construções do verbo *dar*. Cançado, Godoy e Amaral (2012), por exemplo, elaboraram uma proposta geral de classificação dos verbos do Português, tendo como base a classificação de Levin (1993) para as construções verbais do Inglês.

Levin (1993) propôs uma classificação dos verbos do Inglês em 57 grandes classes (com diversas subclasses), de acordo com as características sintáticas e semânticas que esses verbos compartilham. Naquele trabalho, Levin (1993) descreveu as classes verbais como um conjunto de verbos que compartilham um ou mais componentes de significado e apresentam comportamento sintático similar. Para o Português, Cançado, Godoy e Amaral (2012) estão produzindo um catálogo de verbos do Português Brasileiro, seguindo a proposta de classificação verbal de Levin. Algumas daquelas classes, tais como a dos verbos psicológicos, dos verbos causativos e a dos verbos de transferência, serão consideradas nesta pesquisa, haja vista que o verbo *dar* pode constituir predicados desses três tipos.

Para o Português Europeu, há também um recurso léxico-computacional, chamado ViPEr (BAPTISTA, 2012), que é o Léxico-Gramática das construções verbais do PE. Muitas das construções com o verbo pleno *dar* que estão descritas no ViPEr também ocorrem em Português Brasileiro e, portanto, serão analisadas da mesma forma, com base nas suas propriedades estruturais, distribucionais e transformacionais.

Além desses trabalhos, que propõem classificações gerais para as construções verbais de uma língua, recorremos também a análises particulares de um verbo ou uma construção verbal específica, a fim de caracterizar globalmente todas as construções em que o verbo *dar* ocorre.

As construções causativas com o verbo *dar* têm como base a noção de verbo-operador, que foi proposta inicialmente por Gross (1981, p.23-39), distinguindo entre duas situações: (i) o verbo-operador causativo, que introduz um argumento à frase de base, estabelecendo entre este argumento e a frase simples uma relação semântica de causa, podendo, frequentemente, reestruturá-la, modificando as relações sintáticas entre os constituintes (*e.g. Ana tem medo de escuro* [*Causativo*] = *O escuro dá medo em Ana*); e (ii) o verbo-operador de ligação, que também acrescenta um argumento à frase

de base, o qual já está presente na frase elementar, na posição de complemento, mas não modifica o significado da frase de base (*e.g. Ana está sob o controle do Rui* [*VopL*] = *O Rui tem a Ana sob controle*).

Cançado (2010) também faz análise semântica das construções causativas, porém leva em consideração as alternâncias sintáticas que um predicado semântico compartilha com outros da mesma classe. Por adotarmos a proposta teórico-metodológica do Léxico-Gramática, consideramos apenas uma forma sintática de base, que corresponde a frases do *kernel* da língua (HARRIS, 1951); assim, outras estruturas, por exemplo, a construção causativa, são transformações de uma frase de base, mas não são, em si mesmas, frases de base, já que podem ser decompostas em dois ou mais predicados semânticos.

Sobre as expressões fixas, recorremos ao trabalho de Vale (2001), que construiu um dicionário eletrônico de expressões cristalizadas em Português Brasileiro. Naquele trabalho, porém, o autor não levou em consideração as expressões fixas formadas pelos verbos *dar*, *fazer*, *ter* e *estar*. Devido à grande produtividade das expressões fixas com esses verbos, o autor considerou que deveriam ser objeto de estudo em outro trabalho. Neste *e-book*, portanto, pretendemos recensear e classificar as expressões cristalizadas com o verbo *dar*, utilizando a mesma classificação sintática proposta por Vale (2001).

As expressões cristalizadas do Português Europeu já foram descritas por Baptista, Correia e Fernandes (2004), incluindo as que se constroem com o verbo *dar*. As que são encontradas também em Português Brasileiro, e que tenham sido atestadas em *corpora*, serão analisadas e classificadas no Capítulo 6 deste *e-book*.

Citam-se também duas teses de referência para a descrição das construções com o verbo *dar*: Salomão (1990) analisa cerca de 25 construções com *dar*. Contudo, o foco daquele trabalho não é descrever as propriedades formais dessas construções, mas sim explicar fenômenos linguísticos, como a polissemia, a aspectualidade e a modalidade. Essas construções com o verbo *dar*, naquele trabalho, são meramente exemplificativas, portanto, somente são úteis no âmbito desta pesquisa no sentido de fornecer exemplos de construções. A outra tese é de Scher (2004), sobre as construções nominais formadas pelo verbo *dar* e nominalizações terminadas em *-ada*, com uma abordagem gerativa baseada na Morfologia Distribuída. Na tese, a autora faz uma análise minuciosa das construções, mas não visa à exaustividade nem a uma fomalização dos dados voltada para seu processamento automático.

3.2 Identificação e descrição das *CVS*

Identificar construções com verbo-suporte não é uma tarefa fácil, nem manualmente por humanos, nem automaticamente por máquinas. Citemos aqui pelo menos quatro trabalhos que propõem diferentes testes formais para identificar manualmente *CVS*.

Langer (2004) propõe uma bateria de testes para distinguir, por um lado, construções semi-composicionais (que se assemelham ao que denominamos *CVS*) e combinações de verbo-nome que são semanticamente composicionais (construção com verbo pleno, por exemplo) e, por outro lado, *CVS* e expressões idiomáticas. O autor apresenta 15 testes que investigam: (i) a referencialidade do sintagma nominal predicativo; (ii) o esvaziamento semântico do verbo, por meio de testes de substituição do *Vsup* ou da impossibilidade de nominalizar *Vsup*; e (iii) o estatuto dos complementos, já que os complementos de nomes predicativos podem ser extraídos para fora do sintagma nominal e alguns argumentos admitem dupla realização.

Para tanto, Langer (2004) compara construções em Inglês, Francês e Alemão. Apesar de quase todos os testes serem válidos para identificar *CVS* em Português, nenhum deles é considerado necessário e suficiente para definir uma *CVS*. Como o próprio autor afirma nas conclusões, esses testes representam uma aproximação das propriedades semânticas subjacentes às *CVS*, mas todos eles podem admitir exceções. Algumas dessas exceções são apresentadas pelo próprio autor, ao longo do texto, pois alguns testes valem para uma língua e não para outra(s).

O segundo trabalho é o de Mel'čuk (2004), que explica, de forma precisa e rigorosa, o quadro formal de caracterização das construções com verbo-suporte, que havia sido proposto pelo mesmo autor na década de 1960. O autor estabelece cinco regras básicas que podem ser aplicadas a todas as construções com verbo-suporte: (i) o verbo-suporte é semanticamente vazio; (ii) existem apenas três tipos possíveis de verbo-suporte (o mais comum é aquele em que o nome predicativo está na posição de objeto direto, um outro em que o nome predicativo assume a posição de sujeito, e um terceiro tipo bem mais raro em que o nome predicativo ocorre na posição de objeto indireto); (iii) há dois sentidos mais gerais que um verbo-suporte pode assumir (o sentido frásico e o sentido causativo); (iv) há vários sentidos que o nome predicativo pode assumir, mas este só pode ser inferido na combinação com o *Vsup*; (v) existe uma família de verbos semanticamente plenos que possuem o mesmo comportamento sintático que os verbos-suporte.

Além das contribuições teóricas propostas por Mel'čuk (2004), o autor também propõe uma representação bastante sistemática do ponto de vista formal para essas construções. Apesar de concordarmos com grande parte das afirmações, não assumiremos a proposta de Mel'čuk (2004) por três motivos: (i) não consideramos os verbos causativos como um tipo de verbo-suporte; (ii) o autor faz uma excelente análise sintático-semântica, mas não se preocupa com a exaustividade das construções, o que é um de nossos objetivos; e (iii) a representação formal e o nível de análise são extremamente rebuscadas para uma aplicação computacional.

O terceiro trabalho investigado é o de Mejri (2008), que faz uma análise de construções semi-composicionais, distinguindo três tipos de construções que são sintática e semanticamente diferentes: as construções com verbo-suporte, as colocações e as locuções verbais. O autor também reconhece a gradação entre essas construções em termos de fixidez dos elementos, partindo das frases cuja sintaxe é livre, passando pelas construções com verbo-suporte, seguidas das colocações até chegar nas frases fixas.

O autor utiliza os princípios do Léxico-Gramática, tais como, identificar a frase simples como unidade mínima de análise e correlacionar os vários níveis linguísticos, e descreve as propriedades formais dessas construções, a fim de distingui-las. Para identificar as construções com verbo-suporte, Mejri (2008) propõe substituições paradigmáticas nas posições dos argumentos. Isso significa descrever extensionalmente todos os nomes que podem ser selecionados para a posição de sujeito e de complementos em cada construção com verbo-suporte. Essa abordagem, porém, não é nova na descrição do léxico. Gross (1994) já havia tratado das classes de objetos, como grupos de nomes semanticamente semelhantes, que podem comutar paradigmaticamente na frase, tanto na posição de sujeito quanto na posição de complemento.

Por fim, recorremos a uma quarta abordagem, que é atribuída a vários autores (GROSS, 1981; RANCHHOD, 1990; BAPTISTA, 2005b) que, ao longo do tempo, foram acumulando evidências para considerar seis testes sintáticos, que se referem a propriedades definitórias das *CVS*. Dos seis testes, apenas um deve ser considerado necessário e suficiente para a identificação das *CVS*: a estreita relação entre o *Npred* e um de seus argumentos, que não permite que o *Npred* se situe fora da esfera de referência, por exemplo, de seu sujeito. Nesse sentido, o sujeito sintático do verbo-suporte (com o qual o *Vsup* concorda) tem que ser o mesmo agente do predicado expresso pelo *Npred*. Os seis testes serão apresentados e discutidos em pormenores no Capítulo 8, na Parte II.

Após a identificação das *CVS*, preocupamo-nos também com o estado da arte da descrição dessas construções. Há diversos trabalhos descritivos do Português Brasileiro

sobre as construções com verbo-suporte e, mais especificamente, as *CVS* com *dar*. Neves (2000), na *Gramática de Usos do Português*, explica as construções com o verbo-suporte *dar*, buscando a motivação que leva os falantes à escolha por uma construção com verbo pleno ou uma com verbo-suporte; Basilio (2007) descreveu alguns poucos exemplos de construções com os verbos *dar* e *fazer* seguidos de sintagma nominal, identificando os casos particulares das ocorrências com verbo-suporte; Davel (2009) caracterizou as propriedades das construções com a estrutura [*dar SN*], com base na Gramática Funcional; Abreu (2011) também investigou a semântica das construções com verbos-suporte na FrameNet, tratando mais especificamente das construções com os verbos *dar* e *fazer*; Smarsaro e Rocha (2011) discutiram as propriedades de construções impessoais com *dar*, como em *Deu mofo no guarda-roupas* e *Deu piolho na cabeça do Rui*.

Ressalte-se que todos esses trabalhos trazem contribuições teóricas ao estado da arte, mas nenhum deles visa à exaustividade do fenômeno ou a uma formalização voltada ao PLN. Seguindo a proposta do Léxico-Gramática, outros autores buscaram descrever em extensão o maior número possível de predicados nominais, em diferentes línguas.

Para o Francês, por exemplo, Giry-Schneider (1978, 1987) descreveu os predicados nominais com o *Vsup faire* (*fazer*), Labelle (1984), os nomes predicativos que selecionam o verbo-suporte *avoir* (*ter*); Leclère (1971) descreveu predicados nominais formados por vários *Vsup*, tais como, *faire* (*fazer*), *donner* (*dar*) e *avoir* (*ter*); dentre outros trabalhos.

Para o Inglês, Machonis (1988) descreveu, em extensão, os predicados formados pelo verbo-suporte *be Prep* (*ser de*) e, posteriormente, com o verbo-suporte *make* (*fazer*) (MACHONIS, 1991).

Em Português Europeu, destacam-se também as descrições, nos moldes do L-G, de predicados nominais com os verbos *estar com*, *ser de*, *fazer* e *dar*. Ranchhod (1990) descreveu cerca de 1.800 predicados nominais que se constroem com o *Vsup estar Prep*; em seguida, Baptista (2005b) descreveu 2.095 *CVS* com *ser de*; Chacoto (2005) descreveu outros 2.983 nomes predicativos que selecionam o *Vsup fazer*. Como descrições do Léxico-Gramática das *CVS* com o *Vsup dar* em PE, citam-se dois trabalhos: Vaza (1988) e Baptista (1997).

Vaza (1988), em sua dissertação de mestrado, recenseou 1.466 nomes predicativos que selecionam o verbo *dar* em PE e descreveu esses predicados em matrizes do Léxico-Gramática. Naquele trabalho, no entanto, há inconsistências tanto de ordem teórica quanto metodológica. Algumas dessas inconsistências já foram indicadas na seção 2.1.3 e

citem-se outras, a título de exemplo: (i) *dar palpite* é considerado pela autora como uma construção com verbo-operador causativo, e não como uma *CVS*; (ii) em contrapartida, inúmeros casos de construções com verbo-operador causativo são considerados como *CVS*, tais como, *dar pena, dar agonia, dar náusea etc.*; (iii) a autora expande alguns nomes predicativos no diminutivo, considerando *beijo*, *beijoca* e *beijinho* como três entradas diferentes, mas não o faz sistematicamente para os outros *Npred*; (iv) *dar o fora*, no sentido de *sair*, é considerado pela autora como *CVS*, ao passo que nós classificamos essa construção como fixa; (v) *dar à-vontade* também é classificado como uma *CVS*, sendo que, na verdade, *à vontade* é uma locução adverbial de modo e não o complemento de *dar*. Apesar de algumas inconsistências, muitos nomes predicativos foram aproveitados da lista de Vaza (1988), procedendo-se a uma análise mais sistemática e coerente de cada caso.

Já o trabalho de Baptista (1997) apresenta uma análise sistemática das construções conversas com os verbos *dar* (em construções *standard*) e *levar* (em construções conversas). Baptista (1997) analisou 292 nomes predicativos que admitem conversão com *levar*. Falta ainda descrever os demais nomes predicativos, que fazem conversão com *receber* ou *ter*, ou ainda aqueles que não admitem conversão. Posteriormente, Rassi *et al.* (2015) fizeram um estudo contrastivo entre as construções com *dar/levar* em PB e PE. Os resultados daquele trabalho provaram que há mais diferenças do que semelhanças entre as duas variantes do Português, no que tange às construções conversas do *Vsup dar*.

Para o Português Brasileiro também existem trabalhos que descrevem ou explicam o funcionamento do *Vsup dar*, mas esses trabalhos são principalmente de base funcionalista, visando a explicar a motivação para o uso dos *Vsup*, ou o processo de gramaticalização pelo qual esses verbos passam. Os trabalhos que apresentam uma descrição e formalização de *CVS*, nos moldes do L-G, considerando um grande número de construções, são: Barros (2014), que descreveu e classificou cerca de 2.000 predicados nominais com o *Vsup fazer*; e Santos (2015), que descreveu e classificou cerca de 3.000 predicados nominais com o *Vsup ter*.

Somando-se os nomes predicativos que selecionam os verbos *estar com*, *ser de*, *fazer*, *ter* e *dar*, e desconsiderando-se as diferenças entre as duas variantes do Português, é possível criar um recurso léxico-sintático com cerca de 10.000 nomes predicativos, já descritos e formalizados em Português.

3.3 Identificação e/ou processamento automático das *CVS*

A literatura dispõe de, pelo menos, duas abordagens distintas para análise de *CVS*: (i) uma que considera as *CVS* como um bloco único cujos constituintes são relativamente fixos, tais como outras expressões multipalavras (*multiword expressions - MWE*) (CALZOLARI *et al.*, 2002; SAG *et al.*, 2002; DIAB; BHUTADA, 2009); e (ii) uma abordagem 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 sistematicamente certas transformações sintáticas, apesar de algumas serem lexicalmente determinadas (GIRY-SCHNEIDER, 1987; RANCHHOD, 1990; BAPTISTA, 2005b). Neste trabalho, adotamos a segunda abordagem, que reconhece e descreve as relações sintáticas e semânticas entre os constituintes de uma *CVS*.

Grande parte dos trabalhos encontrados na literatura focam apenas na identificação ou extração automática das *CVS* em *corpora*. Pouco se conhece em relação à implementação de *CVS* em *parser* ou análise sintática automática dessas construções. Alguns poucos trabalhos que fizeram o processamento automático de *CVS*, por exemplo, os de Butt (1994), Wang e Ikeda (2008) e Vincze (2012), referem-se a alguma aplicação específica, como é o caso da tradução automática. Naqueles trabalhos, as *CVS* são consideradas cadeias não-analisáveis de constituintes, em um bloco fixo, tal como as palavras compostas ou as expressões cristalizadas.

Já os estudos que visam à identificação ou extração dessas construções são, muitas vezes, baseados em padrões lexicais (expressões regulares, por exemplo), combinados com listas de potenciais candidatos a *Vsup* ou candidatos a *Npred* (GREFENSTETTE; TEUFEL, 1995; DURAN *et al.*, 2011; PÁEZ, 2014).

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 ou adjetivos, 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. Um 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. Por outro lado, restringe-se apenas aos *Npred* que apresentam nominalizações, excluindo da análise os *Npred* autônomos.

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 *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, é comum encontrar trabalhos, como o de Páez (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 difere dessas abordagens por partir dos padrões sintáticos de combinações de categorias gramaticais (*POS tags*) para encontrar *CVS*, por exemplo, [*V N Prep*] (*abrir mão de*), [*V Prep N*] (*deixar de lado*), [*V Det N Prep*] (*virar as costas para*), [*V Det Adv*] (*dar o fora*), [*V Adv*] (*ir atrás*), [*V Prep Adv*] (*dar para trás*), [*V Adj*] (*dar duro*).

Usando esse método, Duran *et al.* (2011) conseguiram identificar 773 predicados complexos, que foram, posteriormente, anotados. Segundo os autores, esses predicados complexos incluem (mas não se limitam a) construções com verbos leves[25]. Consideramos, no entanto, que a utilização de expressões regulares ou combinações de POS não seja adequada para a identificação 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*).

Existem várias outras abordagens que objetivam a identificação, extração ou detecção de *CVS*, também chamadas na literatura de *construções com verbo leve* (ou *light verb*

25 O objeto que Duran *et al.* (2011) e outros autores designam como *verbo leve* é o mesmo que nós designamos, neste trabalho, como *verbo-suporte*. Existem diferenças conceituais entre os dois termos, que serão explicadas no Capítulo 7, mas, por enquanto, considerem-se os termos como sinônimos.

constructions - LVC). Essas abordagens podem ser baseadas: (i) apenas em informação linguística, como o grau de composicionalidade dos constituintes das *CVS*, padrões sintáticos ou anotação de *corpus* com *CVS*; ou (ii) apenas em informação estatística, por exemplo, medidas de associação ou técnicas de co-ocorrência de palavras; ou ainda (iii) em abordagens híbridas que levam em consideração tanto as informações linguísticas quanto estatísticas (STEVENSON; FAZLY; NORTH, 2004; TAN; KAN; CUI, 2006; FAZLY; STEVENSON, 2007; CRUYS; MOIRÓN, 2007; WANG; IKEDA, 2008; DIAB; BHUTADA, 2009; TU; ROTH, 2011; GURRUTXAGA; ALEGRIA, 2011). Os métodos híbridos são os mais comuns hoje em dia para a identificação, extração ou detecção de expressões multipalavra em *corpus*.

Stevenson, Fazly e North (2004) propõem uma medida estatística para mensurar o grau de aceitabilidade das construções com verbo leve, com base em propriedades linguísticas, principalmente na ordenação de classes de complementos semanticamente semelhantes. Tan, Kan e Cui (2006) e Tu e Roth (2011) construíram diferentes sistemas de aprendizagem que incorporam sistematicamente medidas estatísticas informativas e contextos locais específicos. Os dois trabalhos diferem entre si em relação ao conjunto de dados anotados e os conjuntos de verbos considerados para a anotação. Fazly e Stevenson (2007) usam medidas estatísticas linguisticamente motivadas para distinguir subtipos de combinações de verbo-nome. Cruys e Moirón (2007) apresentam um método híbrido baseado em semântica e estatística, com vistas a identificar combinações de [*V Prep N*] em Alemão. Esse método é baseado em preferências de seleção para ambas as categorias: um nome usado como verbo e um verbo usado como nome, nos casos de nominalização. Wang e Ikeda (2008) propõem regras de tradução baseadas em padrões sintáticos, visando à tradução de *CVS* do Japonês para o Chinês. Diab e Bhutada (2009) identificaram os graus de idiomaticidade e classificaram combinações de [*V N*], usando um método supervisionado de aprendizagem de máquina que combina diferentes características linguisticamente motivadas. Gurrutxaga e Alegria (2011) exploraram várias técnicas para extrair automaticamente combinações de [*N V*] em Basco, usando uma abordagem híbrida que considera as complexas combinações de características, tais como, a institucionalização, a não-composicionalidade e a fixidez léxico-sintática dos constituintes.

Vários outros trabalhos que seguem a mesma linha podem ser citados, mas ressalte-se que todos eles consideram *CVS* como um tipo de expressão multipalavra (*MWE*) e nenhum deles analisa as propriedades linguísticas inerentes a essas construções.

Por fim, refira-se ainda ao trabalho de Barreiro *et al.* (2014), que avaliou dois sistemas de tradução automática, o OpenLogos (baseado em regras) e o Google Translate (baseado em métodos estatísticos), 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* 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 rico conhecimento lexical do primeiro sistema.

O conceito de *CVS* que os autores adotaram, naquele trabalho, é diferente do que adotamos aqui. Para Barreiro *et al.* (2014, p.37), "uma *CVS* é uma expressão multipalavra ou um predicado complexo, que consiste em um verbo semanticamente fraco (o verbo-suporte) e um nome predicativo (que é o mais comum), ou um adjetivo predicativo ou um advérbio predicativo"[26].

Para justificar a definição adotada de *CVS*, os autores usam exemplos como: (i) *fazer uma apresentação* (*make a presentation*), que é uma *CVS* formada pelo verbo-suporte *fazer* e o nome predicativo *apresentação*; (ii) *fazer isso simples* (*make it simple*), que também é considerado uma *CVS* formada pelo verbo-suporte *fazer* e o adjetivo predicativo *simples*; (iii) *ser significativo* (*be meaningful*), como *CVS* formada pelo verbo-suporte *ser* e o adjetivo predicativo *significativo*; (iv) *ser conhecido como* (*be known as*) e *estar envolvido em* (*be involved in*), como *CVS* do tipo construção com verbo-suporte adjetival preposicional, por serem formadas por um verbo-suporte (*ser* ou *estar*), um adjetivo predicativo (*conhecido* ou *envolvido*) e uma preposição (*como* ou *em*); dentre outros exemplos. Pela abordagem que adotamos, apenas o exemplo (i) é, de fato, uma *CVS*. No exemplo (ii), *fazer* deveria ser classificado como verbo-operador causativo, que indica *fazer com que isso seja simples* ou *tornar isso simples*. Os exemplos (iii) e (iv), em nossa análise, são construções adjetivais formadas por um verbo de cópula (*ser* ou *estar*), que não é um *Vsup*, e um adjetivo predicativo.

As construções que Barreiro *et al.* (2014) identificam como *CVS* são chamadas por outros autores de predicados complexos (DURAN *et al.*, 2011) ou *expressões multipalavras* (*multiword expressions*) (CALZOLARI *et al.*, 2002; DIAB; BHUTADA, 2009; SAG *et al.*, 2002). Há diferentes objetos linguísticos sendo considerados como um tipo único de construção.

Este é o último capítulo dessa parte introdutória do *e-book*. Nele, oferecemos ao leitor um breve panorama dos estudos linguísticos sobre a classificação dos verbos

26 Tradução minha. Do original: "A *SVC* is a multiword or complex predicate consisting of a semantically weak verb (the support verb), and a predicate noun (most commonly), a predicate adjective, or a predicate adverb" (BARREIRO *et al.*, 2014, p.37).

em Português. Indicamos alguns dicionários (BORBA, 1991; BUSSE, 1994) e trabalhos acadêmicos (CANÇADO; GODOY; AMARAL, 2012; LEVIN, 1993; BAPTISTA, 2012) que propuseram diferentes classificações verbais, sejam elas sintáticas ou semânticas. Em seguida, indicamos também trabalhos que descrevem tipos específicos de construções verbais, tais como, as construções causativas (CANÇADO, 2010; GROSS, 1981) e as expressões fixas (VALE, 2001) etc. Esses trabalhos serão utilizados, posteriormente, para fazer a análise e descrição das construções com o verbo *dar*.

Ainda neste capítulo, apresentamos os principais estudos sobre a identificação e descrição das *CVS*, tanto em Português quanto em várias outras línguas. Por fim, apresentamos o estado da arte do processamento automático de *CVS*. Apesar de haver muitos estudos que propõem a identificação ou extração automática de *CVS* em *corpora*, grande parte desses estudos consideram as *CVS* como expressões multipalavras e não analisam sua estrutura sintática interna.

A partir do próximo capítulo, que analisa as construções em que *dar* funciona como verbo pleno, inicia-se a primeira parte de análise e descrição das construções com o verbo *dar* em Português.

Parte I

Proposta tipológica para o verbo *dar*

4. O verbo pleno *dar*

Os verbos plenos são também chamados de verbos distribucionais, pois são eles os responsáveis pela distribuição dos argumentos, ou seja, são eles que selecionam seus argumentos. Para Travaglia (2003), que estuda processos de gramaticalização, o primeiro estágio dos fenômenos linguísticos, a partir de onde se iniciam os processos de gramaticalização, é aquele em que os verbos têm seu significado lexical pleno e seus objetos se referem tipicamente a um objeto concreto. Com base nessas restrições, reconhecem-se oito subtipos do verbo *dar* como pleno: (i) o verbo com complemento dativo; (ii) com complemento reduzido (ou apagado); (iii) com complemento locativo; (iv) com complemento quantificador; (v) com complemento instrumental; (vi) com complemento não-restrito introduzido pela *prep com*; (vii) o verbo transitivo-predicativo; e (viii) o verbo pleno com complemento espaço-temporal.

4.1 Verbo pleno com complemento dativo

Para que seja considerado como verbo pleno acompanhado de complemento dativo, o verbo *dar* deve apresentar um sujeito humano (*Nhum*), agente, volitivo, que é responsável pela transferência física de um objeto concreto a um beneficiário. A forma plena mais prototípica do verbo *dar* em PB é com complemento dativo:

(4.1) *Um dos islamitas que atacaram o shopping de Nairóbi deu um chocolate a um menino britânico de 4 anos.* [Ex.R]

(4.2) *A Lourdes, psicanalista, deu um computador às sobrinhas.* [Ex.R]

Nesses dois casos, transfere-se a posse do objeto (*chocolate* ou *computador*) das mãos de *N*0 (*Um dos islamitas que atacaram o shopping de Nairóbi* ou *A Lourdes*) para as mãos de *N*2 (*um menino britânico de 4 anos* ou *as sobrinhas*).

Em todas as línguas latinas, esse verbo possui o sentido dativo e é reconhecido por todas as gramáticas e dicionários consultados como verbo bitransitivo que indica dação ou doação, ou seja, sua estrutura argumental requer o preenchimento de três posições argumentais: o sujeito, o objeto direto e o objeto indireto, portanto sua estrutura pode ser formalizada como [*N*0 *Vdar* (*Det*) *N*1 *para N*2], sendo que *N*1 é um objeto concreto. Em PB, sabe-se que a preposição *para*, introduzindo complementos dativos, pode comutar com as preposições *a* e *pra*.

Em outros casos, o verbo *dar* pode ser parafraseado por outro operador verbal com o sentido de *oferecer*, como em:

(4.3) *Neymar deu uma festinha à fantasia em sua casa na noite deste sábado.* [Ex.R]

(4.4) *Minha mãe deu um almoço em casa para toda a nossa família.* [Ex.R]

Há nomes, como *jantar*, *churrasco*, *cervejada*, *feijoada*, *churrascada*, *farofada*, *ceia*, *banquete*, *baile*, *boca livre*, *lanche*, *show*, *recital*, *espetáculo*, *concerto* e outros, que podem ser identificados como tipos de festas. Todos esses nomes são predicativos e podem aparecer em construções com verbo-suporte, mas o *Vsup* de base para essa classe é o *fazer*, e não o *dar*. Quando esses nomes são associados ao verbo *dar*, essas construções são dativas. Em outras palavras, as construções de base seriam:

(4.3a) *Neymar fez uma festinha.* [Ex.C]

(4.4a) *Mirela fez um almoço.* [Ex.C]

Ambos os exemplos se classificam como construções com o verbo-suporte *fazer*, porém, quando o *Npred* é associado ao verbo *dar*, a construção exige o complemento dativo, o qual é selecionado pelo verbo *dar* e não pelo nome predicativo *festinha* ou *almoço*:

(4.3b) *Neymar fez uma festinha e a ofereceu a seus amigos.* [Ex.C]

(4.4b) *Mirela fez um almoço e o ofereceu a seus convidados.* [Ex.C]

Nos casos em que os argumentos são selecionados pelo verbo e não pelo nome, significa que o verbo é pleno. Os exemplos apresentados até agora não geram discordância de análise, mas há exemplos no *corpus* cuja classificação pode ser discutível. A classe dos nomes de documentos, por exemplo, poderia gerar dúvidas de classificação, pois os nomes de documentos possuem propriedades de nomes predicativos – quando associados a um verbo-suporte – ao mesmo tempo que podem se referir a objetos concretos – como complementos diretos em construções com verbo pleno. É o caso de construções como:

(4.5) *Você deu uma procuração para uma advogada justamente para ela fazer tudo sozinha, sem você.* [Ex.R]

(4.6) *Os médicos temem o risco de dar atestado de sanidade a alguém incapaz.* [Ex.R]

(4.7) *Agnelo deu um certificado para a farmacêutica participar de licitação.* [Ex.R]

Os nomes *procuração*, *atestado* e *certificado*, bem como *declaração*, *recibo*, *certidão*, *multa*, *liminar*, *licença-maternidade*, *diploma*, *habeas corpus* e outros podem ser entendidos como substantivos concretos, por serem representados por um documento (objeto físico) que exerce a função de *atestar*, *certificar*, *declarar* etc., porém esses nomes

serão considerados neste trabalho como nomes predicativos, pois são esses nomes que selecionam o sujeito da construção [*dar Npred*].

Há na literatura uma distinção entre nominalizações reais e os usos concretos das formas nominalizadas (GREFENSTETTE; TEUFEL, 1995). Alguns nomes deverbais, como *presente*, *proposta* e os nomes predicativos que coincidem com nomes de documentos, por exemplo, podem se tornar concretos por uma associação metonímica com o nome predicativo.

(4.8) *Zé fez uma proposta à Ana.* [Ex.C]

(4.9) *Zé guardou a proposta na gaveta.* [Ex.C]

Em (4.8), o nome *proposta* é considerado uma nominalização real, ou seja, é um nome predicativo derivado de *propor* e constitui o elemento predicador da sentença; é ele quem seleciona o sujeito e o verbo-suporte *fazer*. Já em (4.9), o nome *proposta* se refere a um objeto físico e consiste em um uso concreto do nome *proposta*, que é apenas um dos argumentos selecionados pelo predicador *guardar*.

Já que esses nomes podem exercer a função de nomes predicativos e serem associados ao verbo-suporte, eles serão discutidos mais profundamente na Parte II.

Vale destacar que, muitas vezes, o objeto direto (complemento dativo) vem acompanhado de advérbios ou expressões adverbiais com valor de modo, tais como, *de brinde*, *de bandeja*, *de lambuja*, *de graça*, *de presente etc*. Nem sempre esses modificadores possuem interpretação literal e são, muitas vezes, confundidos com complementos verbais fixos. Esses modificadores devem ser analisados como expressões cristalizadas adverbiais, porém a construção, como um todo, deve ser interpretada como uma construção com verbo pleno *dar* e complemento dativo, que pode se deslocar de posição, como em:

(4.10) *O zagueiro Gonzales* ***deu de bandeja*** *a bola para o atacante.* [Ex.R]

= *O zagueiro Gonzales* ***deu*** *a bola* ***de bandeja*** *para o atacante.* [Ex.C]

Esses dois exemplos são equivalentes e expressam o mesmo predicado semântico. Em ambos, o verbo *dar* é pleno, responsável pela transferência de um objeto (*a bola*) de um doador (*O zagueiro Gonzales*) a um recipiente (*o atacante*). A expressão fixa adverbial *de bandeja* pode ser um modificador do verbo pleno *dar* e, como todo sintagma adverbial, pode ser deslocado de posição ou ainda pode modificar uma construção nominal.

(4.11) *Mas ainda não sei o que fazer e nem vou dar uma informação de bandeja ao Mano Menezes.* [Ex.R]

Consideram-se construções com o verbo pleno *dar* também aquelas que indicam transferência de objeto, mas não denotam *dação* ou *doação*. O verbo pleno *dar* pode ter cerca de 50 sentidos, incluindo: *vender*, *pagar*, *doar*, *emprestar*, *entregar*, *enviar*, *mandar*, *transferir*, *destinar*, *trocar*, dentre outros.

4.2 Verbo pleno com complemento reduzido/apagado

Há um grupo restrito de construções que apresentam a mesma estrutura sintática de [*N0 Vdar de V inf*], em que o verbo *dar* corresponde semanticamente a *oferecer* e o verbo na posição de *Vinf* pode ser nominalizado, como é o caso de:

(4.12) *Se tem gente com fome, a gente* ***dá de comer****.* [Ex.C]

= *Se tem gente com fome, a gente* ***dá [algo] de comer****.* [Ex.C]

= *Se tem gente com fome, a gente* ***oferece comida****.* [Ex.C]

(4.13) *Se o menino estava com sede, ele* ***dava de beber****.* [Ex.C]

= *Se o menino estava com sede, ele* ***dava [algo] de beber****.* [Ex.C]

= *Se o menino estava com sede, ele* ***oferecia bebida****.* [Ex.C]

Em alguns casos, o complemento dativo não está apagado ou reduzido, o que comprova que essas construções apresentam complemento (explícito) e podem ser definidas intencionalmente; daí classificá-las como construções com verbo pleno.

(4.14) *Vem esse tanto de gente e eu tenho que achar* ***o que dar de jantar****.* [Ex.C]

= *Vem esse tanto de gente e eu tenho que achar* ***o que oferecer como janta****.* [Ex.C]

Nesses exemplos, o argumento *o que* funciona como objeto direto preposto ao verbo distribucional *dar* (ou *oferecer*).

O Dicionário Gramatical de verbos do Português contemporâneo do Brasil (BORBA, 1991, p.364) explica que, nesses casos, "o primeiro complemento pode ser expresso pelo seu adjunto: quando lhe davam [alguma coisa] de comer, comia (ED, 97); Manuel dá de beber a Tonho (GE, 74)".

4.3 Verbo pleno com complemento locativo

Quando o verbo *dar* exige um complemento locativo, introduzido por preposição, também é classificado como um verbo pleno, pois a construção cuja estrutura é [*N0 Vdar para Nloc*] é dotada de significado e pode ser descrita intencionalmente. Essa construção é formada pelo verbo *dar*, seguida da preposição *para* e de um locativo, designado por *Nloc*. Em PB, a preposição *para*, que introduz complementos locativos, pode comutar com a preposição *pra*, usada em contextos menos formais. O elemento locativo expressa local, direção, rumo, e a construção toda indica a direção para onde determinado objeto está voltado, para onde possui vista.

(4.15) *sozinha sob o umbral da porta que* ***dá para o campo****.* [Ex.R]

(4.16) *e fomos grudados nele até a porta que* ***dá para o elevador de serviço****.* [Ex.R]

(4.17) *os quartos ganham algum charme com a varanda que* ***dá para o rio****.* [Ex.R]

(4.18) *Jantar no terraço do hotel, que* ***dá para os jardins****, é um prazer.* [Ex.R]

(4.19) *a Rocinha estava começando a se expandir pela vertente do morro Dois Irmãos, que* ***dá para o condomínio****.* [Ex.R]

(4.20) *Como sonhar com a paisagem na janela, se ela* ***dá para a cinza praça Roosevelt****?* [Ex.R]

O verbo pleno *dar*, nas construções locativas, impõe fortes restrições ao preenchimento lexical tanto do sujeito quanto do complemento. Ambos os argumentos só podem ser instanciados por nomes locativos. A construção toda deve ser interpretada e traduzida para outras línguas com o sentido de "*x* ter a frente voltada para *y*", em que *x* e *y* são representados por *Nloc*.

A segunda estrutura sintática do verbo pleno *dar* com complemento locativo também exige dois argumentos locativos, um na posição sujeito e outro na posição de complemento. Essas construções são representadas por [*N0 Vdar em Nloc*].

(4.21) *A estrada* ***dá em Paris****, na França.* [Ex.R]

(4.22) *Existe um caminho que* ***dá no mar****.* [Ex.R]

(4.23) *Tem uma trilha que* ***dá em uma cachoeira****.* [Ex.R]

A preposição selecionada nessa estrutura é *em* e a construção toda tem a ideia de finalização de um percurso. O verbo *dar* pode ser substituído por *chegar* ou *acabar*.

4.4 Verbo pleno com complemento quantificador

Existem duas construções sintáticas diferentes em que o verbo pleno *dar* exige um complemento quantificador e, em termos semânticos, tem o sentido de suficiência: (i) [*N0 Vdar para Num*], que pode ser reduzido para [*Vdar*], também indicando quantidade suficiente; e (ii) [*N0 Vdar Det N1*], que não indica quantidade, mas modaliza a ideia de suficiência.

As construções com [*N0 Vdar para Num*] indicam que determinada quantidade (de coisas) é suficiente para outra determinada quantidade (de pessoas, por exemplo).

(4.24) *O carro-chefe é o pintado na brasa, que **dá para duas** pessoas.* [Ex.R]

(4.25) *Eu cubro a minha namorada mesmo quando o edredom só **dá pra um**.* [Ex.R]

(4.26) *Do lado esquerdo fica uma garagem que **dá pra 10** carros.* [Ex.R]

A mesma ideia de suficiência, por vezes, apresenta-se em uma construção reduzida da estrutura anterior, resultando apenas na construção [*Vdar*]. Na redução, a preposição *para* e o segundo argumento (*Num*) são apagados. Pelo contexto, é possível inferir que o sentido empregado pelo verbo *dar* é o mesmo que indica suficiência de quantidade:

(4.27) *Que **dá**, **dá**.* [Ex.R]

(4.28) *Coloquei só uma receita, não sei se **dá**.* [Ex.R]

(4.29) É só acrescentar mais água no feijão, aí é certeza que ***dá***. [Ex.R]

Existe ainda uma expressão cristalizada semelhante à construção [*Vdar para N*] que também indica suficiência (ou quantidade suficiente):

(4.30) *O pincel não é a oitava maravilha, mas **dá pro gasto**.* [Ex.R]

A expressão *dar para o gasto* é considerada uma expressão fixa porque todos os seus elementos (verbo, preposição, determinante e nome) são fixos e não admitem variação. Por ser uma *EC*, essa expressão será analisada na seção 6.1.

O segundo tipo de construção com verbo pleno indicando suficiência é formalizado pela estrutura [*N0 Vdar Det N1*] e pode ser exemplificado como:

(4.31) *O caso do estudante de medicina Diogo Nogueira **dá uma tese de doutorado**.* [Ex.R]

(4.32) *Minha vida **dá um livro**.* [Ex.R]

Borba (1991, p.364) se refere a essas construções como "2. Com sujeito causativo [...] 2.2 Com complemento expresso por nome indicativo de produto, significa *constituir*, *formar*". Não consideramos que o sujeito dessas construções seja tipicamente um causativo, mas sim um nome predicativo. Além disso, existem fortes restrições do verbo *dar* ao preenchimento lexical do complemento, que, em geral, é um nome de gênero textual (*livro, tese, romance, filme, epopeia, comédia, etc.*).

4.5 Verbo pleno com complemento instrumental

Nessas construções, o verbo pleno *dar* exige dois complementos preposicionados, um instrumental e outro não-restrito. Em termos semânticos, essas construções têm o sentido de *bater*, como nos exemplos seguintes:

(4.33) ***Deu com*** *a mão* ***na*** *cara do infiel, que ficou a marca vermelha da mão.* [Ex.R]

(4.34) *Então veio Rosangela e lhe* ***deu com*** *um tijolo* ***nas*** *costas.* [Ex.R]

(4.35) *Tyler* ***deu com*** *uma cadeira* ***na*** *cabeça dele.* [Ex.R]

Essas construções são do tipo *Onnn* (com três argumentos), sendo que: (i) o primeiro é o sujeito humano agente; (ii) o segundo argumento, introduzido pela preposição *com*, é preenchido por um nome concreto de instrumento; e (iii) o terceiro argumento, introduzido pela preposição *em*, pode ser preenchido por um nome parte do corpo ou, metonimicamente, por um nome humano. Esse terceiro argumento pode ainda ser instanciado por um objeto-alvo ou nome locativo, como em:

(4.36) *E o lobo foi buscar uma marreta na casa dele e* ***deu com*** *a marreta* ***na*** *casa e destruiu a casa.* [Ex.R]

O ViPEr (BAPTISTA, 2012) classifica esses casos como pertencentes à classe 40, que é formada por construções plenas do tipo [*N*0 *V Prep*1 *N*1 *Prep*2 *N*2]. A glosa indicada no ViPEr para o Inglês é o verbo '*hit*'.

Essas construções podem ter dado origem a *CVS* de atos violentos, tais como:

(4.37) ***Deu uma mãozada*** *na cara do infiel, que ficou a marca vermelha.* [Ex.C]

(4.38) *Então veio Rosangela e lhe* ***deu uma tijolada*** *nas costas.* [Ex.C]

(4.39) *Tyler* ***deu uma cadeirada*** *na cabeça dele.* [Ex.C]

Apesar de serem semanticamente semelhantes, já que ambas indicam atos violentos, essas duas construções são sintaticamente diferentes e devem ser classificadas também de forma diferente. Os exemplos de (4.33) a (4.35) são classificados como construções

com verbo pleno e complemento instrumental, ao passo que os exemplos de (4.37) a (4.39) são classificados como *CVS*. A análise dessas *CVS* será retomada na Parte II.

No *corpus*, encontram-se frases em que o verbo *dar* significa *bater*, mas o complemento instrumental, na posição de *N*1, encontra-se apagado. É o caso de:

(4.40) *Quanto ao rapaz que* ***deu na*** *cara do secretário, sem moral, torço para que ele se arrependa.* [Ex.R]

Vale ressaltar que existe uma expressão cristalizada, em Português, que mantém a mesma estrutura sintática dessas construções plenas:

(4.41) *Gustavo saiu em busca de socorro, mas só* ***deu com a cara na porta****.* [Ex.R]

Essa construção é considerada fixa, e não livre, porque todos os complementos são fixos, além de ter sentido metafórico, que não pode ser inferido pelo significado de seus componentes. Por ser uma expressão fixa, será analisada na subseção 6.1.7.

4.6 Verbo pleno com complemento não-restrito introduzido pela *prep 'com'*

Algumas construções com o verbo pleno *dar* exigem um complemento não-restrito, que é introduzido pela preposição *com*, e indicam o processo de *encontrar algo ou alguém*. Essas construções podem ser representadas pela estrutura [*N*0 *Vdar com N*1]:

(4.42) *Aberta a mala o carrinheiro* ***deu com*** *o corpo de uma mulher cortado em pedaços.* [Ex.R]

(4.43) *Fabian abriu a porta e* ***deu com*** *o menino deitado na cama.* [Ex.R]

Essa construção se distingue das apresentadas na seção 4.5 por três razões: (i) aquelas apresentam três argumentos ([*N*0 *V com N*1 *em N*2]), ao passo que estas apresentam tipicamente dois argumentos ([*N*0 *V com N*1]); (ii) o argumento introduzido pela preposição *com* naquelas construções representa o INSTRUMENT, ao passo que o argumento introduzido pela *prep com* nesta última é um PATIENT; e (iii) semanticamente aquelas indicam um ato violento, ao passo que estas indicam um encontro.

Essa estrutura pode ter sido derivada da estrutura de uma expressão fixa, que é [*N*0 *Vdar de cara com N*1]. Classificamos [*N*0 *Vdar com N*1] como uma construção livre e [*N*0 *Vdar de cara com N*1] como uma expressão fixa porque a primeira construção tem os argumentos livres e significado literal (considerando-se *dar* como sinônimo de *encontrar*), ao passo que a segunda construção tem um complemento *de cara* fixo e seu significado global não é literal. A expressão fixa será analisada na subseção 6.1.7.

4.7 Verbo pleno transitivo-predicativo

O verbo *dar*, em algumas construções, tem o significado de *considerar* e pode entrar em dois tipos de estruturas sintáticas transitivo-predicativas: (i) [*N0 Vdar-se por Adj*]; e (ii) [*N0 Vdar N1 (por + como) Adj*]. Em ambos os casos, o verbo *dar* é distribucional e exige um complemento predicativo, seja do sujeito seja do objeto.

A primeira estrutura apresenta o verbo pronominal (*dar-se*), seguido da preposição *por*, seguida ainda de um adjetivo predicativo do sujeito *N0*. Essa estrutura engloba frases como:

(4.44) *Silveirinha não* ***se deu por vencido***. [Ex.R]

(4.45) *O time de Campos não* ***se deu por satisfeito*** *e buscou o segundo.* [Ex.R]

Poder-se-ia argumentar que essas construções deveriam ser classificadas como expressões fixas, já que os complementos *vencido* e *satisfeito* parecem fixos. No entanto, no *corpus*, encontramos vários outros adjetivos no lugar de *vencido* e *satisfeito*, o que não nos permite considerar que seja uma posição fixa. Seguem-se alguns exemplos:

(4.46) *Por comparecimento espontâneo, entende-se aquele em que o réu* ***se dá por citado*** *nos autos.* [Ex.R]

(4.47) *A cota de nomes indicados pelo Partido dos Trabalhadores (PT) no governo de Robinson Faria já* ***se deu por encerrada***. [Ex.R]

(4.48) ***Deu-se por lido e conferido*** *o acórdão.* [Ex.R]

(4.49) *foi bem pago e* ***se deu por quite***. [Ex.R]

Em todos esses casos, de (4.44) a (4.49), o verbo pronominal *dar-se* pede um complemento indireto, introduzido pela preposição *por* e esse complemento é predicativo do sujeito; daí ser classificado como *verbo transitivo-predicativo*. O verbo *dar* pode comutar com o verbo *considerar*, que também é classificado como transitivo-predicativo. Quando se usa o verbo *considerar-se*, no entanto, a preposição selecionada é *como*.

A segunda estrutura sintática mencionada anteriormente [*N0 Vdar N1 (por + como) Adj*] difere da primeira estrutura nos seguintes aspectos: (i) o verbo não é pronominal; (ii) a preposição que introduz o complemento indireto pode ser *por* ou *como*; e (iii) o *Adj* é um predicativo do objeto e não do sujeito. Essa estrutura pode ser exemplificada por:

(4.50) *Tribunal de Justiça do Estado* ***deu como desaparecido*** *o prefeito de Itapema.* [Ex.R]

(4.51) *Tribunal **deu como provado** que Luís Catarino é psicopata.* [Ex.R]

(4.52) *Em dezembro de 2008, a empreiteira **deu como acabado** o trabalho de assentamento do piso em paver.* [Ex.R]

A posição de *Adj* também pode ser preenchida por um verbo no particípio, como se nota em (4.52).

4.8 Verbo pleno com complemento não-preposicionado

O verbo pleno *dar*, por vezes, exige um complemento não-preposicionado que pode exprimir três valores diferentes: (i) tempo, (ii) distância, ou (iii) existência, conforme atestam os exemplos:

(4.53) *A viagem de carro até São Paulo dá 2 horas e meia.* [Ex.R]

(4.54) *Daqui lá dá uns 8 quilômetros, mais ou menos.* [Ex.R]

(4.55) *Hoje no meu show deu mil pessoas.* [Ex.R]

Essas construções possuem a estrutura sintática [*N*0 *Vdar N*1], sendo que a posição de *N*0 pode ser preenchida ou não. Mesmo quando é preenchida, sofre fracas restrições de seleção, podendo aceitar um nome predicativo, uma frase completiva, uma expressão adverbial ou outras. Já a posição de *N*1 deve ser preenchida por um complemento não preposicionado geralmente introduzido por um modificador numérico. Em (4.53) e (4.54), o verbo *dar* pode comutar com os verbos *ser* e *corresponder a*. Em (4.55), o verbo *dar* tem valor existencial e pode comutar com os verbos *haver* ou *ter*.

Existem outras construções temporais do Português que também foram inseridas nesta classe.

(4.56) *Deu duas horas e meia da tarde e Amy já estava lá na praia esperando.* [Ex.R]

(4.57) *Deu 5 horas, vamos embora que vai fechar a loja.* [Ex.R]

Essas construções, assim como (4.55), poderiam ser classificadas como impessoais, já que o verbo *deu* (singular) não concorda em número com os elementos plurais *horas* (4.56) e (4.57) ou *pessoas* (4.55). É possível, no entanto, encontrar construções muito semelhantes em que o verbo concorda com o suposto sujeito, como se verifica em:

(4.58) *No hospital deram cem pessoas.* [Ex.R]

(4.59) *Deram duas horas e nada do Halim chegar.* [Ex.R]

(4.60) *São apenas 50 minutos de vôo de Barcelona, mas no total deram 5 horas e meia de viagem.* [Ex.R]

Em vez de classificar essas construções como pessoais ou impessoais, a partir do seu sujeito, classificamos todas elas como construções com verbo pleno que podem admitir ou não sujeito e que exigem um complemento não-preposicionado com valor espaço-temporal ou existencial.

Por fim, o verbo *dar* também pode ocorrer como pleno quando indica um resultado numérico de uma operação matemática. O verbo substitui o sinal de igual (=) numa operação matemática. Como exemplos desse tipo, citam-se:

(4.61) *Dois e dois dá exatamente quatro.* [Ex.R]

(4.62) *5 mais 8 dá 13.* [Ex.R]

Esses casos são bastante controlados linguisticamente e não apresentam grande dificuldade na análise.

Este capítulo, que introduz a primeira parte do livro, discutiu o conceito de construção com verbo pleno, ou também chamado verbo distribucional porque é o responsável pela distribuição dos seus argumentos, ou seja, é ele quem seleciona os seus argumentos necessários e obrigatórios. Neste trabalho, foram identificadas oito construções sintáticas diferentes em que o verbo *dar* é classificado como pleno: (i) com complemento dativo (*e.g. Rui deu um livro à Ana*), que pode ter cerca de 50 sentidos diferentes; (ii) com complemento reduzido ou apagado (*e.g. Rui deu de comer à Ana*); (iii) com complemento locativo (*e.g. O rio dá no mar*); (iv) com complemento quantificador (*e.g. A comida dá pra 10 pessoas*); (v) com complemento instrumental (*e.g. Rui deu com uma pedra na cabeça da Ana*); (vi) com complemento não-restrito introduzido pela *prep com* (*e.g. Rui deu com a Ana no parque*); (vii) verbo pleno transitivo-predicativo (*e.g. Rui se deu por satisfeito*); e (viii) verbo pleno com complemento não-preposicionado (*e.g. Deu mil pessoas na festa*).

Na sequência da taxonomia proposta para as construções com o verbo *dar*, no próximo capítulo, discutiremos os casos em que esse verbo é classificado como causativo, apresentando as propriedades formais e as configurações sintáticas dessas construções.

5. O verbo-operador causativo *dar*

O conceito de verbo-operador foi introduzido por Harris (1964) para designar os verbos que acompanhavam as nominalizações. Posteriormente, Gross (1981) redefiniu o conceito, distinguindo, de um lado, os verbos-suporte, e, de outro lado, os verbos-operadores. O conceito de verbo-operador, para Gross (1981, p.23-39), pode se referir ainda a dois tipos distintos de fenômenos e, portanto, dois tipos distintos de verbos: (i) verbo-operador causativo (*VopC*); e (ii) verbo-operador de ligação (*VopL*).

O verbo-operador causativo (*VopC*) tipicamente acrescenta um novo elemento à frase de base, sendo que esse elemento é facilmente identificável como a causa do evento. Distribucionalmente, esse elemento sofre fracas restrições semânticas, admitindo qualquer tipo de argumento (humano, não-humano, evento, frase completiva *etc.*) como causa, na posição de sujeito. Gross (1981) trata mais especificamente do verbo *faire (fazer)* como causativo, mas também cita construções causativas com *deixar*, *meter* e *dar*, seguidos por um verbo no infinitivo. Esse é o caso de:

(5.1) Zé fez # Ana dorme

≡ *O Zé fez a Ana dormir.* [Ex.C]

O segundo tipo de verbo-operador é chamado verbo-operador de ligação (*VopL*), o qual dificilmente modifica o significado da frase de base sobre a qual ele incide. Nas construções com *VopL*, também há a inserção de um elemento à frase de base, porém não é considerado um elemento novo.

(5.2) Zé tem # Ana está sob o controle de Zé

≡ *Zé tem Ana sob seu controle.* [Ex.C]

A noção de *Vop de ligação* já foi descrita para o Português Europeu a propósito das *CVS* com o *Vsup estar prep* (RANCHHOD, 1990).

Neste capítulo, abordaremos apenas o primeiro tipo de verbo-operador, haja vista que o verbo *dar* não funciona como verbo-operador de ligação, apenas como verbo-operador causativo. Foram encontradas no *corpus* cerca de 600 construções com o *VopC dar*, as quais serão discutidas nas seções seguintes.

5.1 Propriedades das construções com o *VopC dar*

Os verbos causativos formam uma classe extensa de vocábulos cujas propriedades transformacionais permitem estabelecer acarretamentos, sejam eles lexicais ou

sintáticos. Verbos como *causar*, *provocar*, *tornar*, *pôr*, *transformar*, *fazer* e *dar*, dentre outros, formam "um conjunto de verbos que se aplicam à frase elementar, acrescentando-lhe um argumento suplementar, e estabelecendo entre este e a frase à qual se aplicam uma relação de natureza tipicamente causal" (BAPTISTA, 2005b, p.202); daí serem nomeados verbos-operadores causativos.

As construções com verbo-operador causativo se distinguem das construções com verbo-suporte por possuírem propriedades sintático-semânticas específicas. Essas propriedades serão analisadas nas subseções seguintes.

5.1.1 Restrições ao preenchimento lexical do argumento na posição sujeito

A posição sujeito da construção com *VopC* sofre poucas restrições distribucionais e pode ser preenchida por elementos de categorias sintáticas variadas, por isso a posição de *N*0 é geralmente assinalada por um nome não-restrito (*Nnr*).

(5.3) *(O preenchimento do formulário + O fato de Rui ter preenchido o formulário + A falta de burocracia + Isso + Rui) deu agilidade ao processo.* [Ex.C]

Considerando o que foi apontado por Gross (1975, p.50), a notação "*Nnr*" não corresponde a uma classe específica de substantivos, mas a uma posição sintática que pode ser preenchida por um substantivo de qualquer natureza, ou uma frase completiva ou uma oração no infinitivo. O teste apresentado por Gross (1975) consiste em substituir a posição sintática por quaisquer dos seguintes grupos nominais: *Paulo* (*Nhum*) ou *este bolo* ([+concreto]) ou *A sinceridade* ([+abstrato]) ou *QueF* (*completiva*) ou *Vinf* (*infinitivo*).

Essa propriedade de *Nnr* é comumente marcada na posição *N*0 dos verbos-operadores causativos, conforme atestam as várias possibilidades de sujeito de (5.4).

(5.4) *(Rui + Esse carro + A infidelidade do Rui + O fato de Rui ter saído + A saída de Rui + Rui ter saído) deu raiva em Ana.* [Ex.C]

Todos esses tipos semânticos do argumento na posição sujeito também podem ser substituídos por *Isso* (*e.g. Isso deu raiva em Ana*), o que não ocorre nas construções com verbo-suporte (*e.g. *Isso deu um tapa em Ana*)[27]. Como forma de atestar sistematicamente essa propriedade em todos os *Npred* que autorizam a construção causativa com *VopC dar*, testamos a aceitabilidade de todas as construções causativas com o sujeito *Isso*.

27 Essa construção somente poderia ser aceitável se o locutor se referir pejorativamente ao agente da ação, que é humano, por meio do pronome *isso*, em vez de *ele* ou *ela*.

Oliveira (1979) tratou dos verbos psicológicos em Português Europeu e constatou que diferentes tipos sintáticos de argumentos podem ser sujeito de verbos psicológicos, e isso se aplica também às construções parafrásticas com o verbo *dar*, nos casos em que ele se classifica como *VopC*.

(5.5) *(Rui + Aquela caveira + O fato de Zé se atrasar + Que o Zé tenha se atrasado)*

(afligiu + emocionou + consternou + amedrontou) Ana. [Ex.C]

É possível construir paráfrases desses verbos psicológicos com construções com verbo-suporte cujo nome predicativo é considerado nome psicológico (ver exemplo 5.6). Por outro lado, também é possível encontrar na língua construções causativas nas quais a posição sujeito não é preenchida por nenhum elemento, assemelhando-se às construções impessoais (ver exemplo 5.7).

(5.6) *(Rui + Aquela caveira + O fato de Zé se atrasar + Que o Zé tenha se atrasado) deu (aflição + emoção + consternação + medo) em Ana.* [Ex.C]

(5.7) *Deu um(a) (aflição + emoção + consternação + medo) em Ana.* [Ex.C]

Algumas construções que apresentam nominalizações com sufixo *-ada* (*V-n-ada*) devem ser classificadas como construções causativas (em vez de *CVS*), por apresentarem *Nnr* na posição sujeito:

(5.8) *(Rui + A televisão + Que o Rui tenha saído + O fato de Rui ter saído + A saída de Rui + Rui ter saído) deu uma (acalmada + aliviada + melhorada) em Ana.* [Ex.C]

Pode-se considerar que as construções como (5.8) sofreram uma transformação, em que a frase de base é: *Ana deu uma (acalmada + aliviada + piorada)*. Na transformação da frase de base para a frase causativa (como aparece em 5.8): (i) foi introduzido o argumento CAUSA na posição sujeito; (ii) o argumento que era sujeito na frase de base (*Ana*) passa a ser complemento na frase causativa (*em Ana*); e (iii) o verbo causativo *dar* "absorveu" um outro verbo *dar*, que é *Vsup* na construção de base.

5.1.2 A comutação do *VopC* com outros verbos causativos

Nos casos em que *dar* é um verbo-operador causativo, ele pode comutar com outros verbos também causativos, já que todos eles possuem propriedades sintáticas e semânticas semelhantes. Levin (1993) listou um conjunto de verbos causativos para o Inglês; Cançado, Godoy e Amaral (2012) fizeram o mesmo para os verbos em Português Brasileiro; e Baptista (2012) classificou 67 verbos do PE como verbos causativos, na classe 01T, no ViPER.

As listas de verbos causativos de Cançado, Godoy e Amaral (2012) e de Baptista (2012) foram intersectadas a fim de obtermos o maior número possível de verbos com comportamento sintático e semântico semelhante ao do *VopC dar*. São exemplos de verbos-operadores causativos: *causar*, *provocar*, *fazer com que*, *implicar*, *ocasionar*, *originar*, *permitir*, *possibilitar*, dentre outros.

Em seguida, a lista de *Npred* que autoriza a construção causativa com *dar* foi intersectada com a lista dos verbos causativos, utilizando grafos do Unitex, a fim de identificar em *corpus* as combinações desses *Npred* com qualquer dos verbos alistados. Os poucos *Npred* que não foram encontrados no *corpus*, associados a algum dos verbos disponíveis, foram consultados individualmente na *web* para garantir a sistematicidade do experimento. A lista dos nomes predicativos que autorizam o *VopC dar* é apresentada no Apêndice B.

5.2 O estatuto das construções causativas

Há divergências teóricas quanto ao estatuto do verbo causativo, pois pode ser classificado como um tipo particular de verbo-suporte ou como uma categoria distinta da categoria dos verbos-suporte, mas que mantêm uma fronteira tênue com estes últimos.

Ruppenhofer *et al.* (2006), ao proporem uma classificação semântica para os verbos-suporte no âmbito da FrameNet, reconhecem vários tipos de predicados-suporte, dentre os quais citam-se: *plain vanilla*, *aspectual*, *ponto de vista*, *registro* e *causativo*. Com relação aos predicados causais, os autores afirmam que:

> O suporte adiciona outro participante e a ideia de *causa* à cena básica. Isso geralmente ocorre emparelhado com um verbo-suporte não-causativo, como em *colocar em um (mau) humor* vs. *entrar em um (mau) humor*; *trazer para o jogo* vs. *entrar no jogo*; *dar dor de cabeça* vs. *ter dor de cabeça*, *etc*. Observe que normalmente nós só marcamos a causa, por exemplo, o objeto de *dar* ou *colocar*, como um elemento de *frame* evocado pelo alvo. Além disso, nós marcamos o sujeito do verbo-suporte quando ele preenche um papel de elemento de *frame* que também é parte do *frame* de base. Este é o caso, por exemplo, do elemento de *frame* COMUNICADOR ocorrendo com o substantivo *telefonema* no *frame* de CONTATO, em frases como: *Eu vou dar-lhe um telefonema quando Mr. Steen voltar à cidade*[28]. (RUPPENHOFER *et al.*, 2006, p.38).

28 Tradução minha. Do original: "The support adds another participant and the idea of causation to the basic scene. These generally occur paired with a non-causative support, as in *put in a (foul) mood* versus *get into a (foul) mood*; *bring into play* vs. *come into play*; *give a headache* versus *have a headache*, etc. Note that normally we only tag the cause, *e.g.*, the object of *give* or *put*, as a frame element of the frame evoked by the target. Additionally, we tag the subject of the support verb when it fills a frame element role that is also part of the basic frame. This is the case, for instance, for the COMMUNICATOR frame element occurring with the noun call in the CONTACTING frame in sentences such as I'll give you a call when Mr Steen is back in town" (RUPPENHOFER *et al.*, 2006, p.38).

Ao tratar da função semântica dos verbos-suporte, Gross (1998) classifica o verbo causativo como um tipo de verbo-suporte. O primeiro argumento (GROSS, 1998, p.31) para justificar que os causativos são também verbos-suporte funciona bem para o Francês, mas não é muito natural em Português; trata-se da extração do complemento nominal, mediante a dupla análise dos complementos, propriedade esta que se aplica às construções com verbo-suporte. Essa propriedade é chamada de *construção clivada* em Português.

(5.9) *Le tennis a un avantage sur la boxe.* [Ex.R]

(5.9a) *Ce règlement donne un avantage sur la boxe au tennis.* [Ex.C]

(5.9b) *C'est sur la boxe que ce règlement donne un avantage au tennis.* [Ex.C]

(5.10) *O tênis tem uma vantagem sobre o boxe.* [Ex.C]

(5.10a) *Este regulamento dá vantagem ao tênis sobre o boxe*[29]. [Ex.C]

(5.10b) *?É sobre o boxe que este regulamento dá uma vantagem ao tênis.* [Ex.C]

(5.10c) *?É ao tênis que este regulamento dá uma vantagem sobre o boxe.* [Ex.C]

As duas formas de extração do complemento preposicional, indicadas em (5.10b) e (5.10c) não parecem muito naturais em Português.

Gross (1998, p.32) reconhece que existe um problema terminológico em relação às construções causativas, porque costuma-se considerar as sequências de [*N0 fazer*] como auxiliar causativo, mas as sequências de [*N0 colocar*], [*N0 dar*] e [*N0 tornar*] como verbos-suporte causativos, por possuírem propriedades semelhantes às construções com outros verbos-suporte.

Pode-se acrescentar ainda um outro argumento em favor da classificação dos causativos como um subtipo de verbo-suporte: a correspondência entre a construção com [*Vsup Npred*] e um verbo causativo morfologicamente associado ao lema do *Npred* também se aplica a algumas construções com verbos causativos:

(5.11) *Rui deu alegria à Ana.* [Ex.C]

(5.11a) *Rui alegrou Ana.* [Ex.C]

(5.12) *O escuro dá medo em Zé.* [Ex.C]

(5.12a) *O escuro amedronta Zé.* [Ex.C]

29 A ordem dos complementos preposicionais (ao tênis e sobre o boxe) foi alterada porque, em Português, é mais natural que o sujeito do nome predicativo, em uma construção de base (O tênis tem uma vantagem sobre o boxe), seja colocado logo após o Npred, em uma construção causativa (Este regulamento dá vantagem ao tênis sobre o boxe).

Apesar de reconhecermos a validade dos argumentos que consideram o verbo-operador causativo como um subtipo de verbo-suporte, adotamos um posicionamento diferente. Consideramos neste trabalho que os *VopC* formam uma classe diferente da classe dos *Vsup*, e não uma extensão destes últimos. O que justifica esse posicionamento é o fato de que "o *VopC* 'absorve' o *Vsup* dessas frases e, de uma forma regular, pode introduzir-lhe certas modificações estruturais. A frase resultante apresenta, então, o aspecto superficial de uma frase simples" (BAPTISTA, 2005b, p.202).

(5.13) *(O Zé + Essa notícia) deu # A Ana tem muita coragem.* [Ex.C]

(5.13a) ≡ *(O Zé + Essa notícia) deu muita coragem à Ana.* [Ex.C]

Apesar de (5.13a) se assemelhar a uma frase simples, ela pode ser desmembrada em um predicado de base (*A Ana tem coragem*) e uma operação de causa que incide sobre essa frase de base. Baptista (2005b, p.202-203) afirma que:

As frases com *Vop* não se confundem com as frases com *Vsup*, na medida em que:

(i) na frase com *Vop*, a relação entre o sujeito gramatical e o resto da construção é de natureza causal;

(ii) na frase com *Vop*, continua a observar-se a relação particular que une o sujeito ao nome predicativo na respectiva frase de base com *Vsup*, mesmo que este último se encontre reduzido; e

(iii) as frases com *Vop* não permitem a formação de *GN*, ou seja, não é possível, a partir das frases com *fazer* e *dar*, acima, obter os *GN=: *o medo do Zé à Ana*, *GN=: *a coragem do Zé à Ana*.

5.3 Tipos de frases causativas

As construções classificadas como causativas se subdividem em quatro categorias semânticas diferentes, baseadas no tipo semântico dos *Npred*: (i) a classe de sentimentos, sensações e emoções (indicada pela etiqueta <SEN>); (ii) a classe das doenças (<DOE>); (iii) a classe dos atributos ou características (<ATR>); e (iv) a classe dos *Npred* resultativos (<RES>). Essas etiquetas são atribuídas a cada um dos 616 *Npred* que ocorrem com o *VopC dar* (ver Apêndice B). Seguem-se dois exemplos representantes de cada uma dessas classes, respectivamente.

(5.14) *Elogios lhe **dão alegria**.* [Ex.C]

(5.15) *A escuridão do quarto me **deu medo**.* [Ex.C]

(5.16) *Comer sal **dá pressão alta**.* [Ex.C]

(5.17) *O peixe cru me **deu dor de barriga**.* [Ex.C]

(5.18) *O advogado **deu celeridade** ao processo.* [Ex.C]

(5.19) *O editor **deu uma cara nova*** à revista. [Ex.C]

(5.20) *A poluição do ar **deu um aumento** no buraco da camada de ozônio.* [Ex.C]

(5.21) *A política internacional **deu uma diminuição** na taxa de juros.* [Ex.C]

Há ainda um tipo peculiar de construção com verbo-operador, que se diferencia desses tipos apresentados anteriormente por ser mais natural a interpretação do sujeito como agentivo do que como causativo. Trata-se de construções como:

(5.22) *(O pai + *O fato de que o pai se casara novamente) deu uma nova madrasta para o seu filho.* [Ex.C]

(5.23) *(Deus + *O fato de que Deus nos ama) deu seu filho unigênito para morrer na cruz por amor a nós.* [Ex.C]

Como se nota pelos exemplos, a interpretação de sujeito agentivo é mais natural do que como sujeito causativo. O verbo-operador *dar*, nesses casos, não pode ser substituído tipicamente por *causar*, como nos exemplos de (5.14) a (5.21), mas pode comutar com outros verbos causativos: em (5.22) pode comutar com *proporcionar* ou *propiciar*, e em (5.23) pode comutar com *fazer com que*. Além disso, é possível desmembrar essas frases em outras construções de base:

(5.22a) *O pai fez com que # O filho tem uma nova madrasta.*

(5.23a) *Deus fez com que # O filho unigênito de Deus morreu na cruz por amor a nós.*

Nesse sentido, essas construções não se confundem com construções com verbo pleno ou com *CVS*.

Na linguagem religiosa, existem construções sintáticas cujo complemento do verbo *dar* é um nome abstrato que corresponde a virtudes, dádivas ou características concedidas por alguma entidade a um experienciador de forma que esse experienciador passe a ser dotado de tal virtude, dádiva ou característica.

(5.24) *Ontem fez um ano que Deus me **deu fé**, e seu amor me tem transbordado o coração.* [Ex.R]

(5.25) *Deus nos **deu a vida eterna**, e essa vida está em seu Filho.* [Ex.R]

Apesar de não serem tipicamente construções com o verbo-operador causativo, serão assim classificadas porque podem comutar com o *VopC fazer com que* e porque não são construções de base. As próximas subseções apresentam os quatro tipos de construções causativas, agrupadas de acordo com o tipo semântico do *Npred*.

5.3.1 Causa sensação ou sentimento

Cerca de 140 construções causativas envolvem o *VopC dar* e um nome predicativo que se refere a sensações, emoções ou sentimentos. As propriedades das construções causativas também se aplicam a essa classe: (i) admite-se qualquer tipo de argumento (*Nnr*) na posição sujeito, que funciona como a causa, em (5.26); (ii) o verbo *dar* pode ser substituído por *causar* ou *provocar*, em (5.27); e (iii) o elemento na posição de complemento indireto é o experienciador, que funciona como sujeito numa frase de base com o *Vsup ter*, *estar com* ou *ser de*, em (5.28).

(5.26) *(A morte do pai + O fato de o pai ter morrido + O pai ter morrido + Isso + Rui) deu muita tristeza à Ana.* [Ex.C]

(5.27) *A morte do pai (deu + causou + provocou) muita tristeza à Ana.* [Ex.C]

(5.28) *Ana (tem + está com + é de) uma grande tristeza.* [Ex.C]

Além das propriedades gerais das construções causativas, pode-se elencar outras características que são próprias dessa classe de *Npred*. A primeira delas se refere ao fato de que grande parte desses nomes predicativos são nominalizações de verbos psicológicos, tais como:

(5.29) *O resultado da prova deu (um + uma) (aborrecimento + alegria + aflição + alívio) na Ana.* [Ex.C]

≡ *O resultado da prova (aborreceu + alegrou + afligiu + aliviou) a Ana.* [Ex.C]

Essa propriedade raramente se observa com os *Npred* da classe de doenças, por exemplo, *caxumba* (**caxumbar*), *sarampo* (**sarampear*)[30].

A grande maioria dos nomes dessa classe admitem o apagamento do elemento CAUSA da posição sujeito. Isso se dá mediante transformações em duas estruturas sintáticas diferentes, como: (i) [*Deu um Npred em N0*], que se assemelha a construções impessoais, em que a causa está elíptica e o complemento indireto figura na posição pós-nominal

30 Ambas as palavras *caxumbar* e *sarampear* possuem algumas poucas ocorrências na *web*, mas elas não se referem às doenças *caxumba* e *sarampo*, respectivamente. As ocorrências da palavra *caxumbar* aparecem apenas na expressão *caxumbar formiga*, mas isso consiste em um erro de grafia de expressão *cachimbar formiga*. A palavra *sarampear* também é usada em contextos que não se referem à doença.

(5.30a); ou (ii) [*Me deu um Npred*], a partir da pronominalização do complemento indireto, que passa para a posição pré-verbal (5.30b).

(5.30) *O resultado da prova deu um alívio na Ana.* [Ex.C]

(5.30a) *Deu um alívio na Ana.* [Ex.C]

(5.30b) *Me deu um alívio.* [Ex.C]

O apagamento do elemento CAUSA ocorre nas frases de superfície (em frases do *corpus*), porém reconhece-se que esse argumento está presente numa estrutura subjacente à frase de superfície; ele está contido na frase de base. Esse argumento, que aparece tipicamente na posição sujeito, pode mudar de posição ou de função na frase.

(5.31) *Saber o resultado da prova me deu uma grande euforia.* [Ex.C]

(Oração subordinada subjetiva substantiva preposta ao verbo)

(5.31a) *Me deu uma grande euforia saber o resultado da prova.* [Ex.C]

(Oração subordinada subjetiva substantiva posposta ao verbo)

(5.31b) *Me deu uma grande euforia (ao + por + em) saber o resultado da prova.* [Ex.C]

(Oração subordinada adverbial causal)

Em (5.31) e (5.31a), o sujeito é o mesmo (*Saber o resultado da prova*), a diferença é que, em (5.31), o sujeito é preposto, enquanto, em (5.31a), o sujeito é posposto ao predicado. Já a estrutura (5.31b) apresenta a mesma informação da CAUSA, mas não como sujeito, e sim como uma oração causal, introduzida por alguma das preposições (*ao* + *por* + *em*).

5.3.2 Causa doença

Pertencem à classe dos nomes predicativos de doenças que admitem o *VopC dar* cerca de 100 substantivos. Esse rol é apenas exemplificativo da classe, pois há um número muito mais extenso de nomes de doenças em Português. Não nos interessa, neste trabalho, elencar todos os nomes de doenças existentes na língua, mas apenas listar um número considerável desses nomes a fim de identificar propriedades gerais à classe.

Nesta classe de *Npred* inserem-se não apenas os nomes específicos das doenças (como *aids*, *câncer*, *pneumonia*, *tuberculose*, *gripe* e outras), mas também os *Npred* que se referem aos efeitos que essas doenças podem provocar no organismo (como *vômito*, *febre*, *enxaqueca*, *infarto*, *diarreia* e outros).

Os nomes pertencentes a essa classe admitem as duas estruturas sintáticas referidas na seção anterior quanto ao apagamento do argumento CAUSA. Na verdade, as construções causativas com nomes de doenças são ainda mais comuns com o apagamento da causa do que as construções em que a causa é explícita. Quando o elemento CAUSA aparece na frase de superfície, impõe restrições ao tempo e modo do verbo *dar*. É muito mais comum encontrar frases no presente do indicativo (*e.g. Transar sem camisinha dá aids*) ou com o modal *poder* (*e.g. Não é só bebida que pode dar cirrose*) do que verbos no pretérito, como ocorria com os *Npred* de sensação, emoção ou sentimento.

As frases causativas cujo *Npred* é nome de doença são mais naturais no presente porque representam afirmações genéricas, que podem ser vistas como verdades universais, e não fatos pontuais.

Além das duas transformações atestadas na seção anterior, os *Npred* da classe das doenças admitem uma terceira estrutura sintática: [*N0 deu Npred*], como em:

(5.32) Água parada ***dá dengue***. [Ex.R]

(5.32a) *?A água parada* ***deu dengue*** *no Rui.* [Ex.C]

(5.32b) ***Deu dengue*** *no Rui.* [Ex.C]

(5.32c) *O Rui* ***deu dengue***. [Ex.C]

Na transformação de (5.32b) para (5.32c), o complemento indireto passa para a posição de sujeito, mantendo-se com o mesmo papel semântico de EXPERIENCIADOR e mantendo equivalência sintática e semântica com outras frases de base cujo *Vsup* elementar é o verbo *ter*.

(5.32d) *O Rui* ***teve dengue***. [Ex.C]

Por serem equivalentes as construções *O Rui* ***deu dengue*** e *O Rui* ***teve dengue***, ambas são consideradas construções com verbo-suporte e, portanto, serão abordadas na Parte II.

Essa transformação é exclusiva da classe das doenças, mas não é autorizada por todos os *Npred* dessa classe. Em Português Brasileiro, cerca de 35 nomes predicativos (tais como *alergia*, *anemia*, *catapora*, *convulsão*, *depressão*, *espasmos*, *febre*, *tremedeira*, *vômito etc.*), além de admitirem a construção causativa com o *VopC dar*, também admitem os dois verbos *ter* e *dar* como *Vsup* elementares em construção com verbo-suporte. Os exemplos seguintes foram retirados de *corpus* a fim de comprovar sua existência:

[*VopC*]: *Sua lição de casa*: *água parada* ***dá dengue***. (= *causar*) [Ex.R]

[*Vsup=:dar*]: *Inclusive a vizinha também* ***deu dengue***. (= *ter*) [Ex.R]

[*Vsup=:ter*]: *O meu esposo já* ***teve dengue***. [Ex.R]

Rassi *et al.* (2014) descrevem esses casos, bem como as dificuldades na tarefa de integrar esses dados em um analisador sintático automático. As estratégias para superar essas dificuldades, bem como a análise desses casos, serão explicitadas na Parte III.

Outras três características podem ser observadas nessa classe de *Npred*: (i) os nomes de doenças são, em geral, nomes autônomos, ou seja, não são morfologicamente associados a verbos ou adjetivos; (ii) admitem com facilidade a presença de nomes aspectuais, tais como, *ataque* ou *crise* (*ataque do coração*, *ataque de asma*, *ataque de tosse*, *crise de labirintite*, *crise de enxaqueca*, *crise de rinite etc.*); (iii) muitos *Npred* de doenças, em frases elementares com o *Vsup ter*, não aceitam determinante (*Ana teve aids* vs. *?*Ana teve uma aids*)[31].

Um trabalho mais profundo acerca dos nomes de doenças deve ser feito no âmbito da Terminologia, já que grande parte desses nomes existe em textos de linguagem especializada da área médica. Os 100 *Npred* aqui considerados – tais como *câncer*, *aids*, *caxumba*, *catapora*, *rubéola*, *etc.* – pertencem à linguagem comum, pois foram retirados de textos jornalísticos.

5.3.3 Atribui propriedade ou característica

A grande diferença entre essa classe de *Npred* e as demais reside na própria semântica dos *Npred*, que, em geral, indicam um atributo, uma propriedade, uma característica ao complemento preposicional. Nas construções causativas desta classe, o verbo *dar* pode comutar com *atribuir*, *conferir*, *imprimir*, mas geralmente não comuta com *causar* ou *provocar*, como se observa em:

(5.33) *O texto constitucional* ***deu legalidade*** *ao processo.* [Ex.C]

(5.33a) *O texto (****atribuiu*** *+* ***conferiu*** *+* ***imprimiu****)* ***legalidade*** *ao processo.* [Ex.C]

(5.33b) *?*O texto constitucional (****causou*** *+* ***provocou****)* ***legalidade*** *ao processo.* [Ex.C]

As três estruturas sintáticas descritas nas seções anteriores ([*Deu Npred em N0*], [*Me deu Npred*] e [*N0 deu Npred*]) não se verificam para essa classe de *Npred*. Em contrapartida,

31 O determinante pode ser admitido se: (i) o nome *aids* for acompanhado de modificador (*e.g. Ele teve uma aids ebolana*) ou (ii) for usado com sentido não literal, numa expressão cristalizada (*e.g. Teve uma aids aqui no servidor, mas a princípio voltou ao normal*).

os nomes de atributos, propriedades e características apresentam particularidades diferentes das demais classes:

1. As construções com *Npred* de atributos aceitam apenas a preposição ***a***, enquanto as construções com *Npred* de sentimentos, sensações ou emoções aceitam ambas as preposições ***a*** e ***em***[32]:

(5.34) *Rui deu (clareza + coerência + informatividade) (ao + *no) texto.* [Ex.C]

(5.35) *Rui deu (medo + alegria + raiva) (à + na) Ana.* [Ex.C]

2. É possível identificar certas regularidades nos sufixos com que os *Npred* se constroem, em decorrência dos processos derivacionais que lhes dão origem. Por exemplo, os nomes derivados de adjetivos, com os sufixos *-dade*, *-ância* e *-ência*, exprimem, eles próprios, qualidades. Já os nomes com sufixo *-mento* exprimem propriedades típicas de ações ou processos. Segue-se o agrupamento de alguns nomes predicativos cujos sufixos são mais comuns:

- *-dade*: *espiritualidade, espontaneidade, exclusividade, lealdade, legalidade, expressividade, flexibilidade, fragilidade, graciosidade, homogeneidade, hospitalidade, igualdade, imobilidade, inferioridade*;
- -ção: *aceleração, afeição, agitação, amplidão, animação, coloração, conotação, decoração, determinação, emancipação, entonação, evolução, exaltação, iluminação, movimentação, precisão, projeção, sofisticação*;
- -ância/-ência: *abrangência, aparência, autossuficiência, cadência, coerência, consciência, consistência, eficiência, elegância, extravagância, exuberância, fluência*;
- *-mento*: *arejamento, caimento, encantamento, distanciamento, fortalecimento, nascimento, movimento, rejuvenescimento, seguimento, sentimento*;
- *-tude*: *amplitude, concretude, magnitude*;
- *-ura*: *brancura, desenvoltura, espessura, estrutura, formosura, moldura, postura, textura, tessitura, ternura*;
- *-ismo*: *brilhantismo, dinamismo, favoritismo, ineditismo, otimismo, simbolismo*;
- *-ez/-eza*: *avidez, beleza, clareza, destreza, esperteza, firmeza, flacidez, fluidez, fortaleza, leveza, ligeireza, liquidez, lucidez, nitidez, nobreza, rapidez*;
- *-or*: *amargor, esplendor, rigor, sabor, teor*;

32 Vale ressaltar que a preposição típica das construções com verbo-suporte em PB é a preposição ***em***, que, por vezes, pode alternar com a preposição ***a***. Já em PE, a preposição típica dessas construções é ***a***, podendo, por vezes, alternar ou ser substituída pela preposição ***em***.

Para um estudo aprofundado sobre as propriedades específicas dos *Npred* de cada grupo de sufixos, ver Rassi e Vale (2013).

3. O complemento indireto das construções com *Npred* de propriedades é, principalmente, nome não-humano (*N-hum*), ao passo que os complementos de *Npred* de sentimentos e doenças é tipicamente nome humano (*Nhum*), como demonstram os exemplos seguintes:

(5.36) *Rui deu (inteligibilidade + coerência + legitimidade) ao (processo + *Zé).* [Ex.C]

(5.37) *Rui deu (medo + alegria + raiva) ao (*processo + Zé).* [Ex.C]

Essas três propriedades nos permitem afirmar que o grupo de nomes predicativos que atribuem uma propriedade ou característica a um objeto é relativamente homogêneo, além de ser bastante produtivo. O Apêndice B mostra cerca de 360 *Npred* dessa classe que entram em construções causativas com o *VopC dar*.

5.3.4 Verbo causativo-resultativo

Na classificação proposta por Girju (2003), as construções causativas podem ser categorizadas em três grupos: (i) simples, que se referem aos verbos que mantêm uma relação causal entre o sujeito e o objeto, como os verbos *causar* e *provocar*; (ii) resultativas, que implicam uma relação causal e uma relação de efeito entre o sujeito e o objeto, como o verbo *reduzir*; e (iii) instrumentais, que possuem, no próprio verbo, uma parte da causa, como o verbo *envenenar*.

Na classe das construções causativo-resultativas, inserem-se também as extensões dos verbos resultativos. Se o verbo *reduzir* na construção (5.38) é um verbo causativo-resultativo, também as frases equivalentes com os *VopC fazer* (5.38a) e *dar* (5.38b) devem ser assim classificadas:

(5.38) *O Banco do Brasil* ***reduziu*** *a taxa de juros de operações de crédito.* [Ex.R]

(5.38a) *O Banco do Brasil* ***fez uma redução*** *da taxa de juros de operações de crédito.* [Ex.C]

(5.38b) *O Banco do Brasil* ***deu uma redução*** *na taxa de juros de operações...* [Ex.C]

Ao estender as propriedades das construções com verbos resultativos às construções com o *VopC dar*, encontramos um bom número de nomes predicativos que são nominalizações de verbos causativo-resultativos, tais como: *abatimento*, *acréscimo*, *adiamento*, *adiantamento*, *adicional*, *amortização*, *abono*, *aumento*, *lucro*, *prejuízo*, *rendimento* e outros.

(5.39) *Conseguimos chegar a esse valor porque implantamos a Rede Cegonha, o que nos* ***deu um acréscimo*** *de R$ 1 mil em cada vaga.* [Ex.R]

Alguns desses nomes predicativos admitem dois tipos distintos de construções: (i) a construção causativo-resultativa, por ter como sujeito um argumento CAUSA e como complemento um BENEFICIÁRIO, como descrita em (5.39); e (ii) a construção com o verbo-suporte *dar*, como uma variante do verbo-suporte *ter* na mesma frase de base.

(5.40) *A loja que comprei* ***deu um rendimento*** *anual de 5%.* [Ex.R]

≡ *A loja que comprei* ***teve um rendimento*** *anual de 5%.* [Ex.C]

(5.41) *Oktoberfest* ***deu lucro*** *de 1,062 milhão.* [Ex.R]

≡ *Oktoberfest* ***teve lucro*** *de 1,062 milhão.* [Ex.C]

Essas frases são ambíguas na medida em que podem ser interpretadas como construções causais, em que o beneficiário não está explícito, ou serem interpretadas como frases de base em que o próprio sujeito é também o beneficiário. Esses casos (5.40) e (5.41) serão explicados detalhadamente na Parte II, por serem construções com verbo-suporte.

Neste capítulo, introduzimos o conceito de verbo-operador causativo, que se distingue do verbo-operador de ligação e do verbo-suporte por apresentar propriedades particulares. As propriedades definitórias das construções causativas com o verbo *dar* são: (i) a posição sintática de sujeito, em construções causativas, deve ser preenchida por um argumento tipicamente causal; e (ii) o verbo-operador causativo *dar* pode comutar com outros verbos causativos, tais como *causar*, *resultar em*, *provocar*, *fazer com que*, dentre outros.

Neste capítulo também discutimos o estatuto dessas construções que, para alguns autores, são um subtipo de construções com verbo-suporte, e, para outros, as construções causativas possuem um estatuto exclusivo. Em seguida, classificamos as construções causativas em quatro grupos, de acordo com o tipo semântico dos nomes predicativos: (i) sensação, emoção ou sentimento; (ii) doença; (iii) característica ou propriedade; e (iv) nomes predicativos que expressam resultado de alguma ação ou processo.

Até agora foram analisadas construções com o verbo pleno e com o verbo-operador causativo. No próximo capítulo, abordaremos os casos de expressões cristalizadas e provérbios que são construídos com o verbo *dar*.

6. As frases fixas

É necessário distinguir as frases fixas das construções livres porque, de maneira geral, o significado global de uma frase fixa não corresponde aos significados de seus constituintes. As frases fixas possuem estrutura sintática semelhante à estrutura de frases livres, porém as frases fixas devem ser identificadas e analisadas como um bloco único de significado, cujas posições sintáticas devem ser sempre preenchidas por unidades lexicais específicas. Isso significa que as frases fixas (tanto provérbios quanto *ECs*) são constituídas por palavras e sintagmas tais como as frases livres, porém devem ser entendidas como uma expressão complexa cujo significado pode ser ou mais ou menos metafórico.

Consideraremos neste capítulo dois tipos distintos de frases fixas: as Expressões Cristalizadas (*EC*) e os provérbios. Nas *ECs* verbais, o verbo e pelo menos um dos seus argumentos são cristalizados juntos, ou seja, a composição é distribucionalmente invariável, ou o complemento nominal sofre pequenas variações lexicais. Geralmente, o sujeito de *ECs* é distribucionalmente livre e sua seleção depende tanto do verbo quanto do significado global da combinação entre o verbo e seus complementos fixos, por exemplo:

(6.1) *Projeto de interferir na Eurásia já* ***deu com os burros na água****.* [Ex.R]

Por outro lado, os provérbios são tipicamente sentenças fixas completas, em que todos os seus elementos são fixos, inclusive o sujeito. Conforme aponta Gross (1982, p.161), enquanto o sujeito das *ECs* geralmente varia e pode ser definido intencionalmente por restrições distribucionais, os provérbios têm a posição sujeito necessariamente preenchida por um elemento fixo.

(6.2) ***Todos os caminhos dão em Roma****.* [Ex.R]

Ressalte-se que, assim como nas *ECs*, pode haver pequenas variações lexicais ou formais entre os constituintes. No caso do provérbio (exemplo 6.2), o verbo *dar* pode comutar com os verbos *levar*, *conduzir* e *chegar*, podendo variar também a preposição locativa (*em* ou *a*).

A segunda propriedade que distingue provérbios e expressões fixas, segundo Lopes (1992, p.10), é que os provérbios "sempre têm um valor semântico autônomo em termos comunicativos, diferentemente dos idiomas[33], que são apenas constituintes de sentenças e

33 O termo *idioma* é usado pela autora para se referir ao mesmo objeto que nós nomeamos, neste trabalho, como *expressão cristalizada*.

nunca ocorrem como uma sentença completa[34]". Nesse sentido, os provérbios substituem sentenças inteiras (com predicação completa), enquanto as Expressões Cristalizadas substituem apenas sintagmas (sintagma verbal, sintagma nominal, sintagma adverbial *etc.*).

Nas duas seções seguintes apresentaremos, respectivamente, as estruturas de Expressões Cristalizadas e de provérbios constituídos com o verbo *dar*, encontradas no *corpus*.

6.1 As expressões cristalizadas

A abordagem das expressões cristalizadas (*EC*) – ou expressões fixas (*EF*) – representa um problema sério para a Linguística teórica porque remete ao conceito, ainda inexato, de *palavra*, ou *unidade lexical*. Além disso, remonta à questão da delimitação entre sintaxe e léxico, no que concerne à definição de itens lexicais com constituição maior do que um vocábulo. Isso quer dizer que o conceito de léxico deve ser tomado de forma muito mais inclusiva do que como um único vocábulo. Essa é a proposta do Léxico-Gramática.

As *ECs* – também chamadas de *expressões idiomáticas*, *idiomatismos*, *idiotismos*, *colocações*, *frasemas*, *idiomas* e outras – merecem destaque nos estudos linguísticos porque representam um problema significativo para o processamento automático de línguas naturais, pois uma vez que o significado global de uma *EC* não pode ser dado pelo significado de suas partes, um sistema computacional precisa reconhecer a *EC* como um vocábulo único, e não como a combinação de vários vocábulos. Quando uma expressão possui significado global correspondente ao significado de suas partes, dizemos que possui significado composicional.

O critério da composicionalidade, no entanto, não pode ser o único aspecto a ser observado para caracterizar determinada expressão como cristalizada, pois existem diferentes graus de composicionalidade e isso não pode ser medido categoricamente. Observe-se, por exemplo, as expressões *dar o melhor de si* e *dar tudo de si*. Apesar de não serem completamente transparentes e literais, elas possuem um alto grau de composicionalidade. O problema é que também é comum encontrar em *corpus* variações como:

34 Tradução minha. Do original: "[...] always have an autonomous semantic value in communicative terms, unlike idioms that are only constituents of sentences and may never occur as a full sentence" (LOPES, 1992, p.10).

(6.3) *sou Proativa todas as minhas experiências de trabalho nesse longo tempo de 26 anos de carteira assinada sempre* ***dei o melhor de si****.* [Ex.R]

A falta de concordância entre o sujeito de *dei* e a pessoa designada pelo pronome *si* atesta um certo grau de fixidez da expressão *dar o melhor de si*, ainda que seja no limite da composicionalidade.

Existem vários tipos de *ECs*: verbais, substantivais, adjetivais e adverbiais. A diferença principal entre os quatro tipos de *ECs* consiste basicamente no fato de as *ECs* verbais poderem ser substituídas por um verbo (ou locução verbal), ao passo que as *ECs* adverbiais podem ser substituídas por um advérbio (*Adv* ou locução adverbial), e assim por diante. Nesta pesquisa serão analisadas apenas as *ECs* verbais, ou seja, expressões que apresentam sempre um verbo como componente fixo obrigatório. São exemplos de *ECs* verbais:

(6.4) *Desprestigiado, o aloprado ameaça* ***dar com a língua nos dentes****.* [Ex.R]

(6.5) *Você entende que* ***dar a mão à palmatória*** *não te diminui, te evolui.* [Ex.R]

(6.6) *Cansei de* ***dar murro em ponta de faca*** *e vou entrar na dos africanos.* [Ex.R]

As *ECs* verbais se distinguem dos outros tipos por se referirem a ações ou processos e, nesse sentido, substituem a posição sintática do verbo. No *corpus* foram encontradas também *ECs* adjetivais e adverbiais cuja estrutura exige a presença do verbo *dar* e de outros elementos fixos ou relativamente fixos. Como exemplos de *EC* adverbial e *EC* adjetival, construídas com o verbo *dar*, podemos citar (6.7) e (6.8), respectivamente:

(6.7) *Uma festa de arromba, com Zé Pedro de DJ e sorvete* ***a dar com pau****.* [Ex.R]

(6.8) *Enquanto isso, as estações da periferia ficam* ***ao deus-dará****.* [Ex.R]

A expressão destacada em (6.7) é considerada um advérbio de intensidade, podendo ser substituída pelo advérbio *muito* preposto ao nome *sorvete* (*Uma festa de arromba, com Zé Pedro de DJ e* ***muito*** *sorvete*); já a segunda expressão, em (6.8), é classificada como adjetivo que poderia ser substituído pelos adjetivos *abandonadas* ou *desassistidas* (*Enquanto isso, as estações da periferia ficam* ***desassistidas***).

O conceito de expressão cristalizada está longe de ser consensual na literatura, porém muitos teóricos consideram que existem características das *ECs* que devem ser satisfeitas para que sejam consideradas *ECs*. Tendo analisado várias propostas anteriores (GROSS, 1988; GROSS, 1996; VALE, 2001; BAPTISTA; CORREIA; FERNANDES, 2004; LAMIROY, 2008), consideraremos como expressão cristalizada toda sequência de palavras que satisfaça os três critérios seguintes:

i) deve apresentar, no mínimo, dois elementos fixos. No caso das *ECs* verbais, um deles deve ser o verbo e o outro pode ser pertencente a qualquer outra categoria gramatical;

ii) seu significado deve ser não-composicional;

iii) não permite a inserção de outros elementos entre os seus constituintes.

Esses critérios servem, em larga medida, para distinguir entre *ECs* e construções livres. Distinguir entre *ECs* e *CVS*, no entanto, é mais complexo. Em trabalho anterior (RASSI, 2015), adotamos os seguintes critérios para distinguir as *ECs* das *CVS*: (i) a existência de complementos fixos; (ii) a fixidez do determinante; e (iii) o sentido (literal ou figurado) dos complementos. A Figura 4 apresenta a chave dicotômica para a distinção entre essas categorias:

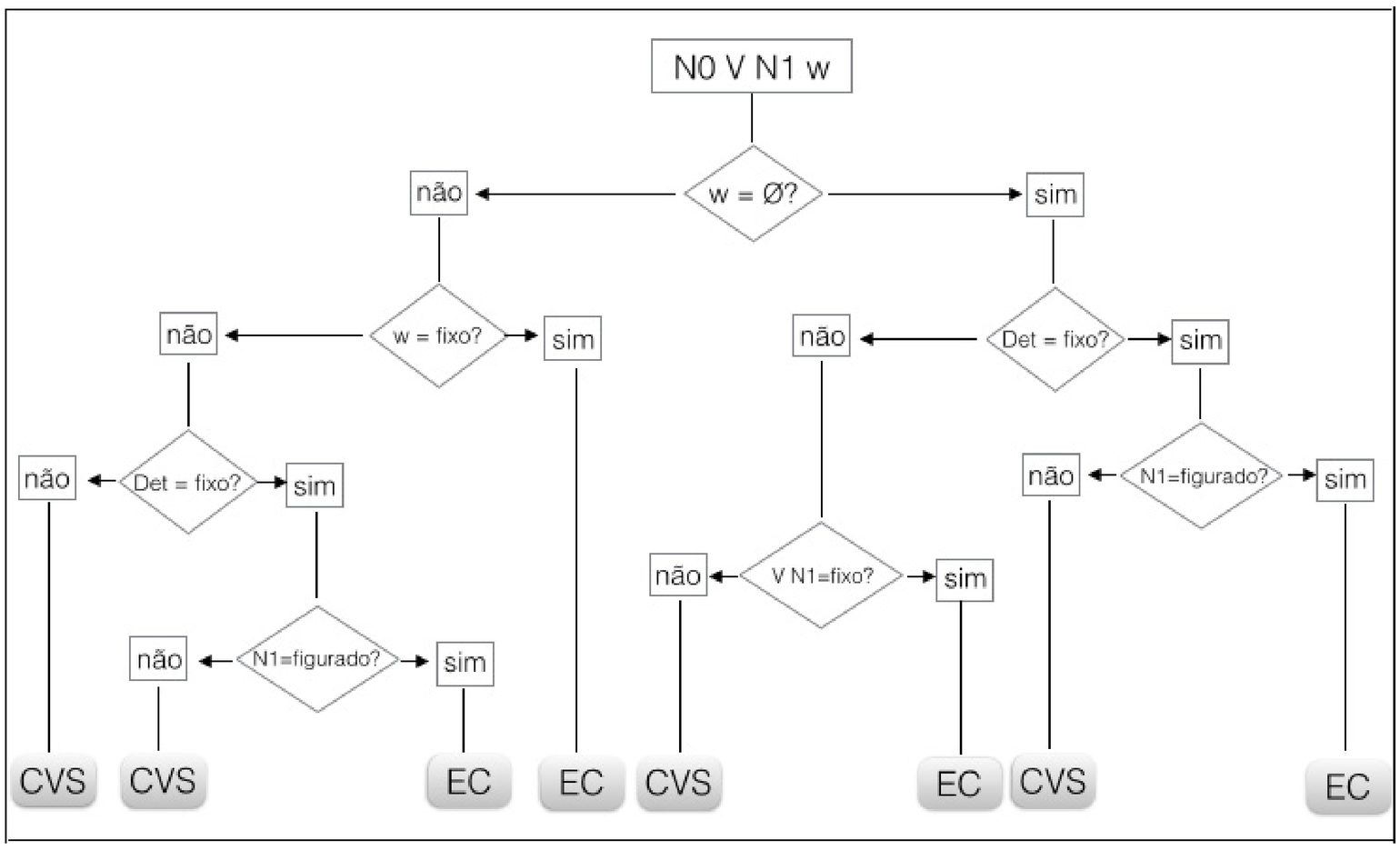

Figura 4 – Chave dicotômica para a distinção entre *EC* e *CVS*
Fonte: Elaborado pela autora e já publicado em Rassi, Baptista e Vale (2015, p.226) Essa figura deve ser interpretada da seguinte forma:

1. Se a construção possui outro(s) complemento(s):

   [1.1] e, se algum dos complementos é fixo, então trata-se de uma *EC*; [1.2] mas, se o complemento não é fixo, então: [1.2.1] se o determinante é livre, é uma *CVS*; [1.2.2] mas, se o determinante é fixo:

   [1.2.2.1] e N1 tem sentido literal, então é *CVS*; [1.2.2.2] e N1 tem sentido figurado, então é *EC*;

2. Se a construção não possui outro(s) complemento(s):

   [2.1] e, se o determinante é livre:

      [2.1.1] e os dois constituintes (verbo e nome) são fixos, então é *EC*;

      [2.1.2] mas os dois constituintes (verbo e nome) são livres, então é uma *CVS*;
      [2.2] mas, se o determinante é fixo:

      [2.2.1] e, se o nome tem uma interpretação literal, então trata-se de uma *CVS*;
      [2.2.2] e, se o nome tem uma interpretação figurada, então é uma *EC*.

Para testar a funcionalidade da chave dicotômica, apresentamos dois casos aparentemente semelhantes, mas que são diferentes quanto à sua classificação: (*fazer, cinema*) e (*fazer, filme*). Esses casos são ainda interessantes por salientarem as diferenças entre as variantes Brasileira e Europeia. No Português Europeu, além da interpretação literal significando "ser ator ou realizador de cinema", a expressão *fazer um filme* também tem um valor figurado quando empregada no sentido de "pôr-se a imaginar o que poderia acontecer", geralmente com uma conotação negativa (ou disfórica), sendo considerada, nessa situação, uma *EC*. Em Português Brasileiro, *fazer filme* permite apenas a interpretação literal, pelo que corresponde a uma *CVS*. Pode-se comparar a *EC fazer um filme*, em PE, com as expressões *fazer teatro*, *fazer drama* e *fazer cena*, em PB, que, em certas situações, significam "agir com exagero".

A construção *fazer filme*, na perspectiva tanto do cineasta quanto do ator, é considerada uma *CVS*, pois, na ausência de outros complementos, o *Npred filme* admite determinante livre. Já a expressão *fazer cinema* também não possui outros complementos, além do verbo e do nome, então, seguindo a orientação da chave dicotômica, verifica-se se o determinante é livre ou fixo:

(6.9) *Eva fez (um + muitos + o + seu + Ø) filme(s).* [Ex.C]

(6.10) *Eva fez (*um + *muitos + *o + *seu + Ø) cinema(s).*[35] [Ex.C]

O exemplo (6.9) admite livremente qualquer determinante, ao passo que (6.10) exige a ausência de determinante, ou seja, tem determinante fixo. Além da fixidez do determinante, o nome *cinema* é usado metonimicamente no lugar de "filme exibido no cinema", o que nos permite afirmar que *fazer cinema* é uma *EC*, ao passo que *fazer filme* é uma *CVS*, salvo as exceções explicitadas anteriormente.

35 Alguns desses determinantes podem introduzir a palavra cinema, mas com interpretações diversas de "atuar como atriz ou ator em cinema". Essas construções podem indicar literalmente "construir um cinema" ou serem expressões idiomáticas semelhantes a "fazer drama", ou ainda os determinantes definido e indefinido podem ser aceitos, se acompanhados de modificador.

Seguindo a proposta do Léxico-Gramática, utilizaremos as mesmas designações das classes sintáticas já identificadas para Expressões Cristalizadas em outras línguas, incluindo a classificação das *ECs* do Português Europeu (BAPTISTA; CORREIA; FERNANDES, 2004) e do Português Brasileiro (VALE, 2001). Naqueles trabalhos, foram identificadas várias classes sintáticas, mas, nesta pesquisa, apresentaremos apenas aquelas em que se enquadram as *ECs* com *dar*.

Ressalte-se que não existe ainda a descrição e/ou formalização das *ECs* verbais com o verbo *dar* em Português Brasileiro. O trabalho de Vale (2001), que recenseou cerca de 3.500 *ECs* verbais em PB, não abrange *ECs* com os verbos *ser, estar, ficar, fazer, ter* e *dar*. O autor reconhece a importância da descrição dessas *ECs*, mas não as incluiu naquele trabalho devido à grande produtividade das construções com esses verbos (VALE, 2001, p.28-29). Já o trabalho de Baptista, Correia e Fernandes (2004), que também recensearam cerca de 3.500 *ECs* para o PE, apresenta 141 *ECs* verbais formadas com o verbo *dar*. Destas, apenas 34 expressões também existem na variante brasileira.

A Tabela 2 apresenta as nove classes em que se subdividem as *ECs* com o verbo *dar*, bem como sua estrutura sintática característica, um exemplo representativo da classe e o número de *ECs* pertencentes a cada classe. As oito primeiras classes foram retiradas das classificações de Baptista, Correia e Fernandes (2004) e Vale (2001). A nona classe foi criada para abarcar uma estrutura sintática que não constava naqueles trabalhos.

Classe	**Estrutura**	**Exemplo**	**nº**
C1	*N0 V C1*	*Rui deu o último suspiro*	36
CDN	*N0 V (C de N)1*	*Rui deu fé do assalto*	7
CP1	*N0 V Prep C1*	*Rui deu de ombros*	17
C1PN	*N0 V C1 Prep N2*	*Rui deu as costas para Ana*	67
CNP2	*N0 V N1 Prep C2*	*Ana deu à luz um bebê*	1
C1P2	*N0 V C1 Prep C2*	*Rui deu nome aos bois*	24
CPP	*N0 V Prep C1 Prep C2*	*Ana deu com os burros n'água*	15
CPPN	*N0 V C1 Prep C2 Prep C3*	*Ana deu parte do Rui à polícia*	2
N0=Ø	*V Det C1 Prep N2*	*Deu a louca em Ana*	4
TOTAL			173

Tabela 2 – Classificação das *ECs* com *dar*
Fonte: Elaboração própria

No total, foram encontradas 173 *ECs* nos *corpora* pesquisados. As propriedades sintáticas das nove classes são descritas a seguir:

- C1: é formada por *ECs* que possuem apenas um complemento fixo, não preposicionado (*N*0 *V C*1);
- CDN: é formada por *ECs* que possuem dois complementos, sendo que o primeiro é fixo e o segundo é um complemento livre, introduzido pela preposição *de*, e são representadas sintaticamente por (*N*0 *V* (*C de N*)1);
- CP1: é formada por *ECs* com apenas um complemento fixo, o qual é introduzido por preposição (*N*0 *V Prep C*1);
- C1PN: é formada por *ECs* com dois complementos, sendo que o primeiro é fixo, não preposicionado, e o segundo é livre e introduzido por preposição (*N*0 *V C*1 *Prep N*2);
- CNP2: é formada por *ECs* constituídas de dois complementos, sendo que o primeiro deles é livre e não-preposicionado, enquanto o segundo complemento é fixo e preposicionado (*N*0 *V N*1 *Prep C*2). A única *EC* representante dessa classe é *Ana deu à luz um bebê*;
- C1P2: é formada por *ECs* com dois complementos fixos, sendo que o primeiro é um complemento direto e o segundo é preposicionado (*N*0 *V C*1 *Prep C*2);
- CPP: é formada por *ECs* constituídas de dois complementos fixos, ambos preposicionados (*N*0 *V Prep C*1 *Prep C*2);
- CPPN: é formada por *ECs* com três complementos, sendo um não-preposicionado e os outros dois preposicionados (*N*0 *V C*1 *Prep C*2 *Prep C*3).
- N0=Ø : é formada por *ECs* destituídas de sujeito, ou seja, a expressão começa com um verbo e possui um complemento fixo, não-preposicionado, e um complemento livre preposicionado (*V Det C*1 *Prep N*2).

O Apêndice C deste livro contêm as expressões cristalizadas verbais formadas com o verbo *dar* em Português do Brasil. Essas informações estão dispostas em uma matriz do Léxico-Gramática.

6.1.1 Classe C1

A classe C1 é composta por 36 *ECs*, formadas pelo verbo *dar* e um complemento direto fixo, selecionando sempre um *Nhum* para preencher a posição de sujeito. Essas *ECs* não possuem outro(s) complemento(s) obrigatórios, nem fixos nem livres, mas,

em geral, aparecem no *corpus* acompanhadas de outros sintagmas ou orações que exprimem o contexto.

(6.11) *Mariah Carey* ***deu bandeira*** *em concerto no Japão.* [Ex.R]

(6.12) *A presidente Dilma* ***deu as regras do jogo*** *da reforma que empreende neste início de segundo mandato.* [Ex.R]

Essas *ECs* poderiam ser usadas, em construções de base, como *Mariah Carey deu bandeira* e *Dilma deu as regras do jogo* sem a necessidade de outros complementos essenciais.

As *ECs* desta classe podem ter um nome simples preenchendo a posição de complemento direto fixo, como em (6.11), ou um nome composto na mesma posição de *C*1, tal como em (6.12) ou nos seguintes exemplos:

(6.13) *O primeiro presidente sul-africano negro, Nelson Mandela,* ***deu o seu último suspiro*** *e descansou.* [Ex.R]

(6.14) *Quatro anos depois, ele* ***deu a volta por cima*** *e ergueu a taça nos Estados Unidos.* [Ex.R]

(6.15) *Cingapura* ***deu o pulo do gato*** *e prova que tamanho não é documento.* [Ex.R]

Os complementos *último suspiro*, *volta por cima* e *pulo do gato*, apesar de serem compostos por mais de uma palavra, assumem a posição de *C*1 na estrutura [*N*0 *V C*1], da mesma forma como ocorre nas palavras compostas.

O que caracteriza as *ECs* desta classe e as diferencia de algumas *CVS* com interpretação metafórica é a fixidez do determinante, combinado com a interpretação figurada de *C*1. Observe-se, por exemplo, as diferenças sintáticas e semânticas entre (6.16), (6.17) e (6.18):

(6.16) *Reynaldo Gianecchini confirma que* ***deu um fora em*** *Carla Bruni.* [Ex.R]

(6.17) *No programa desde domingo (30), Tatiele Polyana, até então a última eliminada do BBB 14,* ***deu um fora*** *e levou uma cutucada de Faustão.* [Ex.R]

(6.18) *Val Kilmer* ***deu o fora da*** *produção (Sorte dele), dando lugar para George Clooney.* [Ex.R]

Os exemplos (6.16) e (6.17) significam, respectivamente, "rejeitar determinada pessoa" e "passar vergonha". Nos dois casos, essas construções admitem livremente qualquer tipo de determinante (artigo definido, artigo indefinido, acompanhado ou não

de modificador, pronome demonstrativo, ausência de determinante *etc.*). Por aceitarem determinante livre, são classificados como *CVS*. Já o exemplo (6.18), que significa "sair de determinado lugar, ir embora", exige como determinante o artigo definido *o*. Como o determinante é fixo e a expressão tem sentido metafórico, a construção é classificada como *EC*.

Há 10 *ECs* da classe C1 com comportamento sintático-semântico peculiar por dois motivos: (i) o complemento que ocupa a posição de *C*1 é um advérbio ou nome com o significado metafórico funcionando como advérbio (em geral, os advérbios *certo* e *errado*); e (ii) a posição sujeito dessas *ECs* pode ser ocupada por uma oração completiva ou não ser preenchida. Seguem-se alguns exemplos:

(6.19) *Era pra ser um Yakisoba, mas* ***deu zica****.* [Ex.R]

(6.20) *Tentei anexar um arquivo mas* ***deu bode****.* [Ex.R]

Ressalte-se que, mesmo que a posição sujeito não esteja preenchida por nenhum elemento, esse argumento está presente em frases de base.

A única exceção à restrição ao preenchimento lexical do sujeito ocorre com a expressão *dar pau*, que exige um *N-hum* na posição sujeito, além de admitir propriedades transformacionais específicas:

(6.21) *Meu aparelho mais uma vez* ***deu pau*** *e agora não liga mais.* [Ex.R]

(6.21a) ***Deu pau no*** *meu aparelho e agora não liga mais.* [Ex.C]

Outros exemplos dessa subclasse da C1 são: *dar certo*, *dar errado*, *dar bode*, *dar zebra*, *dar galho*, *dar chabu*, *dar zica*, *dar pau*, *dar samba* e *dar pizza*.

Por fim, há ainda duas outras expressões desta classe com uma característica particular: o verbo *dar* pode assumir a forma pronominal *dar-se*. A posição de complemento fixo (*C*1) pode ser preenchida pelos advérbios *bem* ou *mal*, conforme demonstram os exemplos:

(6.22) *A ESPN apostou na transmissão do Super Bowl e* ***se deu bem****.* [Ex.R]

(6.23) *Foi se exibir bancando o valentão e* ***se deu mal****.* [Ex.R]

Essas expressões se distinguem da expressão *A Ana se dá bem com Zé* porque esta última exige um complemento obrigatório (*com N*2) e admite a simetria. Essa *EC* pertence à classe C1PN e será analisada na subseção 6.1.4.

6.1.2 Classe CDN

Fazem parte desta classe 7 *ECs* com o verbo *dar*. A análise das expressões desta classe será feita em três etapas: (i) a primeira verifica expressões do tipo *dar fé de*, *dar tino de* e *dar falta de*; (ii) a segunda etapa investiga expressões como *dar pinta de*, *dar parte de* e *dar uma de*; e (iii) a terceira etapa analisa a estrutura da expressão *dar-se conta de que F*. Analisemos as três primeiras *ECs*:

(6.24) *Enquanto os índios reivindicavam áreas na Amazônia, a gente nunca* ***deu fé do*** *decreto de demarcação [em vigor].* [Ex.R]

(6.25) *Assim esteve, sem* ***dar tino do*** *tempo que passava.* [Ex.R]

(6.26) *Gerente* ***deu falta do*** *dinheiro e descobriu o furto após verificar as imagens das câmeras de segurança.* [Ex.R]

Essas expressões têm a estrutura [*N*0 *V* (*C de N*)1], em que a posição sujeito é ocupada por um *Nhum*; o complemento (*C de N*)1 possui a posição de *C* fixa (*fé*, *tino*, e *falta*, respectivamente), e a posição de *N* é livre (*decreto*, *tempo* e *dinheiro*, respectivamente). A posição de *N* também poderia ser preenchida por uma oração completiva, em todos os três casos.

(6.27) *Ao passarem por uma Igreja, Ricardo* ***deu fé de que*** *o seio palpitante de Ludovina o empurrava.* [Ex.R]

(6.28) *Quando a sentinella* ***deu tino de que*** *se aproximava alguém, cumpriu a praxe militar.* [Ex.R]

(6.29) *Ocorre que, por usar pouco o cartão,* ***não deu falta de que*** *havia perdido, só deu conta quando chegou uma fatura em sua residência.* [Ex.R]

Os significados das duas primeiras *ECs* é "perceber". Já a *EC dar falta de* indica "perceber a ausência, a falta de".

O segundo grupo de expressões da classe CDN, também representado pela estrutura [*N*0 *deu* (*C de N*)1], pode ter o complemento *C*1 instanciado por *parte*, *pinta* ou *uma*. A posição de *N*0 é instanciada por *Nhum* e a posição de *N*1 pode ser instanciada por alguns nomes ou adjetivos, sob certas restrições semânticas.

(6.30) *Vieira da Costa, juiz de direito da comarca,* ***deu parte de doente*** *para não presidir jury.* [Ex.R]

(6.31) *Não é careca, muito menos barrigudo, não* ***dá pinta de playboy*** *e não é truta de ninguém.* [Ex.R]

(6.32) *O zagueirão do PSG se enrolou e* ***deu uma de Joel Santana****.* [Ex.R]

O significado dessas três *ECs*, resguardadas pequenas diferenças semânticas, é "fingir ser, fazer parecer que é", que equivalem ao verbo '*to pretend*' do inglês.

A expressão *dar parte de* foi classificada por Baptista, Correia e Fernandes (2004) na classe C1, reconhecendo como nomes compostos os complementos *parte de doente* e *parte de fraco*. No Português Brasileiro, é possível encontrar ainda outros nomes que podem preencher a posição *N*1, e não apenas *doente* e *fraco*, tais como:

(6.33) *o Tutucha no caminho* ***deu parte de valentão****.* [Ex.R]

(6.34) *Quando Ângela* ***deu parte de pronta****, despedimo-nos e o doutor ofereceu a casa ao namorado.* [Ex.R]

Assim, essas expressões serão classificadas na classe CDN porque: (i) admitem certa liberdade ao preenchimento lexical de *N*1; e (ii) assemelham-se, tanto pela forma quanto pelo significado, a outras *ECs* que fazem parte da classe CDN, como *dar pinta de* e *dar uma de*.

A expressão *dar pinta de N* também admite certa liberdade ao preenchimento lexical da posição de *N*1, conforme demonstram alguns exemplos retirados de *corpus*:

(6.35) *Juliano Cazarré, vai* ***dar pinta de MC*** *na trama de Favela Chiqui.* [Ex.R]

(6.36) *Vestido Calvin Klein* ***pra dar pinta de Audrey Hepburn****.* [Ex.R]

(6.37) *Concordo que um homem não deve desmunhecar e* ***dar pinta de baitola****.* [Ex.R]

(6.38) *O rapaz quis* ***dar pinta de malandro****.* [Ex.R]

Com relação à expressão *dar uma de*, ressalte-se que a *EC* mais característica do grupo é *N*0 *deu uma de (João + Joãozinho) sem braço*, em que *C*1 é ocupado pela palavra *uma* e a posição de *C*2 é ocupada pelo substantivo *joão sem braço* ou *joãozinho sem braço*. Essa expressão, por conter dois complementos fixos, foi introduzida na classe C1P2.

A expressão *dar uma de (João + Joãozinho) sem braço* deu origem a várias outras expressões cuja posição de *C*2 (*joão sem braço*) deixa de ser fixa e passa a ser livre, mas com algumas restrições. A posição de *C*2 pode ser ocupada por alguns outros nomes ou adjetivos.

(6.39) *Cristovam Buarque (PT-DF)* ***dá uma de tucano****.* [Ex.R]

(6.40) *Ela fica* ***dando uma de meiga*** *pra ver se arruma casamento.* [Ex.R]

(6.41) *Finalmente alguém* ***deu uma de sábio*** *na reunião.* [Ex.R]

Nas *ECs* formadas por [*N*0 *Vdar* (*uma de N*)1], em que *uma* está preenchendo a posição do *C*1, os elementos *dar uma de* são extremamente fixos, não admitindo qualquer variação. A posição de sujeito dessa *EC* também é distribucionalmente livre, mas possui restrições semânticas: deve ser um *Nhum*, que seja capaz de "fingir" ('*to pretend*'). O complemento *N*1, apesar de sofrer fortes restrições de seleção, é considerado um complemento livre. Essa posição só pode ser preenchida por um adjetivo ou um substantivo do qual se possa extrair uma característica que funcione como qualificativa. Todas as construções com [*N*0 *Vdar* (*uma de N*)1] poderiam ser parafraseadas por outra expressão cristalizada com o verbo *fazer*, como em:

(6.39a) *Cristovam Buarque (PT-DF)* ***se faz de tucano****.* [Ex.C]

(6.40a) *Ela fica* ***se fazendo de meiga*** *pra ver se arruma casamento.* [Ex.C]

(6.41a) *Finalmente alguém* ***se fez de sábio*** *na reunião.* [Ex.C]

Também pertence a esta classe CDN a expressão *dar-se conta de que*. Essa expressão se difere das demais *ECs* da classe nos seguintes aspectos: (i) o verbo é pronominal (*dar-se*); e (ii) a posição de *N*1 é ocupada por uma frase completiva (*QueF*), como se demonstra em:

(6.42) *Num momento de lucidez, ele* ***se deu conta de que*** *havia um padrão no seu comportamento.* [Ex.R]

(6.43) *Ela* ***se deu conta de que*** *precisaria de ajuda para levá-lo para a telona.* [Ex.R]

Apesar de haver diferenças formais entre essa expressão e as outras da classe, elas apresentam estruturas sintáticas parecidas. Na falta de uma classe mais apropriada, a expressão *dar-se conta de que* foi inserida nesta classe.

6.1.3 Classe CP1

Essa classe é composta por expressões cuja estrutura sintática é *N*0 *V Prep C*1.

Fazem parte desta classe 17 *ECs*, que podem ainda ser subdivididas em três grupos: (i) 9 *ECs* formadas por alguma preposição (*de*, *em*, *para*, *por*) e um nome ou pronome na

posição de *C*1; (ii) 4 *ECs* cuja posição de *C*1 é preenchida por uma oração subordinada substantiva objetiva direta, reduzida de infinitivo, que pode ou não ser introduzida pela preposição *em*; e (iii) 4 *ECs* formadas pelo verbo *dar*, seguido da preposição *a*, seguida de um verbo no infinitivo, que sofre fortes restrições semânticas.

O primeiro grupo pode ser representado por expressões como:

(6.44) *Peão que não* ***dá no couro****... a mulher foge com o Ricardão.* [Ex.R]

(6.45) *O marido* ***deu no pé*** *e nunca mais apareceu.* [Ex.R]

(6.46) *O sommelier* ***deu de ombros*** *e mandou vir a garrafa.* [Ex.R]

Nos três exemplos, *C*1 é preenchido por um substantivo parte-do-corpo (*couro*, *pé* e *ombros*), assim como a expressão *dar nas vistas*, mas nem todas as *ECs* desse grupo possuem a mesma característica. Algumas têm o *C*1 preenchido por pronomes, como *dar por si* (pronome oblíquo) e *não dar em nada* (pronome indefinido). Outras têm o *C*1 preenchido por outros nomes comuns, como *dar pro gasto* e *dar em pizza*, ou ainda por um advérbio, como *dar pra trás* (advérbio de lugar).

O segundo grupo de *ECs* da classe CP1 é representado por [*N*0 *Vdar (o + no) que (falar + pensar)*]. O verbo *dar* pode ser seguido ou não da preposição *em*, que é seguida pelo componente fixo *o que*, que, por sua vez, é seguido de verbo no infinitivo (*Vinf*). Há restrições semânticas em relação aos verbos que podem assumir essa posição de *Vinf* dentro da completiva. As únicas formas encontradas no *corpus* ocorrem com os verbos *falar* e *pensar*:

(6.47) *Paulo inovou no processo de criação dos suínos e* ***deu no que falar****.* [Ex.R]

(6.48) *A ação já foi realizada na UFLA, UFOP e UFSJ e* ***deu o que pensar*** *nos mestrandos e doutorandos.* [Ex.R]

Esporadicamente outros verbos (inclusive o próprio *Vdar*) assumem a posição de *Vinf*, com frequência muito mais baixa e todas elas podem ser representadas por [*N*0 *Vdar (o + no) que Vinf*].

(6.49) *O trabalho voluntário* ***deu o que fazer****.* [Ex.R]

(6.50) *A discussão já* ***deu o que tinha que dar****.* [Ex.R]

(6.51) *ele estava se destacando no mercado musical nacional e* ***deu no que deu****.* [Ex.R]

Ressalte-se que a posição *N*0 nunca é preenchida por *Nhum*, aceitando apenas *N-hum* ou completiva como sujeito.

O terceiro e último grupo da classe CP1 é representado pela expressão [*Vdar a entender*], que pode apresentar variações, como [*Vdar a ver*], [*Vdar a crer*] e [*Vdar a conhecer*]. A forma mais frequente é [*Vdar a entender*].

(6.52) *Lula **dá a entender** que tem mais opções petistas do que Palocci.* [Ex.R]

(6.53) *Ele **dava a entender** sua mágoa diante da negação dos cargos.* [Ex.R]

(6.54) *Os oficiais comunistas **deram a conhecer** a nacionalidade de 15 prisioneiros.* [Ex.R]

(6.55) *Os diálogos do filme **dão a ver** um homem de grande inteligência.* [Ex.R]

(6.56) *E ele **dava a crer** que estava possuído pelo demônio.* [Ex.R]

Essas expressões exigem, além do *Vinf*, que é fixo, um outro complemento, que é livre, e geralmente é preenchido por uma oração completiva, como em (6.52) e (6.56), ou um nome predicativo, como em (6.53) e (6.54).

6.1.4 Classe C1PN

Essa classe é a mais numerosa contendo 67 *ECs* com a estrutura [*N*0 *V* (*Det*) *C*1 *Prep N*2]. Em geral, as posições de *N*0 e *N*2 são preenchidas por *Nhum*, ao passo que *C*1 é o complemento fixo não preposicionado que acompanha o verbo *dar*, também fixo.

(6.57) *A atriz foi clicada com Marlon Teixeira e discretamente **deu dedo para** o paparazzo.* [Ex.R]

(6.58) *As duas gostaram do sujeito e ele aparentemente **deu corda para** elas.* [Ex.R]

(6.59) *Lewis Hamilton **deu o troco em** seu companheiro de Mercedes.* [Ex.R]

Em casos mais raros, outros tipos semânticos podem preencher as posições de *N*0 e *N*2, como se observa nos exemplos:

(6.60) *O ribeirão não **dá pé para** Emília.* [Ex.R]

(6.61) *Você **deu as costas para** o futuro porque esqueceu do passado.* [Ex.R]

(6.62) *Nas vitrines da capital, o vermelho **deu lugar ao** branco.* [Ex.R]

Em (6.60), *N*0 é preenchido por um *N-hum*, com fortes restrições de seleção, ou seja, precisa ser preenchido necessariamente por um nome do tipo *piscina, rio, ribeirão, mar, lago, lagoa etc*. Em (6.61), *N*2 é preenchido por um *N-hum*, com fracas restrições de seleção, podendo também ser preenchido por *Nhum*. Já em (6.62), *N*0 e *N*2 precisam ser preenchidos pelo mesmo tipo semântico de argumento, por exemplo, duas cores ou dois locais ou dois nomes humanos *etc*.

Assim como na classe C1, é possível que o complemento *C*1 da classe C1PN seja um nome composto, como em (6.63), (6.64) e (6.65), ou ainda seja uma coordenação de nomes cuja ordem é fixa, como em (6.66), (6.67) e (6.68).

(6.63) *O dentista* ***deu um chá de cadeira nos*** *clientes porque seu carro estava com o pneu furado.* [Ex.R]

(6.64) *O governador Silval Barbosa* ***deu um banho de água fria em*** *grupos políticos.* [Ex.R]

(6.65) *o comandante dos Lobos do Norte* ***deu uma injeção de ânimo em*** *seus atletas.* [Ex.R]

(6.66) *Erasmo era o fodão, que Tim Maia o ensinou a tocar violão e* ***lhe deu casa comida e roupa lavada****.* [Ex.R]

(6.67) *Depois que o rinovírus foi embora,* ***deu tchau e bênção ao*** *meu corpinho.* [Ex.R]

(6.68) *Pelas macaquices, trejeitos, e meneos, em que* ***dá sota, e ás aos*** *das outras espécies.* [Ex.R]

Outros nomes compostos podem assumir a posição de *C*1 nesta classe; são eles: *carta branca, chega-junto, chega-pra-lá, colher de chá, palavra de apoio, sinal verde, banho de loja, banho de entusiasmo, voz de prisão, lição de vida, lição de moral, pé na bunda* e *cartada final*.

Nesta classe, existem três *ECs* que exigem a negação obrigatória: *não dar a mínima para, não dar pelota para, não dar tchum para*.

(6.69) *Equipe do Vasco* ***não deu a mínima para*** *a sua torcida.* [Ex.R]

(6.70) *Ainda no campo futebolístico, Dodô* ***não deu pelota para*** *a conquista da Copa das Confederações.* [Ex.R]

(6.71) *Entrei na loja todo roto, o cara* ***nem me deu tchum****.* [Ex.R]

Há outras expressões com o mesmo sentido de "não dar atenção" que também exigem a negação obrigatória, tais como [*N0 Neg Vdar (ideia + papo + assunto) para N2*]. Esses mesmos complementos (*ideia*, *papo* e *assunto*) também entram em expressões cristalizadas afirmativas (que não exigem a negação obrigatória), mas com outro sentido.

(6.72) *Ei potranca, chega aí / Vou te* ***dar uma idéia****, chega aí / Se quer dançar, se divertir / Pra lá e pra cá, pra sacudir* [Ex.R]

(6.73) *Mano Brown, representante máximo da cultura independente, do "faça você mesmo", do "não jabá", do "sem rabo preso"* ***deu papo reto****.* [Ex.R]

(6.74) *Essa foto foi tirada na famosa festa do apê, que bombou o prédio inteiro e* ***deu assunto pra mais de mês****.* [Ex.R]

Esta classe C1PN apresenta ainda um outro conjunto de *ECs* com conotação sexual, formadas pelo verbo *dar*, um determinante fixo (normalmente artigo definido) e um nome parte do corpo referente ao ânus, conforme atestam os exemplos de (6.75) a (6.77). Alerte-se que os dois parágrafos que se seguem apresentam linguagem chula, de baixo calão e pornográfica.

(6.75) *Mulher* ***deu o rabo*** *pro jovem amante.* [Ex.R]

(6.76) *Coroa gostosa* ***deu a bunda*** *pro genro.* [Ex.R]

(6.77) *Andressa Urach afirma que* ***deu o boga*** *para Cristiano Ronaldo.* [Ex.R]

Esse tipo de expressão, apesar de ser frequente em *corpus* de língua oral (e na *web*), não consta em nenhum dos dicionários pesquisados. Essas expressões de baixo calão, assim como os palavrões, muitas vezes são suprimidas de documentos formais, materiais didáticos e até de páginas da *web*, em nome da moral e dos "bons costumes" (LABOREIRO; OLIVEIRA, 2014). Como este trabalho se propõe a descrever as construções com o verbo *dar*, e esse tipo de construção existe na língua, então faremos menção a esses casos, sem maiores delongas.

Nessas *ECs*, a posição do *C1* pode ser preenchida por uma grande diversidade de nomes que representam a parte do corpo ânus, tais como: *cu*, *rodela*, *forebis*, *brioco*, *furico*, *anel*, *anel de couro*, *biscoito*, *rosca*, *rosquinha*, *botão*, *rabeta*, dentre outros. Todas essas *ECs* têm o sentido metafórico de "fazer sexo anal". Em situações mais raras, encontra-se também a expressão *dar a buceta* ou somente *dar* (sem complemento) com conotação sexual de "fazer sexo". O complemento *buceta* (ou *boceta*) admite algumas variantes lexicais que também indicam o órgão sexual feminino, tais como, *perereca*, *periquita*

etc. Já a expressão *Vai dar meia hora de cu* é proferida geralmente em tom de ordem ou pedido, no mesmo sentido que outras *ECs*, tais como, *Vai encher o saco do bode, Vai plantar batata, Não enche o saco*, todas elas com o significado de "Não me importune".

Por fim, há uma expressão na classe C1PN que apresenta uma particularidade: o verbo *dar* é pronominal (*dar-se*). Trata-se da construção [*N*0 *Vdar-se bem com N*1], em que a expressão *dar-se bem* significa "relacionar-se bem".

(6.78) *O que fazer quando você* ***não se dá bem com*** *a madrasta.* [Ex.R]

O advérbio *bem*, que preenche a posição de *C*1, pode ser apagado da expressão, mantendo-se o significado. A estrutura sintática da construção passa a ser:

[*N*0 *Vdar-se com N*1]

(6.79) *Mas o mundo o amaria menos, considerando-se que há muita gente que* ***não se dá com*** *gays por aí.* [Ex.R]

As duas estruturas sintáticas da mesma expressão admitem a propriedade da simetria (BAPTISTA, 2005b), que permite que dois argumentos de um predicado troquem de posição relativa ou sejam coordenados, sem que isso altere o significado das frases resultantes, *e.g. Ana se dá bem com Zé = Zé se dá bem com Ana = Ana e Zé se dão bem = Zé e Ana se dão bem*. A simetria pode ser observada tanto em expressões fixas, como neste exemplo, quanto em construções verbais, nominais e adjetivais.

Essas expressões são consideradas *ECs* porque a posição de *C*1 só pode ser preenchida por um único elemento, ou admitir, em raros casos, algumas variações. Assim, a expressão é composta por um verbo (*dar-se*), um complemento fixo (*bem*) e seleciona dois argumentos do tipo *Nhum*, sendo que o segundo é introduzido pela preposição *com*. A fixidez de *C*1 pode ser comprovada pelo fato de que a forma oposta dessa expressão não se dá com a alteração do advérbio *bem* para *mal*. A negação mais característica dessa construção se faz por meio da inserção de um advérbio de negação à construção *dar-se bem* (*e.g. Ana não se dá bem com Zé*); raramente a negação dessa construção se faz com a antonímia do advérbio *bem* (*e.g. ?Ana se dá mal com Zé*).

A expressão *dar-se mal*, em oposição a *dar-se bem*, na ausência de complementos, possui sentido figurado, com os respectivos significados de "cometer um ato falho, sem sucesso" e "ter êxito, ser bem-sucedido, conquistar o que se almeja". Essas *ECs*, por não apresentarem complemento, foram inseridas na classe C1 e analisadas na subseção 6.1.1.

6.1.5 Classe CNP2

Conforme apontado anteriormente, esta classe possui apenas uma *EC* com o verbo *dar*, apesar de permitir certas variações sintáticas.

(6.80) *Grávida aos 70 anos do marido de 80,* ***deu à luz*** *o menino Amritwani e a menina Barsaat.* [Ex.R]

A expressão *dar à luz N* é representada pela estrutura sintática [*N*0 *Vdar N*1 *Prep C*2], em que *Prep* é instanciado pela preposição *a* e o complemento *C*2 é instanciado pelo nome *luz*. A alternância da ordem entre *N*1 e *Prep C*2 não configura variação sintática. Há, no entanto, algumas possibilidades de variação sintática dessa *EC*, encontradas em *corpus*.

(6.80a) *Uma parteira afirma ter feito o parto de uma mulher que* ***deu luz a*** *um lagarto na Indonésia.* [Ex.R]

(6.80b) *A mãe, a operadora de caixa Aparecida da Silva Godoy,* ***deu a luz à*** *Sofia.* [Ex.R]

(6.80c) *A apresentadora do SBT, Patrícia Salvador,* ***deu a luz a*** *uma menina, Bárbara.* [Ex.R]

Essas formas da *EC dar à luz N*, apesar de muito frequentes em *corpus*, são consideradas desviantes da norma padrão, segundo as gramáticas normativas. Na verdade, trata-se de uma anástrofe que é lexicalmente determinada pela fixidez dos constituintes.

6.1.6 Classe C1P2

Nesta classe, constam 24 *ECs*, cuja estrutura básica é [*N*0 *V C*1 *Prep C*2]. Em geral, essas expressões exigem um *Nhum* na posição de sujeito e não possuem nenhum outro complemento além de *C*1 e *C*2, como se observa nos seguintes exemplos:

(6.81) *Felipão, pelo jeito,* ***deu a mão à palmatória****.* [Ex.R]

(6.82) *O próprio Batista denunciou,* ***deu nome aos bois****.* [Ex.R]

Assim como nas outras classes, a posição de *C*2 também pode ser preenchida por um nome composto, como nos exemplos:

(6.83) *Dilma não apenas colocou Aécio de joelhos diante da questão, como também* ***deu um tapa com luva de pelica*** *nos eleitores de Alckmin.* [Ex.R]

(6.84) *Todos sabemos que, hoje em dia, dar a volta do "magalhães" é mais barato do que* ***dar a volta ao bilhar grande****.* [Ex.R]

(6.85) *Não fique aí remando contra a maré,* ***dando murro em ponta de faca****.* [Ex.R]

Nesses três exemplos, o complemento *C*2 é preenchido por substantivos compostos, como *luva de pelica*, *bilhar grande* e *ponta de faca*. Em outros casos, a posição de *C*2 também pode ser preenchida por um verbo, em vez de nome, como se observa em:

(6.86) *Nem todos têm coragem de* ***dar a cara pra bater****.* [Ex.R]

(6.87) *Champagnat compreendeu esta espécie de chantagem e não* ***deu o braço a torcer****.* [Ex.R]

A segunda expressão (6.87) aparece frequentemente em *corpus* na forma negativa, acompanhada por um advérbio (*não*, *nunca*, *jamais*), porém a negação obrigatória não é considerada uma propriedade característica dessa *EC*, já que há casos em que a expressão aparece na forma afirmativa.

Apenas uma das expressões desta classe tem a propriedade da negação obrigatória:

(6.88) *Ronaldo, que* ***não dá ponto sem nó****, busca três objetivos.* [Ex.R]

A preposição que aparece entre os complementos *C*1 e *C*2 é bastante variada. No *corpus*, há registros das seguintes preposições: *a*, *com*, *de*, *em*, *para*, *por* e *sem*, como demonstram os exemplos:

(6.89) *O presidente Castanheira* ***deu a cara a tapa*** *e assumiu a responsabilidade dos jurados.* [Ex.R]

(6.90) *Se ninguém quer fazer gol, Müller* ***dá conta do recado****.* [Ex.R]

(6.91) *Eu* ***dei um tiro no pé*** *ao recusar compradores para minha casa.* [Ex.R]

(6.92) *A pressão diária dos servidores sobre o governo está* ***dando pano pra manga****.* [Ex.R]

(6.93) *Mas por conta das restrições a Alice* ***deu um tiro pela culatra****.* [Ex.R]

(6.94) ***dei a volta por cima*** *e vi nesse site que não sou a única que sofreu bullying.* [Ex.R]

Por fim, indica-se ainda uma última expressão que pertence a esta classe:

(6.95) *Agora ele vem* ***dá uma de Joãozinho sem braço*** *e falar que a matéria não tinha tempo hábil para discussão.* [Ex.R]

Nessa expressão, o elemento *uma* ocupa a posição do primeiro complemento, enquanto o elemento *joãozinho sem braço* ocupa a posição de segundo complemento. As variantes dessa *EC* que admitem um elemento livre na posição de *C*2 (*e.g. N*0 *dar uma de N*1) foram tratadas na classe CDN.

A classe C1P2 é uma das mais controladas linguisticamente, pois as *ECs* que a compõem apresentam pouca ou nenhuma variação sintática, umas em relação às outras. O estatuto dessas construções enquanto *ECs* também não gera dúvidas, já que elas possuem pelo menos dois complementos fixos, além do verbo, que também é fixo. Seguindo a proposta da chave dicotômica (ver Figura 4), se a construção apresenta mais de um complemento fixo, deve ser classificada como *EC*.

6.1.7 Classe CPP

Esta classe engloba 15 *ECs* formadas por um sujeito, geralmente humano, seguido do verbo *dar* e de dois complementos preposicionados, com a estrutura [*N*0 *Vdar Prep C*1 *Prep C*2].

(6.96) *Interferir na Eurásia já* ***deu com os burros n'água****.* [Ex.R]

(6.97) *O fato é que alguém* ***deu com a língua nos dentes****, a polícia estava de tocaia, por isso tudo deu errado.* [Ex.R]

(6.98) *Pedro Cabral* ***deu com os costados na*** *antiga Ilha de Vera Cruz.* [Ex.R]

(6.99) *Um morador passou mal e foi lá no posto e* ***deu com a cara na porta****.* [Ex.R]

Essas *ECs* possuem a mesma estrutura sintática [*N*0 *Vdar com C*1 *em C*2], mas significados bastante variados: o exemplo (6.96) significa "não conseguir o que se desejava, apesar do esforço feito"; já (6.97) significa "falar demais ou falar o que não deveria"; enquanto (6.98) significa "chegar a algum lugar"; e (6.99) significa "chegar a um local e não ser atendido ou não haver ninguém no local".

As quatro expressões apresentam sujeito distribucionalmente livre e dois complementos fixos: *C*1 (*burros*, *língua*, *costados* e *cara*) e *C*2 (água, *dentes*, *Nloc* e *porta*). A terceira expressão, em particular, apresenta certa liberdade ao preenchimento de *C*2, mas ainda com fortes restrições, já que essa posição só pode ser preenchida por um nome locativo (*Nloc*).

Conforme apontado na seção 4.5, a expressão *dar com a cara na porta* deve ser considerada como *EC* porque tanto o verbo quanto os complementos (*C*1 e *C*2) e os determinantes são fixos.

(6.100) *Um morador passou mal e foi lá no posto e* ***deu com a cara na porta****.* [Ex.R]

Por outro lado, existem construções livres (ver seção 4.5) com a mesma estrutura sintática, porém com significado composicional, em que o verbo *dar* tem o sentido de *bater*, como se verifica em:

(6.101) *Descuidei, e ele* ***deu com*** *a cabeça* ***na*** *quina da mesa!* [Ex.R]

Essas construções possuem complementos livres, admitindo fracas restrições de seleção ao seu preenchimento.

O exemplo (6.100) apresenta duas diferenças em relação ao exemplo (6.101): (i) em (6.100), os complementos *cara* e *porta* são fixos, portanto a expressão é representada como [*N*0 *V Prep*1 *C*1 *Prep*2 *C*2], ao passo que (6.101) é representada como [*N*0 *V Prep*1*N*1 *Prep*2 *N*2]; e (ii) o significado de (6.100) é idiomático, ao passo que o significado da construção livre (6.101) é literal.

Considera-se que os complementos *cara* e *porta* são fixos porque, se forem substituídos por outras variantes, passam a ter o sentido literal de "bater com algo em algo", como se observa nos dois exemplos seguintes:

(6.102) *O famoso delegado* ***deu com*** *o ombro* ***na*** *porta, caindo no chão e desarrumando a sua peruca.* [Ex.R]

(6.103) *A garota caiu e* ***deu com*** *a cara* ***na*** *lama.* [Ex.R]

A expressão fixa admite pequenas variações, como (*cara* + *fuça* + *nariz*), na posição de *C*1, e (*porta* + *parede*), na posição de *C*2.

Em situações muito raras, a expressão *dar com a cara na porta* pode ter significado composicional e indicar literalmente "bater na porta com a cara", como se observa pelo exemplo seguinte:

(6.104) *Virou rapidamente pro lado e* ***deu com a cara na porta*** *do armário de cima aberto – Ai! – colocou a mão na testa, onde tinha batido.* [Ex.R]

A expressão *dar de cara com* não possui exatamente a mesma estrutura sintática das *ECs* analisadas nesta seção, porém será inserida nesta classe CPP por falta de outra

mais adequada. A diferença entre elas está no preenchimento de *C*2, que, em geral, é um complemento fixo (*e.g. dar com a cara na* ***porta****, dar com a língua nos* ***dentes****, dar com os burros na* água), mas pode ser livre, no caso de *dar de cara com* ***N***.

Além das *ECs* já analisadas, também fazem parte da classe CPP expressões como:

(6.105) *A unidade da Major Sertorio* ***dá de mil a zero*** *na da ITU, em atendimento, simpatia e prestatividade.* [Ex.R]

(6.106) *A versão de terno branco* ***dá de cem a zero*** *na versão de colante verde ridículo.* [Ex.R]

(6.107) *A comida aqui* ***dá de dez a zero*** *nas refeições que eu fiz nos outros restaurantes franceses da cidade.* [Ex.R]

Nessas expressões, o verbo *dar* tem o significado de *ganhar*, comparando numericamente duas entidades. O primeiro complemento fixo pode ser instanciado por um número (*mil*, *cem* ou *dez*, ou esporadicamente outros números redondos) e o segundo complemento fixo é necessariamente preenchido pelo número *zero*.

As duas entidades comparadas são livres, ou seja, há fracas restrições semânticas para seu preenchimento, desde que sejam entidades de um mesmo tipo (*e.g. duas lojas, duas pessoas, duas marcas etc.*). O verbo e o complemento preposicional *de (mil + cem + dez)* podem admitir pequenas e restritas variações, mas o complemento preposicional *a zero* é fixo, o que desautoriza construções como:

(6.105a) **A unidade da Major Sertorio* ***dá de mil e trezentos a três*** *na da ITU, em atendimento, simpatia e prestatividade.* [Ex.C]

(6.106a) **A versão de terno branco* ***dá de cento e vinte a dez*** *na versão de colante verde ridículo.* [Ex.C]

(6.107a) **A comida aqui* ***dá de dez a oito*** *nas refeições que eu fiz nos outros restaurantes franceses da cidade.* [Ex.C]

O verbo *dar* autoriza pequenas e restritas variações lexicais, tais como *ganhar* e *bater*.

Essas expressões cristalizadas podem ser assim formalizadas: [*Vdar de C*1 *a C*2] e admitem a redução (ou apagamento) do segundo complemento fixo:

(6.105b) *A unidade da Major Sertorio* ***dá de mil*** *na da ITU, em atendimento, simpatia e prestatividade.* [Ex.C]

(6.106b) *A versão de terno branco* ***dá de cem*** *na versão de colante verde ridículo.* [Ex.C]

(6.107b) *A comida aqui* ***dá de dez*** *nas refeições que eu fiz nos outros restaurantes franceses da cidade.* [Ex.C]

O apagamento do segundo complemento fixo (*a zero*) é autorizado, já que é totalmente fixo e não admite variação. A expressão reduzida deve ser formalizada como [*Vdar de C*1].

6.1.8 Classe CPPN

Esta classe é formada por apenas duas *ECs* com o verbo *dar*, as quais podem ser representadas pela estrutura [*N*0 *V C*1 *Prep C*2 *Prep C*3]. São elas:

(6.108) *Cristina* ***deu parte de Brito à polícia*** *e foi se refugiar no Palácio Grão-Pará.* [Ex.R]

(6.109) *O pai ficou tão contente que* ***deu a mão da filha em casamento*** *e batizou a guloseima de pão do Toni.* [Ex.R]

Nos dois casos, a posição de *C*2 pode ser preenchida por um grande número de nomes, porém será considerado como complemento fixo, já que esses nomes sofrem fortes restrições de seleção ao seu preenchimento. Em (6.108), por exemplo, *C*2 deve ser preenchido por um nome humano da pessoa acusada por determinado delito, o que também poderia ser referido pelos nomes *ladrão, bandido, assassino, delinquente, etc*. Já em (6.109), a posição *C*2 é preenchida pelo nome *filha*, mas também poderia ser instanciada pelo nome próprio da filha, o que permite que vários nomes próprios estejam nessa posição.

A expressão referida em (6.108) significa que "*N*0 denunciou *C*2 à polícia". Dado que nem sempre a identidade do bandido (*C*2) é revelada, é possível também inserir outros nomes (*N-hum*) na posição de *C*2, tais como *furto, roubo, briga, ocorrido, assalto, crime, caso, sequestro, abuso, estupro, episódio etc*. Todos eles são nomes eventivos que substituem, por um processo metonímico, o nome do delinquente pelo delito cometido.

Por vezes, o último complemento (*C*3) pode ser apagado nas frases de superfície, como ocorre em (6.110):

(6.110) *Ela já* ***deu parte do ex-marido*** *por diversas vezes.* [Ex.R]

Ressalte-se, no entanto, que, em uma construção de base, o terceiro complemento (à polícia) está presente e é obrigatório.

A mesma estrutura [*N*0 *deu parte de C*2 *a C*3] pode ocorrer como [*N*0 *deu parte de N*2 *a N*3], de forma que a construção deixa de ter sentido figurado e passa a ter o sentido literal de "transferir a posse do objeto *N*2 das mãos de *N*0 a *N*3". Observe-se, por exemplo:

(6.111) *Ela deu parte do dinheiro a ídolo com câncer.* [Ex.R]

Não se pode confundir essa construção plena com a expressão cristalizada expressa em (6.108), apesar de terem superficialmente a mesma configuração sintática. Em (6.111), o verbo *dar* é distribucional, indicando a "doação do dinheiro a alguém". A expressão *parte de*, que é um complemento fixo em (6.108), passa a ser um quantificador em (6.111), indicando uma "parcela de, um pouco de".

Também não se deve confundir a *EC* [*dar parte de C*2 *a C*3] com a *EC* [*dar parte de*] pertencente à classe CDN, já analisada na subseção 6.1.2.

6.1.9 Classe N=Ø

Esta classe foi criada para abarcar expressões cristalizadas que não apresentam sujeito em suas estruturas de base. Em geral, essas *ECs* são formadas por um verbo (*Vdar*), seguido de determinante fixo (*Det*), seguido de um complemento fixo (*C*1), que, por sua vez, é seguido de uma preposição (*Prep*) e de um complemento livre (*N*2). Essas expressões podem ser assim formalizadas: [*V Det C*1 *Prep N*2]. Trata-se de *ECs* como:

(6.112) ***Deu a louca em*** *Justin Bieber?* [Ex.R]

(6.113) *O treinador Nadélio Rocha lamentou a derrota e disse que em certo momento* ***deu um branco na*** *equipe.* [Ex.R]

(6.114) *Quando* ***deu a hora de*** *ir para casa, eu fui.* [Ex.R]

(6.115) *Hoje não* ***deu tempo de*** *efetuar o pagamento de parte do salário.* [Ex.R]

As duas primeiras expressões, (6.112) e (6.113), exigem a preposição *em*, ao passo que as duas últimas, (6.114) e (6.115), exigem a preposição *de*. Além disso, as primeiras selecionam um *Nhum* para a posição de *N*2, ao passo que as últimas selecionam tipicamente uma oração infinitiva (*V infw*) para a mesma posição.

Há outras propriedades que distinguem os dois tipos de *EC* dentro da classe residual (*N*0=Ø), por exemplo, o fato de (6.112) e (6.113) aceitarem o pronome oblíquo de primeira pessoa *mim* na posição de *N*2 e esse pronome poder passar para a posição pré-verbal como pronome pessoal do caso reto *me*:

(6.112a) *Deu a louca em mim.* [Ex.C]

(6.112b) = *Me deu a louca.* [Ex.C]

(6.113a) *Deu um branco em mim.* [Ex.C]

(6.113b) = *Me deu um branco.* [Ex.C]

Essa transformação não é autorizada nas expressões *dar a hora de V infw* e *dar tempo de V infw*. Apesar de as *ECs* da classe residual não terem estruturas e/ou propriedades idênticas, serão inseridas na mesma classe por possuírem uma característica em comum: a inexistência do sujeito (*N0=Ø*).

6.2 Os provérbios

Provérbios, ditos populares, ditados populares, ditados, prolóquios, adágios, parábolas e *máximas* são apenas alguns dos termos utilizados na literatura para se referirem a objetos semelhantes: frases cujos constituintes são fixos, admitindo pouca ou nenhuma mobilidade, e podendo apresentar uma, duas ou até três proposições. Assim como as *ECs*, o significado das partes de um provérbio não necessariamente condiz com o significado global da sentença, a qual deve ser interpretada em seu sentido metafórico.

O maior problema em relação à identificação de provérbios é que eles possuem a mesma estrutura sintática e as mesmas palavras comuns das frases livres, porém eles possuem geralmente um significado não-composicional e, por isso, devem ser reconhecidos, não como uma sequência de palavras livres, mas como uma unidade complexa, formada por palavras, sintagmas e proposições fixas.

O que diferencia as Expressões Cristalizadas dos provérbios são as restrições para preenchimento lexical da posição de sujeito. Os provérbios têm a posição sujeito obrigatoriamente preenchida por um elemento fixo, enquanto as *ECs* possuem sujeito distribucionalmente livre. Além disso, os provérbios funcionam pragmaticamente com um cunho moral ou social em cada cultura.

Apesar de haver tantas nomenclaturas designando o mesmo tipo de objeto linguístico – e deve-se considerar que há diferenças conceituais entre esses termos – adotaremos neste trabalho o termo *provérbio* para englobar todas essas modalidades de sentenças fixas.

Quando os provérbios são introduzidos por marcas enunciativas (*e.g. como dizem por aí, como dizia a minha avó, dizem por aí que, costuma-se falar que, etc.*), é mais

fácil identificar essas estruturas, pois as marcas enunciativas podem ser descritas extensivamente. No entanto, geralmente não há marcas introdutórias desses discursos, o que constitui uma dificuldade para encontrá-los em textos.

O estudo, a catalogação e a classificação dos provérbios são necessários não apenas para a linguística teórica – no sentido de prover a descrição do sistema de uma língua – mas também por sua aplicação prática em sistemas de Processamento Automático de Línguas Naturais (PLN).

Segundo Conenna (2000b), as parêmias[36] são de interesse não apenas da Paremiologia Linguística, mas também da Linguística Informática ou Linguística Computacional. As ferramentas e aplicações de PLN, tais como tradutores automáticos, sumarizadores automáticos, simplificadores textuais, sistemas de auxílio à escrita, dentre outros, precisam de léxicos computacionais que possam ser lidos e interpretados automaticamente pela máquina. Nesse sentido, a proposta de classificar os provérbios e descrever seus constituintes poderá subsidiar a construção de dicionários eletrônicos ou outros recursos léxico-computacionais que permitam que a máquina interprete um provérbio não como uma frase livre, mas como uma frase fixa.

Em trabalho anterior (RASSI; BAPTISTA; VALE, 2014), propusemos uma classificação sintática para os provérbios, com base em 12 classes, considerando-se os seguintes critérios de classificação, nesta ordem:

1. o número de proposições (uma, duas ou três proposições ou unidades proposicionais);
2. a coordenação (nos provérbios com mais de uma proposição);
3. a ordem das proposições (principal e subordinadas), no caso de provérbios com mais de uma proposição;
4. a ordem dos constituintes, nos provérbios com apenas uma proposição;
5. a existência de construções impessoais;
6. a negação obrigatória.

A Tabela 3 apresenta a classificação estabelecida naquele trabalho (RASSI; BAPTISTA; VALE, 2014).

36 A autora usa o termo *parêmia* para designar o mesmo objeto a que nos referimos, neste trabalho, pelo nome de *provérbio*.

Classe	Estrutura	Exemplo	nº
P1F1	Ø V w	*Não há crime sem lei*	20
P1F2	*N0 V cop Adj w*	*A carne é fraca*	53
P1F3	*N0 V w*	*O hábito (não) faz o monge*	80
P1F4	*N0 Neg V w*	*Burro velho não aprende línguas*	53
P1F5	*Prep Ni N0 V w*	*Para bom entendedor, meia palavra basta*	45
P2F1	*F1 Conjs-comp F2*	*Mais vale um pássaro na mão do que dois voando*	39
P2F2	*F1 Conjc F2*	*A palavra é de prata e o silêncio é de ouro*	71
P2F3	*N1, N2*	*Tal pai, tal filho*	48
P2F4	*Qu- F1 F2*	*Quem tem boca vai a Roma*	90
P2F5	*F1 Conjs F2*	*Os amigos são muitos quando grande é a abastança*	20
P2F6	*Conjs F2, F1*	*Quando a esmola é demais, o santo desconfia*	28
P3	*F1, F2, F3*	*Um é pouco, dois é bom, três é demais*	47
Total			**594**

Tabela 3 – Classificação formal dos provérbios brasileiros
Fonte: Adaptada de Rassi, Baptista e Vale (2014, p.4)

A primeira coluna apresenta a classe; a segunda indica a configuração formal[37] (estrutura sintática) dos provérbios; a terceira demonstra um exemplo característico; e, por fim, a última coluna indica o número de provérbios inseridos em cada classe.

Dos 594 provérbios extraídos automaticamente de *corpus* (RASSI; BAPTISTA; VALE, 2014), apenas 26 são formados pelo verbo *dar*, tais como:

(6.116) *Dê tempo ao tempo.* [Ex.R]

(6.117) *Dai a Cesar o que é de César.* [Ex.R]

A maioria dos provérbios formados por *dar* possuem uma única proposição, como os exemplos anteriores. Apenas três provérbios possuem duas proposições:

37 Leia-se *Adj* para adjetivo; *Conjc* para conjunção coordenativa; *Conjs* para conjunção subordinativa; *Conjs-comp* especificamente para conjunção subordinativa comparativa; F_1, F_2 e F_3 para a primeira, segunda e terceira proposição (ou cláusula), respectivamente; N_0 para o sujeito; N_i para um nome em qualquer posição sintática; *Neg* para negação obrigatória; *Prep* para preposição; *Qu-* para os pronomes interrogativos *quem* ou *o que*; *V* para verbo; *Vcop* para o verbo de cópula *ser*; *w* para qualquer sequência não especificada de complementos; e Ø para qualquer posição sintática vazia.

(6.118) *Cana na fazenda dá pinga; pinga na cidade dá cana.* [Ex.R]

(6.119) *Se queres conhecer o vilão, dê-lhe o bastão.* [Ex.R]

(6.120) *Se a vida lhe der um limão, faça uma limonada.* [Ex.R]

O primeiro exemplo é um provérbio com duas proposições declarativas, enquanto os demais são constituídos de duas proposições, sendo que a primeira é uma oração condicional e a segunda, uma proposição declarativa. O provérbio (6.119) admite variações tanto do verbo *dar* quanto do nome *bastão*, como demonstram os exemplos:

(6.119a) *Se queres conhecer o vilão, põe-lhe o pau na mão.* [Ex.C]

(6.119b) *Se queres conhecer o vilão, dê-lhe o poder na mão.* [Ex.C]

O provérbio (6.120) também admite variação do nome *limão*, em um jogo de palavras e sentidos: *Se a vida lhe der um limão, faça uma caipirinha.*

Há ainda um provérbio que possui três proposições: *Se conselho fosse bom, ninguém dava; vendia.*

É possível encontrar outras regularidades dentre os provérbios constituídos pelo verbo *dar*. Por exemplo, alguns deles exigem a negação obrigatória, tais como:

(6.121) *Aço ruim não dá bom fio.* [Ex.R]

(6.122) *Deus não dá asa à cobra.* [Ex.R]

(6.123) *Deus não dá nozes a quem não tem dentes.* [Ex.R]

(6.124) *Dinheiro não dá em árvore.* [Ex.R]

(6.125) *Jabuticaba não dá em cacho.* [Ex.R]

Os provérbios (6.124) e (6.125) são bastante semelhantes sintática e semanticamente. Um outro provérbio que também exige a negação obrigatória é: *Dinheiro não dá felicidade*. Apesar de considerarmos que essa forma do provérbio, com o verbo *dar*, é uma variante do provérbio *Dinheiro não traz felicidade*, ainda assim a forma proverbial com *dar* é bastante recorrente no *corpus*.

Assim como ocorre nos provérbios (6.122) e (6.123), há vários outros que fazem referência a *Deus*, tanto com negação obrigatória, quanto em proposições afirmativas, como:

(6.126) *Deus dá o frio conforme a roupa.* [Ex.R]

(6.127) *Quem dá aos pobres empresta a Deus.* [Ex.R]

Essa segunda estrutura proverbial (6.127), introduzida pelo pronome *Quem*, também aparece em outros provérbios, como se demonstra por meio dos exemplos:

(6.128) *Quem dá aos pobres paga a conta do motel.* [Ex.R]

(6.129) *Quem dá pão dá criação.* [Ex.R]

(6.130) *Quem fala demais dá bom dia a cavalo.* [Ex.R]

O primeiro exemplo (6.128) poderia ser considerado uma variante de (6.127), porém nós os classificamos como provérbios diferentes, já que não há nenhum elemento lexical comum aos dois predicados (*empresta a Deus vs. paga a conta do motel*).

Por fim, para concluir essa breve explanação sobre os provérbios constituídos pelo verbo *dar*, indica-se uma última configuração formal de provérbios: [*Prep Ni N0 V w*], onde se inserem:

(6.131) *Em terra de saci, calça comprida dá pra dois.* [Ex.R]

(6.132) *Em terra de sapo, mosquito não dá rasante.* [Ex.R]

Em Rassi, Baptista e Vale (2014), identificamos 45 provérbios com essa estrutura, dos quais dois apresentam o verbo *dar*.

Esta seção não teve o intuito de se aprofundar na análise e descrição dos provérbios em Português. Nosso objetivo foi basicamente apresentar a categoria de *provérbios*, que é um subtipo de *frase fixa*, já que existem construções em Português que se classificam nessa categoria. Um estudo mais aprofundado sobre os tipos sintáticos e a classificação dos provérbios em Português pode ser encontrado em Rassi, Baptista e Vale (2014). A lista dos provérbios que contêm o verbo *dar* consta no final do Apêndice C.

Ao longo deste capítulo, descrevemos dois tipos de frases fixas: as expressões cristalizadas e os provérbios. No *corpus* adotado, foram encontradas 173 expressões cristalizadas com o verbo *dar*, as quais foram classificadas e descritas a partir de suas estruturas sintáticas. Também foram encontrados 45 provérbios constituídos pelo verbo *dar*, os quais foram inseridos numa classificação previamente estabelecida para provérbios em Português.

Primeiramente, definimos os critérios formais a serem adotados para identificar as expressões cristalizadas. Depois, as 173 *ECs* foram distribuídas em nove classes de acordo com: o número de argumentos fixos, o tipo semântico dos argumentos, o tipo de preposição e a estrutura sintática das expressões. Cada classe foi sistematicamente analisada e descrita ao longo deste capítulo. Por fim, os provérbios também foram classificados e descritos seguindo critérios sintáticos.

Este capítulo encerra a primeira parte deste livro, que consiste em uma proposta taxonômica para as construções com o verbo *dar*. Essa taxonomia poderá posteriormente ser adaptada e expandida para a descrição de outros verbos da Língua Portuguesa. A partir do próximo capítulo, introduziremos a Parte II do livro, que trata especificamente das construções em que o verbo *dar* é classificado como verbo-suporte e é selecionado por um nome predicativo.

Parte II

Construções com o verbo-suporte *dar* e nome predicativo

7. O conceito de *construção nominal com verbo-suporte*

Neste trabalho, adotamos a noção de verbo-suporte e verbo-operador da gramática transformacional de Harris (1991). Para identificar uma construção com verbo-suporte (*CVS*), é necessário identificar tanto os verbos que podem funcionar como *Vsup* quanto os nomes predicativos, que se distinguem dos demais nomes por possuírem propriedades de nomes eventivos e serem, tipicamente, substantivos abstratos, podendo ser ou não derivados de algum verbo (*abraço/abraçar*) ou adjetivo (*predileção/predileto*), ou ainda de outro nome (*cajadada/cajado*).

Há diferentes testes que podem ser usados para identificar as construções com verbo-suporte, como a substituição da construção com *Vsup* por um verbo pleno correspondente (como *dar um abraço = abraçar*, ou *dar um beijo = beijar*); a estreita relação entre o *Npred* e um de seus argumentos (por exemplo, *O Rui deu um beijo na Ana*, interditando a construção **O Rui deu um beijo do Zé na Ana*); as restrições sobre os determinantes (como *Ana deu uma passeada no parque*, interditando a construção **Ana deu minha passeada no parque*); dentre outros testes. Todos os testes se baseiam na relação que existe entre o *Npred* e o *Vsup* que ele seleciona ou entre o *Npred* e os argumentos que ele seleciona. Para definir *CVS*, portanto, é necessário definir *verbo-suporte* e *nome predicativo*. Esses dois termos serão discutidos nas próximas subseções.

7.1 O conceito de verbo-suporte

A noção de verbo-suporte tem sido usada há muito tempo, sob diferentes perspectivas teóricas e várias terminologias diferentes. Neste trabalho, adotamos a noção proposta na gramática transformacional de operadores. Harris (1964, p.216-217) propôs o conceito de *verbo-operador* e o nomeou como um tipo particular de construção com o operador de nominalização "U", ligando frases como *Ele estuda eclipses = Ele faz estudos sobre eclipses*.

Posteriormente, Harris (1976, 1978, 1982), adotando uma visão não-orientada das transformações, considerou esses dois tipos de frases como formas elementares do *kernel* da língua, parafrásticas e equivalentes. Foi, no entanto, Maurice Gross (1981) quem primeiro elaborou a definição de verbo-suporte que será utilizada neste trabalho; é uma definição que resistiu ao teste do tempo, por um processo de acumulação de evidências em um grande número de línguas, abrangendo diferentes famílias linguísticas (Românico, Grego moderno, Koreano, Kikongo), diversos tipos de construções e poder explicativo teórico.

Essa definição "moderna" do conceito de *construção com verbo-suporte* leva em consideração a noção de predicado semântico, que é um conceito-chave para muitas análises, pois divide os enunciados do discurso em suas unidades de significado.

Na verdade, tendo como base a noção harrissiana de operador linguístico, Gross (1981) postula que a unidade de sentido não é a palavra, mas a frase simples (ou frase elementar), já que a expressão linguística é determinada por um elemento predicativo (por exemplo, um verbo pleno), juntamente com todos os seus argumentos essenciais (sujeito e eventuais complementos essenciais). Nesse sentido, a relação entre frases semelhantes (*Ele estuda eclipses* e *Ele faz estudos sobre eclipses*) é de natureza parafrástica pelo fato de que elas não apenas possuem as mesmas restrições de seleção dos argumentos, mas também porque possuem o mesmo predicado semântico: no primeiro caso, o verbo *estudar*; no segundo caso, o nome predicativo *estudo*.

O fato de que o verbo (*estudar*) e o nome predicativo (*estudo*) são morfologicamente relacionados e que as frases podem ser consideradas parafrásticas tem um estatuto especial no quadro teórico harrissiano: trata-se de uma transformação unária, ou seja, uma operação formal que transforma uma frase em outra, sem, contudo, alterar a informação contida na frase, estabelecendo, assim, classes de equivalência entre frases. Enquanto a informação veiculada pelo par de sentenças é a mesma, os valores gramaticais podem mudar de forma sistemática, ou seja, as duas frases expressam o mesmo predicado semântico, mas com pequenas diferenças de aspecto, modalidade ou valor estilístico. Isso leva a uma definição do estatuto gramatical do verbo-suporte nessas frases.

O verbo-suporte funciona como uma espécie de "auxiliar" do nome predicativo, pois é ele quem carrega as marcas gramaticais de número, pessoa, tempo, modo e aspecto verbal, as quais não podem ser veiculadas pelo nome predicativo, devido à sua própria morfologia. Como resultado do seu estatuto "auxiliar" em relação ao nome predicativo, o verbo-suporte não é considerado um verbo distribucional, portanto não se pode defini-lo intencionalmente no escopo dos seus argumentos. Os *Vsup* se distinguem dos verbos plenos porque estes últimos são verbos distribucionais e podem ser definidos intencionalmente. Nas construções com verbo pleno, é o verbo que seleciona os argumentos nas frases elementares. Já nas *CVS*, é o nome predicativo que seleciona tanto os seus argumentos quanto o verbo-suporte mais apropriado.

Os verbos-suporte são selecionados pelos nomes predicativos, e não vice-versa. Naturalmente, a escolha do verbo-suporte é por vezes intimamente relacionada ao predicado semântico expresso pelo substantivo. Por exemplo, os substantivos que

exprimem ações envolvendo duas entidades humanas são, muitas vezes, construídos com os verbos-suporte *fazer* ou *dar*, enquanto substantivos que exprimem propriedades ou estados são mais frequentemente construídos com o verbo-suporte *ter*.

As *CVS* são também referidas na literatura como *construções com verbo leve* por diversos autores (SCHER, 2004; DURAN; ALUÍSIO, 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.

Neste texto, adotamos o termo *verbo-suporte* por considerarmos que ele serve, de fato, para "suportar" as marcas verbais de tempo, modo, número, pessoa e aspecto. Vale ressaltar que, diferentemente de outros autores, consideramos que o conceito de verbo-suporte não é apenas sintático, possuindo também sua contraparte semântica[38].

A segunda motivação para a escolha do termo *verbo-suporte* se refere ao fato de que é um conceito mais controlado linguisticamente, ou seja, há propriedades formais bem definidas que o distinguem de outras categorias verbais. Os autores que optam pelo termo *verbo leve* consideram-no, em geral, semanticamente vazio ou esvaziado de sentido, mas não há como mensurar o grau ou nível de preenchimento ou esvaziamento semântico de um verbo, portanto essa definição é subjetiva e intuitiva.

O conceito de verbo leve não é consensual na literatura. Há autores que pregam desde o esvaziamento completo do verbo até alguns processos de dessemantização, em que o verbo perde parte de seus traços semânticos. De fato, há diferentes tipos de verbos leves (ou verbos-suporte): Ruppenhofer *et al.* (2006) distinguem cinco classes de verbo-suporte (*plain vanilla*, *aspectual*, *ponto de vista*, *registro* e *causativo*); Gross (1998) identifica outras cinco classes, sendo algumas coincidentes (*aspecto*, *causativo*, *verbos suporte intensivos*, *verbos suporte conjuntivos* e *modalidades especializadas*), além de tratar das relações semânticas entre verbos-suporte.

A literatura oferece inúmeras definições para verbo-suporte, que vão desde conceitos estritamente sintáticos, baseados na composição estrutural das sentenças formadas por *Vsup*, ou na correlação necessária entre a *CVS* e um verbo pleno morfologicamente associado ao *Npred*, até abordagens discursivas que tratam da motivação para o uso do *Vsup* e os diferentes efeitos de sentido causados por seu uso ou não.

No mesmo sentido, existem trabalhos de base gerativista sobre os *Vsup*, dentre os quais se destacam as pesquisas de Scher (2004, 2005, 2006) sobre o verbo *dar* e as nominalizações em *-ada*. Há estudos funcionalistas sobre os verbos-suporte, que

38 Veja-se, por exemplo, Gross (1998), que trata das funções semânticas dos verbos-suporte.

investigam os usos dessas construções, a intenção do falante e a motivação linguística para tais usos em diferentes registros e diferentes gêneros textuais; citam-se como exemplos os trabalhos de Radford (1988) e Neves (1996, 2000, 2002), dentre outros. A abordagem utilizada nesta pesquisa, por outro lado, é de base estruturalista, apoiada pelo modelo do Léxico-Gramática, seguindo a mesma abordagem teórica proposta por Gross (1975, 1981), Leclère (1971), Giry-Schneider (1978, 1987), Labelle (1984), Ranchhod (1990) e outros.

7.2 O conceito de nome predicativo

Os nomes predicativos, à semelhança dos verbos e dos adjetivos predicativos, são predicadores de frases, ou seja, são operadores que selecionam argumentos. Os substantivos podem funcionar como argumentos de um predicador verbal (como em ***Ana*** *beijou* ***Rui***) ou como predicadores que selecionam outros substantivos como argumentos (*Ana deu um* ***beijo*** *em Rui*). Nas duas frases, os substantivos *Ana* e *Rui* são argumentos. A diferença entre elas é que, na primeira frase, o elemento predicador é um verbo (*beijou*), enquanto, na segunda frase, o elemento predicador é um substantivo (*beijo*). Quando o predicador da frase é um substantivo, ele é chamado de *nome predicativo*.

A característica principal do nome predicativo, em construções com verbo-suporte, é a relação particular que ele mantém com o sujeito da frase. Em *Ana deu um beijo em Rui*, existe uma relação entre o *Npred beijo* e o sujeito *Ana*, que é da mesma natureza que a relação entre o verbo *beijar* e o sujeito *Ana* na frase *Ana beijou Rui*. Essa propriedade é definitória das construções com verbo-suporte e nome predicativo, pois não se aplica a outros substantivos que não sejam predicativos. Outras propriedades dos *Npred* - e, por extensão, das *CVS* - serão apresentadas no Capítulo 8.

Tipicamente, os nomes predicativos são substantivos abstratos eventivos. Os *Npred* obedecem às mesmas regras morfológicas de flexão e derivação de todos os outros substantivos (com sufixos nominais típicos *-ção*, *-mento*, *-dade*, *-ismo etc.*), mas sintático e semanticamente funcionam como outros predicadores, podendo indicar ação, processo, ação-processo, estado ou outros.

Os *Npred* podem ser de dois tipos: (i) os nomes predicativos deverbais, deadjetivais ou denominais, ou seja, aqueles que são morfologicamente associados a verbos, adjetivos ou outros nomes, respectivamente, e são o resultado de nominalizações (como *murro/esmurrar*, *tiro/atirar*, *sinal/sinalizar*, *coragem/corajoso*, *competência/competir/competente*, *mãozada/mão*, *cajadada/cajado*); ou (ii) nomes predicativos autônomos,

ou seja, aqueles que não possuem a mesma raiz de verbos e/ou adjetivos (como *carona*, *greve*, *sermão*, *tilt*, *toco*, *sopapo*, *piparote*).

Uma das propriedades das *CVS* consideradas essenciais por alguns autores (LANGER, 2004; TU; ROTH, 2011) é a equivalência entre a nominalização da *CVS* e um verbo pleno, ou seja, somente são considerados como *CVS* os casos em que há equivalência morfológica entre o nome predicativo e um verbo pleno com a mesma raiz, *i.e. participação* ≡ *participar*. Já foi provado (GROSS, 1981), no entanto, que outros nomes predicativos que não possuem equivalência morfológica com verbos ou adjetivos possuem as mesmas propriedades das nominalizações, ou seja, devem ser considerados igualmente como predicadores em *CVS*. Nesse sentido, a existência da nominalização não é uma propriedade necessária para classificar a construção como *CVS*.

Além disso, há também nomes predicativos que são o resultado de uma nominalização, mas não devem ser classificados como *CVS* porque não constituem predicados de base. O nome predicativo *medo*, por exemplo, é morfologicamente associado ao verbo *amedrontar* e ao adjetivo *medroso*; no entanto, esse *Npred* nem sempre constitui uma *CVS*. Quando associado ao verbo-suporte *ter* (*Ana tem medo de cobra*), é uma *CVS*, porém, quando associado ao *VopC dar* (*A cobra deu medo em Ana*), não é uma *CVS*, e sim, uma construção causativa.

Em suma, consideramos, ao contrário de outros autores (LANGER, 2004; TU; ROTH, 2011), que a equivalência entre a nominalização e um verbo pleno não é uma propriedade necessária nem suficiente das *CVS*. Pode ser considerada uma característica geral, pois se aplica à maior parte dos casos, mas não é uma propriedade definitória das *CVS*.

Em relação aos *Npred* que selecionam o verbo-suporte *dar*, existe uma grande produtividade de *Npred* deverbais, que são derivados de verbos acrescidos do sufixo *-da* (*andar + -da = andada*; *mexer + -da = mexida*, *fugir + -da = fugida*). Essas nominalizações em *-da* podem indicar: (i) o aspecto verbal – como *dar uma passada*, *dar uma lida*, *dar uma olhadinha*; ou (ii) *bater com determinado objeto*, em que se acrescenta o sufixo *-da*, não ao verbo, mas ao nome do instrumento (*dar uma martelada* ≡ *bater com um martelo*, *dar uma marretada* ≡ *bater com uma marreta*). Por outro lado, há também *Npred* que são derivados de adjetivos e, portanto, são considerados como deadjetivais (*clareada* deriva de *claro*, *madurada* deriva de *maduro*). Há ainda um conjunto de *Npred*, em geral com sufixo *-ada*, que são derivados de objetos concretos ou nomes de partes do corpo (*dar uma cotovelada* ≡ *bater com o cotovelo*, *dar uma cinzeirada* ≡ *bater com um cinzeiro*). Os tipos semânticos dos *Npred* serão discutidos detalhadamente no Capítulo 10.

Este capítulo introduziu a segunda parte do livro, apresentando a definição de construção com verbo-suporte. As *CVS* são compostas por um verbo-suporte e um nome predicativo, que formam uma unidade de sentido. Assim, este capítulo discutiu os conceitos de *verbo-suporte* e de *nome predicativo*, a partir de várias perspectivas teóricas, tendo adotado as definições fornecidas pelo modelo do Léxico-Gramática.

O próximo capítulo apresentará as propriedades gerais definitórias das *CVS*, discutindo quais dessas propriedades se aplicam às *CVS* com *dar* e algumas especificidades sobre essas propriedades gerais.

8. Propriedades gerais das CVS

As construções com verbo-suporte possuem estrutura sintática aparentemente semelhante à das construções com verbo pleno. O que as distingue são as propriedades sintáticas que se aplicam às *CVS* e não às construções com verbo pleno. A principal propriedade é a existência de uma relação semântica particular que liga o nome predicativo a algum de seus argumentos, tipicamente o nome na posição sujeito[39]. Essa relação é da mesma natureza semântica como a que liga o sujeito a um verbo pleno comum. Além dessa relação, considerada como principal, há outras propriedades também características das *CVS* e que serão apresentadas ao longo deste capítulo.

Serão consideradas seis principais propriedades gerais das construções com verbos-suporte: (i) a relação particular entre o *Npred* e algum de seus argumentos; (ii) as restrições sobre os determinantes; (iii) a descida do advérbio; (iv) a dupla análise dos complementos preposicionais; (v) a redução do *Vsup* e a formação de grupo nominal; e (vi) a possibilidade de variação do *Vsup*.

Apesar de serem propriedades gerais das *CVS*, existem particularidades entre os diferentes verbos-suporte ou entre as combinações de *Vsup* e *Npred*. De fato, as *CVS* são de tipos muito variados, sendo que a principal característica que as agrupa é a presença de um verbo-suporte elementar e um nome predicativo. Há, portanto, uma grande diversidade de objetos e fenômenos nesse conjunto a que nos referimos como *construções com verbo-suporte*.

8.1 Relação particular entre o *Npred* e um de seus argumentos

Essa propriedade consiste na estreita relação que existe entre o nome predicativo e tipicamente o nome que ocupa a posição sujeito. Essa relação é da mesma natureza semântica que a que liga um verbo pleno ao seu sujeito. Observe-se:

(8.1) *A Ana deu um beijo no Rui.* [Ex.C]

(8.1a) ≡ *A Ana beijou o Rui.* [Ex.C]

A relação semântica existente entre os nomes beijo e Ana, em (8.1), é semelhante à relação entre o verbo beijou e o nome Ana, em (8.1a). Ana é o sujeito AGENTE de beijo, assim como o é de beijar. Assim, consideramos que quem seleciona o sujeito Ana é o verbo beijou, em (8.1a), mas o nome beijo, em (8.1).

39 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*).

Essa propriedade faz com que seja impossível colocar o nome predicativo sob o escopo de referência de qualquer outra entidade diferente do próprio sujeito. Ou, em outras palavras, "dada a relação particular que se observa entre o sujeito do *Vsup* e o nome predicativo, este nome não pode receber determinantes que o situem fora da esfera de referência do sujeito" (BAPTISTA, 2005b, p.24). Assim, frases como *?*A Ana deu o meu beijo no Rui* ou *?*A Ana deu o beijo do Zé no Rui*, se aceitáveis, só podem ser interpretadas no sentido de "transmitir ao Rui o beijo que eu ou o Zé lhe mandara" ou "beijar o Rui da mesma forma como o Zé o faz", por exemplo. Nesse sentido, o sujeito da frase não é apenas o sujeito gramatical do verbo-suporte *dar*, mas é também o AGENTE do nome predicativo *beijo*.

Essa relação particular entre o *Npred* e algum de seus argumentos ocorre tipicamente com o argumento que ocupa a posição sujeito. Existem casos, porém, em que essa relação se verifica entre o *Npred* e seu complemento, como nos casos de construções conversas. Na transformação da construção *standard* (*e.g. A Ana deu um beijo no Rui*) para a construção conversa (*e.g. O Rui recebeu um beijo da Ana*), os argumentos sujeito e complemento trocam de posição relativa. Isso faz com que a relação semântica particular que existe entre o *Npred* e o sujeito, na construção *standard*, passe a existir entre o *Npred* e o complemento, na construção conversa.

Em ambos os casos, essa relação é da mesma natureza semântica que a que liga o verbo ao sujeito, numa construção ativa (*e.g. A Ana beijou o Rui*), ou a que liga o verbo ao agente da passiva, numa construção passiva (*e.g. O Rui foi beijado pela Ana*). Tanto na voz ativa quanto na passiva, *Ana* é o agente da ação de *beijar*. Uma discussão mais aprofundada sobre as construções conversas será apresentada na subseção 9.3.2.

A propriedade apresentada nesta seção é a única considerada necessária e suficiente para a identificação das *CVS*. Outras propriedades indicativas de *CVS* serão abordadas nas próximas seções.

8.2 Restrições sobre os determinantes

Essa propriedade diz respeito às restrições que o *Npred*, juntamente com seu *Vsup*, impõe sobre o determinante que o introduz. Inicialmente, é preciso delimitar o conjunto de elementos que constituem a classe dos determinantes. São considerados determinantes os artigos definidos, os indefinidos, os pronomes possessivos, os demonstrativos, os indefinidos, os numerais e o determinante zero (Ø), que corresponde à ausência de determinante. Também é possível que esses determinantes se combinem entre si (*e.g.*

os seus, alguns dos, aquela sua etc.) ou se combinem com modificadores (*Modif*), como é o caso do artigo indefinido acompanhado de *Modif* (*e.g. A Ana deu **um** beijo **longo** no Rui*) ou o artigo definido acompanhado de *Modif* (*e.g. A Ana deu **os sinceros** parabéns ao Rui*).

Como regra geral, os *Npred*, em construções com verbo-suporte, admitem livremente qualquer determinante, porém essa regra pode ser limitada pela propriedade anterior (da relação particular entre o *Npred* e um de seus argumentos). Nesse sentido, os pronomes possessivos que violem essa relação são interditados, como se verifica em:

(8.2) *A Ana deu (um + *o meu + *o teu) beijo ao*[40] *Rui*. [Ex.C]

Isso seria possível em construções com o verbo pleno *dar*.

(8.3) *A Ana deu (um + o meu + o teu) livro ao Rui.* [Ex.C]

O uso do pronome possessivo é uma forma de determinar o referente do *Npred*. Quando esse possessivo não é correferente do sujeito do *Npred*, nas *CVS*, então esse pronome é interditado, como ocorre em (8.2). Isso ocorre porque o uso dos possessivos é uma forma de realização do complemento *de N*. Os possessivos de (8.2) poderiam ser reformulados como um complemento *de N*, como ocorre em:

(8.2a) **A Ana deu o beijo **do Zé** no Rui*. [Ex.C]

Essa construção é interditada porque o complemento *do Zé* não é correferente do sujeito *A Ana*, tal como ocorre com os pronomes possessivos em (8.2).

Já os pronomes demonstrativos não são considerados determinantes apropriados para os *Npred* em *CVS* porque apontam para um referente que não está contido na frase de base.

(8.4) **A Ana deu (esse + este + aquele) beijo no Rui.* [Ex.C]

Essas construções podem ser consideradas aceitáveis – e, de fato, são usuais – quando se considera um contexto linguístico maior que a frase simples, já que os demonstrativos apontam para um referente já conhecido ou para um referente exterior ao texto. Como o Léxico-Gramática reconhece como unidade de análise linguística a frase simples, então (8.4) será julgada como inaceitável, já que não é possível recuperar os referentes desses demonstrativos dentro dessa frase de base.

40 A preposição preferencial nesta e em várias outras construções com o verbo-suporte *dar*, no Português Brasileiro, é a *Prep em* (*e.g. A Ana deu um beijo no Rui*), porém a preposição *a* também é aceitável e será utilizada, neste caso, para que se possa comparar os exemplos (8.2) e (8.3).

Esse problema é semelhante ao dos artigos definidos. *A priori*, os definidos não são determinantes apropriados para os *Npred* porque eles também remetem para um referente exterior à frase de base.

(8.5) **A Ana deu a organizada nos materiais.* [Ex.C]

Ao admitir o artigo definido, perde-se a correferência entre o *Npred* e o sujeito. O referente de *organizada* só pode ser recuperado fora dos limites dessa frase de base. Assim, considera-se que (8.5) poderia ser aceitável em uma interpretação particularizante, a partir da redução de uma oração relativa (*e.g. a organizada que lhe fora solicitada pelo chefe*), mas essa análise ultrapassa o escopo desta pesquisa.

Por outro lado, quando o definido é associado a um *Modif*, a construção é aceitável, já que o modificador muitas vezes explica ou restringe a determinação imposta pelo artigo definido.

(8.6) *A Ana deu a organizada final nos materiais.* [Ex.C]

O uso do artigo definido é autorizado pelo *Modif final*, que restringe a referência de *organizada*.

Conforme foi apontado no início desta seção, o determinante do *Npred* é, via de regra, livre, o que é indicado pelo artigo indefinido ou qualquer outra forma de indefinição (pronomes indefinidos, numerais cardinais e outros). Em raros casos, porém, o *Npred* exige um determinante que não seja indefinido. Observe-se, por exemplo:

(8.7) *O padre deu (a + *uma) (excomunhão + extrema unção) ao fiel.*[41] [Ex.C]

Os *Npred* tais como *excomunhão* ou *extrema unção* se referem a ações ou processos pontuais, que ocorrem uma única vez. Esses casos exigem a presença do artigo definido. Por outro lado, há também outros *Npred* que admitem o determinante indefinido com mais dificuldade. Observe-se, por exemplo:

(8.8) *A Ana deu (os + seus + uns + Ø) parabéns ao Rui.* [Ex.C]

(8.9) *O médico deu (*a + *sua + *uma + Ø) alta ao paciente.* [Ex.C]

Em (8.8), o *Npred parabéns* admite qualquer tipo de determinante, mas é bastante raro com o determinante indefinido. Já em (8.9), o *Npred alta* não admite artigo definido, indefinido nem possessivo, mas exclusivamente o determinante zero (Ø).

41 Em contextos muito específicos, o determinante indefinido pode introduzir os nomes predicativos *excomunhão* e *extrema unção*, mas desde que a preposição que introduza o complemento seja *em*, e não a (*e.g. O bispo de Bauru deu uma excomunhão naquele Padre Beto* [Ex.R] ou *O padre realmente deu uma extrema unção em Senna* [Ex.R]).

Note-se que, mesmo nos casos em que o artigo indefinido não é o mais característico, ainda assim é possível usá-lo, por exemplo, associado a um modificador, como é o caso de:

(8.10) *A Ana deu uns sinceros parabéns ao Rui.* [Ex.C]

(8.11) *O médico deu uma alta provisória ao paciente.* [Ex.C]

Assim, o determinante indefinido (*uns parabéns* e *uma alta*) é considerado aceitável quando acompanhado de um modificador, que é obrigatório nesses casos. Por vezes, a presença do determinante indefinido obriga também a presença de um modificador obrigatório. Isso acontece com a grande maioria dos *Npred* que selecionam o verbo-suporte *ser de* (compare-se, por exemplo, **A Ana é de uma inteligência vs. A Ana é de uma inteligência extraordinária*).

Já que, via de regra, as *CVS* admitem determinante livre, então os casos específicos em que o *Npred* exige um tipo específico de determinante podem ser vistos como um princípio de fixidez, ou seja, um indício de cristalização. Os critérios formais para a classificação de certas combinações de [*V N*] como expressões fixas foram abordados no Capítulo 6.

Além dos tipos já mencionados, também podem funcionar como determinantes os numerais cardinais e os pronomes indefinidos, ao lado do determinante zero que acompanha *Npred* plural, como se demonstra em:

(8.12) *A Ana deu (três + alguns + muitos + Ø) beijos no Rui.* [Ex.C]

Os cardinais (como *dois, três etc.*), os pronomes indefinidos (como *alguns, muitos, todos etc.*) e a ausência de *Det* acompanhando *Npred* no plural são diferentes formas de indefinição do determinante, portanto, via de regra, também deveriam ser aceitos livremente nas *CVS*. O problema, no entanto, é que a aceitabilidade desses indefinidos depende da natureza (contável ou incontável) do *Npred*. Observe-se, por exemplo:

(8.13) **A Ana deu (três + algumas + muitas + Ø) organizadas no material.* [Ex.C]

A interdição dos determinantes indefinidos plurais se deve ao fato de que o *Npred organizada* é um substantivo incontável e, portanto, não pode ser admitido no plural. Aqui também se poderia recorrer à inserção de um *Modif* obrigatório. Em certos casos, os indefinidos plurais se tornam aceitáveis, quando associados a um modificador.

(8.14) *A Ana deu várias organizadas sucessivas no material.* [Ex.C]

O processo indicado pelo *Npred organizada* pode ser duplicado, triplicado, enfim, repetido ao longo do tempo, o que o torna um processo frequentativo e não pontual. Isso permite que o pronome indefinido, desde que acompanhado por um *Modif* obrigatório, seja aceitável com nomes predicativos não-contáveis, como é o caso de *organizada*.

Até o momento, tratamos das restrições sobre os determinantes apenas em construções *standard*. Em construções conversas, sabe-se, porém, que os determinantes definidos introduzindo *Npred* também são naturalmente aceitáveis, como ocorre em:

(8.15) *A Ana recebeu (uma + a) ajuda do Rui.* [Ex.C]

O *Det* definido é autorizado, nas construções conversas, porque o complemento *do Rui*, na verdade, é uma redução da oração relativa (*e.g. A Ana recebeu a ajuda que o Rui lhe deu*). Dadas as diferenças entre construções *standard* e conversas, considera-se que a distribuição dos determinantes depende, em grande medida, da combinação entre o *Vsup* e o *Npred*, e não apenas de um ou de outro.

Nesta seção, indicamos alguns dos problemas referentes ao fenômeno da determinação, mas não nos aprofundaremos nessas questões porque a determinação é um fenômeno complexo, como um predicado de segunda ordem, na medida em que os determinantes definidos são sempre uma reestruturação (ou redução) de orações relativas.

Em suma, dois fatores devem ser considerados com relação à propriedade da restrição sobre os determinantes: (i) não é possível inserir um determinante (demonstrativo, possessivo, artigo definido) que situe o nome predicativo fora da esfera de referência do sujeito da construção; e (ii) é a combinação [*Vsup Npred*] quem impõe as restrições sobre o tipo de determinante que a construção seleciona.

8.3 Descida do advérbio

A propriedade da "descida do advérbio" (GIRY-SCHNEIDER, 1987, p.31-32) se refere à possibilidade de o advérbio modificador do verbo ou do adjetivo descer para a posição de modificador do nome predicativo, sob a forma do adjetivo com a mesma raiz. Essa "descida" remete à posição hierárquica da dependência entre os elementos da frase. Numa construção com verbo pleno, o advérbio está numa relação direta com o verbo e, portanto, está no mesmo nível hierárquico de dependência. Já numa construção com verbo-suporte, em que o *Npred* (e não mais o verbo) é o núcleo da frase, o adjetivo correspondente está numa posição hierárquica de dependência inferior ao lugar estrutural ocupado pelo verbo; daí a designação "descida do advérbio".

Essa propriedade se verifica em relação aos modificadores frequentativos (*MFs*), que podem assumir a forma de advérbio com sufixo *-mente* (*e.g. frequentemente, constantemente*) ou de um adjetivo modificador do nome predicativo (*e.g. frequente, constante*). Esses *MFs* podem ser: (i) advérbios de modo, como *calmamente*; (ii) advérbios de tempo, como *imediatamente*; (iii) advérbios de frequência, como *repetidamente*; ou ainda (iv) advérbios de aspecto, como *longamente*. Para testar a propriedade da "descida do advérbio", são preferíveis os advérbios que modificam o predicado, e não a frase toda. Além disso, o advérbio e o adjetivo são morfológica, sintática e semanticamente associados, como se verifica, por exemplo, em:

(8.16) *A Ana beijou longamente o Rui.* [Ex.C]

(8.16a) ≡ *A Ana deu um beijo longo no Rui.* [Ex.C]

A equivalência semântica entre esse par de exemplos indica que a modificação exercida pelo advérbio não se altera quando se transforma o predicado verbal em predicado nominal; altera-se apenas a categoria gramatical do modificador, que deixa de ser um advérbio e passa a ser um adjetivo. A equivalência entre advérbios e adjetivos se verifica em numerosos pares de {*Adv, Adj*}, sobretudo naqueles que exprimem valor aspectual frequentativo (*frequentemente* ≡ *frequente*, *constantemente* ≡ *constante* etc.).

(8.17) *A Ana beija frequentemente o Rui.* [Ex.C]

(8.17a) ≡ *A Ana dá beijos frequentes no Rui.* [Ex.C]

Essa propriedade é definitória das *CVS* porque não é admitida em construções com verbo pleno:

(8.18) *A Ana dá livros frequentemente para o Rui.* [Ex.C]

(8.18a) **A Ana dá livros frequentes para o Rui.* [Ex.C]

A construção (8.18a) não é aceitável porque os *MFs* estão ligados a operadores, e *livros* não é um operador (nem sequer um nome eventivo), e sim, argumento do verbo pleno *dar*. O fato de que *beijos*, em (8.17a), aceita o *MF frequentes* reforça o estatuto predicativo desse nome.

Essa propriedade só pode ser atestada em construções cujo nome predicativo seja uma nominalização. Os casos de *Npred* autônomos não podem ser comparados em termos da descida do advérbio porque não existe um verbo a ser modificado pelo advérbio, como é o caso do *Npred sermão*:

(8.19) *A Ana deu um longo sermão no Rui.* [Ex.C]

(8.19a) *?*A Ana sermonou longamente o Rui.* [Ex.C]

Além disso, em alguns casos, o verbo e seu advérbio modificador formam um composto. É o caso de advérbios que expressam pontos de vista, tais como aparecem nas combinações *apoiar moralmente*, *golpear mortalmente*, *assediar sexualmente*, dentre outras. Quanto mais fixa é a combinatória entre o verbo e seu advérbio, maior será o entrosamento entre o *Npred* e o adjetivo, considerando-se como *Npred* compostos, por exemplo, *apoio moral*, *golpe mortal* e *assédio sexual*.

A partir dos dados analisados neste trabalho, conclui-se que todas as *CVS* com *dar* cujo *Npred* é uma nominalização autorizam a descida do advérbio.

8.4 A dupla análise dos complementos preposicionais

Nas *CVS* em que o nome predicativo seleciona um complemento preposicionado, esse complemento pode ser extraído de duas formas diferentes: (i) como um constituinte isolado (8.20a) ou (ii) em conjunto com o nome predicativo (8.20b):

(8.20) *A Ana deu um sermão ao Rui.* [Ex.C]

(8.20a) *Foi ao Rui que a Ana deu um sermão.* [Ex.C]

(8.20b) *Foi um sermão ao Rui que a Ana deu.* [Ex.C]

A possibilidade de extrair o complemento preposicionado de duas formas diferentes chama-se *dupla análise dos complementos preposicionais*. Note-se que, nas construções com verbo pleno, a única análise autorizada é a extração do complemento preposicionado isoladamente:

(8.21) *A Ana deu um livro ao Rui.* [Ex.C]

(8.21a) *Foi ao Rui que a Ana deu um livro.* [Ex.C]

(8.21b) *?*Foi um livro ao Rui que a Ana deu.* [Ex.C]

Nas *CVS*, o complemento preposicionado pode ser extraído como um argumento isolado ou pode formar, junto com o nome predicativo, um constituinte único. Nas construções com verbo pleno, o complemento preposicionado (dativo) não forma constituinte único com o objeto direto (neste caso, o *livro*).

Ressalte-se que essa propriedade só pode ser verificada em *CVS* que possuam complemento preposicional. As construções com apenas um argumento, na posição sujeito (*e.g. A Ana deu um espirro*), não podem ser testadas com relação a essa propriedade porque elas não possuem complemento preposicional.

8.5 Redução do *Vsup* e formação de grupo nominal

Essa propriedade consiste em transformar o predicado nominal (*CVS*) em um sintagma nominal (ou grupo nominal). Forma-se um grupo nominal, cuja cabeça é o nome predicativo, a partir da redução (ou apagamento) do verbo-suporte.

(8.22) *A Ana deu um beijo no Rui.* [Ex.C]

[*Rel*] = *O beijo que a Ana deu no Rui* <...> [Ex.C]

[*RedVsup*] = *O beijo da Ana no Rui* <...> [Ex.C]

Conforme aponta Baptista (2005b, p.26), "na medida em que o *Vsup* é apenas um auxiliar gramatical do nome predicativo, ele pode, sob certas condições, ser reduzido sem que tal acarrete perda de informação importante". Nesse caso, o verbo *dar*, por ser suporte, foi reduzido ou apagado. Em contrapartida, nos casos de construções com verbo pleno, apenas a relativização é autorizada; o apagamento do verbo para a formação de grupo nominal não o é.

(8.23) *A Ana deu um livro ao Rui.* [Ex.C]

[*Rel*] = *O livro que a Ana deu ao Rui* <...> [Ex.C]

[*RedVsup*] = **O livro da Ana ao Rui* <...> [Ex.C]

Como o verbo pleno é significativo do ponto de vista semântico, ele não pode ser apagado, o que interdita o grupo nominal **O livro da Ana ao Rui*. A relativização é um processo comum às duas construções, porém o apagamento do verbo é interditado nas construções com verbo pleno.

O grupo nominal formado a partir da redução do *Vsup* (*O beijo da Ana no Rui*) pode funcionar, assim, como argumento sujeito, como em (8.24), ou complemento de outros predicados, como em (8.25).

(8.24) *O beijo da Ana no Rui foi visto por todo mundo.* [Ex.C]

(8.25) *O Pedro não gostou do beijo da Ana no Rui.* [Ex.C]

8.6 Possibilidade de variação do *Vsup*

Os verbos-suporte, por servirem apenas para veicular marcas verbais e não serem predicadores, podem frequentemente ser substituídos por outros verbos também suporte, sem alterar o significado essencial da frase de base. *Npred*.

(8.26) *A Ana (deu + pregou + enfiou + meteu + sentou) um tapa no Rui.* [Ex.C]

A alternância entre os diferentes verbos-suporte não altera o predicado semântico de base, já que o predicado é expresso pelo *Npred* e não pelo *Vsup*. Também não altera as relações que se estabelecem entre o nome predicativo e seus argumentos nem as propriedades sintáticas e semânticas da construção. Nesse sentido, independentemente do verbo selecionado, todas essas formas frasais são consideradas igualmente frases de base.

As variantes dos verbos-suporte podem acarretar diferenças aspectuais e/ou estilísticas entre as frases, mas, de modo geral, não lhes altera o significado global. No caso de (8.26), há diferenças estilísticas entre *dar*, que é mais neutro, e as outras variantes (*pregar*, *enfiar*, *meter* e *sentar*), mas essas diferenças não são de ordem semântica.

Em geral, consideram-se verbos-suporte elementares aqueles que são mais neutros do ponto de vista aspectual e/ou estilístico, e consideram-se variantes de *Vsup* aqueles que introduzem diferentes matizes aspectuais e/ou estilísticos. As variantes de *Vsup* (por vezes, chamadas *extensões do verbo-suporte*) são muito numerosas e, por isso, difíceis de recensear. A literatura, de maneira geral, aborda apenas os verbos-suporte elementares, por considerar que são verbos "esvaziados de sentido". Neste trabalho, foram consideradas 14 variantes aspectuais/estilísticas do verbo-suporte *standard dar* (*aplicar*, *atribuir*, *bater*, *conceder*, *estar com*, *fazer*, *impor*, *ministrar*, *pedir*, *prestar*, *ser de*, *ter* e *tirar*) e outras cinco variantes aspectuais/estilísticas de verbos conversos ao *dar* (*levar*, *pegar*, *receber*, *ter* e *tomar*).

Os predicados nominais com o *Vsup fazer* e com o *Vsup ter* já foram descritos, respectivamente, por Barros (2014) e Santos (2015). Assim, muitos nomes predicativos que aceitam as variantes *fazer* e *ter* já foram incluídos naqueles trabalhos. Ainda assim, na descrição dos predicados nominais com o *Vsup dar*, esses *Npred* também são descritos, já que eles selecionam também o *Vsup dar* em frases de base. Em trabalhos futuros, pretende-se intersectar as três matrizes para produzir um único recurso léxico-sintático, excluindo os dados duplicados.

Neste capítulo, expusemos as seis principais propriedades das *CVS*. Essas propriedades podem ser atestadas e reproduzidas por meio de testes sintáticos que comparam construções com verbo pleno e construções com verbo-suporte. Essas propriedades foram sendo observadas, ao longo do tempo, por diferentes autores que analisaram *CVS* em diversos idiomas. A partir do acúmulo de evidências, Ranchhod (1990) foi quem primeiro sistematizou as propriedades anteriormente elencadas por Gross (1981), Giry-Schneider (1987) e outros autores do Léxico-Gramática.

As propriedades gerais das *CVS* apresentadas e discutidas neste capítulo são: (i) a relação particular entre o *Npred* e algum de seus argumentos; (ii) as restrições sobre os determinantes; (iii) a descida do advérbio; (iv) a dupla análise dos complementos preposicionais; (v) a redução do *Vsup* e a formação de grupo nominal; e (vi) a possibilidade de variação do *Vsup*.

A partir do próximo capítulo, trataremos especificamente das propriedades sintático semânticas das *CVS* com o verbo *dar*. Trata-se de propriedades específicas sobre o tipo de argumentos, de preposição, de determinante, os papéis semânticos dos argumentos, as transformações sintáticas admissíveis, dentre várias outras.

9 Propriedades sintático-semânticas das construções com *dar*

Apresentamos, neste capítulo, as principais propriedades formais, distribucionais e transformacionais dos nomes predicativos que se constroem com o *Vsup dar*. É preciso definir rigorosa e sistematicamente as propriedades que serão descritas para cada predicado, pois são essas propriedades que definem o tipo de construção e a classificação a ser atribuída a cada uma. Para tanto, fizemos um levantamento sistemático das propriedades observadas em outras construções com outros verbos-suporte (RANCHHOD, 1990; GIRY-SCHNEIDER, 1978; GIRY-SCHNEIDER, 1987; VAZA, 1988; CHACOTO, 2005; BAPTISTA, 2005b; BARROS, 2014) e selecionamos aquelas que se aplicam às construções com o *Vsup dar*.

As propriedades sintático-semânticas a serem descritas, neste Capítulo, são: (i) propriedades formais, como o número de argumentos, especificidades sobre a inserção de quantificadores, o tipo de determinante selecionado pelo *Npred* e o tipo de preposição que introduz os complementos essenciais; (ii) propriedades distribucionais, como o tipo semântico dos argumentos (sujeito e complemento) - humano, não-humano, parte-do-corpo, nome locativo, frase completiva ou factiva - e os papéis semânticos dos argumentos; e (iii) propriedades transformacionais, como a simetria, a conversão, a passiva e a nominalização. Por fim, explicamos também os critérios para proceder ao desdobramento lexical de algumas entradas na matriz.

Muitas das propriedades aqui analisadas são estritamente sintáticas, porém usamos a designação *propriedades sintático-semânticas* por três motivos: (i) os traços semânticos [+humano] (*Nhum*) e [-humano] (*N-hum*) que restringem o tipo de argumento são características semânticas; (ii) os papéis temáticos dos argumentos também são propriedades semânticas; e (iii) algumas propriedades transformacionais requerem uma análise, não só sintática, mas também semântica, por exemplo: algumas construções só admitem conversão com os verbos *ganhar* e *receber*, enquanto outras só admitem conversão com os verbos *levar* e *tomar*; a diferença entre essas restrições é também semântica e não apenas sintática.

9.1 Propriedades formais

9.1.1 Número de argumentos

Definir o número de argumentos que as construções com verbo-suporte possuem não é uma tarefa simples. A própria identificação de argumentos essenciais em construções verbais já apresenta uma série de dificuldades, as quais se estendem para as construções nominais.

Cançado (2009) discute as noções de *argumentos*, *complementos* e *adjuntos*, definindo *argumentos* por critérios estritamente semânticos, e os conceitos de *complementos* e *adjuntos*, por critérios sintáticos. Segundo a autora, "os argumentos são estabelecidos em uma estrutura semântico-lexical e são associados, na estrutura sintática, às posições de sujeito, de complemento e, também, de adjuntos". A argumentação da autora é bastante válida, porém, neste trabalho, adotamos uma visão bem mais estreita do conceito de *argumentos*, considerando apenas o sujeito e os complementos (se houver).

Para construir e analisar as frases elementares, adotamos o *Princípio da Máxima Projeção dos Argumentos*, que considera o maior número possível de argumentos de um predicado, desde que esses argumentos não sejam adjuntos (chamados semanticamente de circunstanciais). Quando tratamos do número de argumentos do predicado, referimo-nos apenas aos elementos obrigatórios, por exemplo, excluem-se os circunstanciais de tempo, de modo, de lugar, alguns de companhia, bem como as orações adverbiais de causa, de tempo, de condição *etc*.

Na linguagem em uso, alguns argumentos que fazem parte da estrutura argumental de um predicado encontram-se implícitos ou elípticos, pois podem ser depreendidos a partir do contexto. Assim, quando recorremos a *corpus*, um predicado com dois ou três argumentos em sua estrutura argumental, muitas vezes, aparece em textos reais com dois ou um argumento(s). No entanto, para a estrutura de base das frases simples, consideramos todos os argumentos que tal predicado pode selecionar.

Por vezes, recorremos à estrutura de base do predicado verbal equivalente ao predicado nominal e associamos a rede argumental do verbo ao nome. Numa estrutura de superfície, por exemplo, a frase nominal *Zé deu uma ordem à Ana* parece ter sua rede argumental completa, porém, se recorrermos à frase verbal correspondente, teremos *Zé ordenou à Ana que ela voltasse cedo*. Isso pode sugerir (mas não obriga) que, numa predicação de base, o nome predicativo *ordem* exige um complemento frásico correspondente ao conteúdo da ordem, além dos dois complementos humanos: um na posição de sujeito e

outro na posição de complemento. Por considerarmos o máximo possível de argumentos, nomes predicativos como *ordem* exigem três argumentos.

Vale ressaltar que não existe nenhuma obrigatoriedade, no modelo do Léxico-Gramática, para que as construções verbal e nominal equivalentes possuam exatamente a mesma estrutura argumental ou o mesmo número de argumentos, mas decidimos considerar o maior número possível de argumentos, levando em conta o *Princípio da Máxima Projeção Argumental*. Isso significa que, se determinado argumento puder aparecer em alguma frase de superfície, então ele será considerado como essencial na frase de base.

Independentemente da relação de nominalização que se possa estabelecer entre a construção verbal e a nominal, elas são independentes e podem apresentar diferenças entre si. O estudo sistemático dessas diferenças é o objeto de estudo do Léxico-Gramática. Note que pode haver diferenças meramente gramaticais, como a escolha da preposição, ou ainda diferenças de natureza distribucional.

As construções nominais com o verbo *dar* apresentam três estruturas argumentais diferentes: predicados com um, dois ou até três argumentos, como se apresenta, respectivamente, em:

(9.1) *A Ana deu uma cambalhota.* (um argumento) [Ex.C]

(9.2) *A Ana deu um abraço no Rui.* (dois argumentos) [Ex.C]

(9.3) *A Ana deu o apelido de "rei" ao Rui.* (três argumentos) [Ex.C]

Um dos tipos mais recorrentes de *CVS* com três argumentos é aquele em que o nome predicativo seleciona dois nomes humanos para as posições de *N*0 e *N*1 e seleciona uma frase completiva para a posição de *N*2, como:

(9.4) *A Ana deu a autorização ao Rui para se casar.* [Ex.C]

Nos casos em que o predicado nominal seleciona três argumentos, sendo *N*0=:*Nhum*, *N*1=:*Nhum* e *N*2=:*QueF*, muitas vezes, o argumento *N*1=:*Nhum* é correferente com o sujeito da frase completiva expressa em *N*2=:*QueF*, como ocorre na seguinte transformação.

(9.4) *A Ana deu a autorização ao Rui para se casar.* [Ex.C]

≡ *A Ana deu a autorização para o Rui se casar.* [Ex.C]

A segunda frase é uma transformação da primeira. Na primeira frase, todos os argumentos do *Npred autorização* estão explícitos e podem ser desmembrados (*N0=:Ana*, *N1=:ao Rui* e *N2=:para se casar*). Na segunda frase, a preposição *a* do primeiro complemento se funde com a preposição *para* do segundo complemento e a frase de superfície parece ter apenas dois argumentos (*N0=:Ana* e *N1=:para o Rui se casar*).

Por serem correferentes, poder-se-ia interpretá-los como um argumento único (*N1 =:QueF*) em que se fundem o *Nhum* e a completiva (*QueF*); no entanto, consideraremos essas situações como três argumentos distintos, pois: (i) eles podem ser desmembrados em dois argumentos e, pelo *Princípio da Máxima Projeção Argumental*, devemos considerar o número máximo possível de argumentos para cada predicado; e (ii) algum deles pode ser apagado na estrutura de superfície, o que comprova mais uma vez que são argumentos distintos.

Ressalte-se também que, quando o *Npred* exige ou aceita um elemento restritivo quantificador (*Quantif*), este elemento não é considerado obrigatório, portanto não é argumento, e sim adjunto. Os quantificadores aparecem nas frases de base na matriz (ver Apêndice A) dentro de colchetes angulares (<>) para indicar que não são argumentos obrigatórios nas *CVS*, mas tornam a frase de base mais natural. Há diferentes tipos de *Quantif*: (i) valor monetário (9.5); (ii) percentagem (9.6); e (iii) valor temporal (9.7), ou outros.

(9.5) *A loja deu um desconto <de R$ 3,00> no valor do produto.* [Ex.C]

(9.6) *A loja deu uma promoção <de 50%> em todo o estoque.* [Ex.C]

(9.7) *Ana deu um prazo <de 3 meses> para o Rui se decidir.* [Ex.C]

Uma análise mais profunda sobre os casos de *Npred* que aceitam a inserção de quantificadores será feita na subseção 9.1.2.

9.1.2 Inserção de expressões de quantidade

Os nomes predicativos que expressam grandezas ou unidades de medida podem ser associados a expressões de quantidade (*Quantif*), em construções com verbo-suporte (BAPTISTA; RANCHHOD, 1998), como ocorre em:

(9.8) *A água está com uma temperatura de 30°C.* [Ex.C]

(9.9) *O avião está a uma altitude de 10.000 pés.* [Ex.C]

(9.10) *Essa piscina tem uma profundidade de 3m.* [Ex.C]

A forma *Npred prep Quantif* – instanciada nos exemplos acima, respectivamente, por *temperatura de 30°C*, *altitude de 10.000 pés* e *profundidade de 3m* – pode ser reestruturada internamente e adquirir a forma de *Quantif de Npred*, indicada nos exemplos a seguir.

(9.8a) *A água está com 30°C de temperatura.* [Ex.C]

(9.9a) *O avião está a 10.000 pés de altitude.* [Ex.C]

(9.10a) *Essa piscina tem 3m de profundidade.* [Ex.C]

Ocorre uma situação semelhante com alguns nomes predicativos associados ao verbo *dar*, como, por exemplo, nas construções *dar uma entrada* (9.11) e *dar um sinal* (9.12)[42].

(9.11) *Zé deu uma entrada de R$50.000,00 pela casa.* [Ex.C]

= *Zé deu R$50.000,00 de entrada pela casa.* [Ex.C]

(9.12) *Zé deu um sinal de R$50,00 pelo produto.* [Ex.C]

= *Zé deu R$50,00 de sinal pelo produto.* [Ex.C]

A questão que se levanta é: essas construções devem ser consideradas construções com verbo-suporte ou construções com verbo pleno cujo objeto direto é a expressão de quantidade e o complemento preposicionado (*de entrada* e *de sinal*) é uma locução adverbial?

Analisaremos cinco propriedades sintáticas listadas por Borillo (1985) e por Baptista e Ranchhod (1998) para determinar o estatuto dessas construções: (i) possibilidade de formação de *GN*, tais como as outras *CVS*; (ii) possibilidade de redução do elemento quantificador; (iii) possibilidade de redução do *Npred*; (iv) aceitabilidade de frases interrogativas substituindo o *Quantif* ou o *Npred*; e (v) mobilidade da posição sintática do *Npred*.

A formação de grupo nominal complexo (*GN*) é uma propriedade geral das construções com verbo-suporte e que se verifica também nas construções com *Quantif de Npred*. A transformação da frase para a constituição do grupo nominal complexo passa por duas etapas: (i) a relativização da frase de base (com a inserção de uma oração principal); e (ii) o apagamento do verbo-suporte *dar*.

42 Ressalte-se que há um desdobramento lexical dos nomes *entrada* e *sinal*, por constituírem construções sintáticas diferentes desta. Não estamos tratando aqui das construções *dar uma entrada* ≡ *entrar em Nloc* ou *dar um sinal* ≡ *fazer um sinal/gesto/aceno*, mas sim de construções em que *entrada* e *sinal* indicam parte do pagamento de um valor total, de forma que possam ser restringidas por um quantificador.

(9.11) *Zé deu uma entrada de R$50.000,00 pela casa.* [Ex.C]

[*Rel*] = *A entrada de R$50.000,00 que Zé deu não foi suficiente.* [Ex.C]

[*RedVsup*] = *A entrada de R$50.000,00 do Zé não foi suficiente.* [Ex.C]

(9.12) *Zé deu um sinal de R$50,00 pelo produto.* [Ex.C]

[*Rel*] = *O sinal de R$50,00 que Zé deu foi abatido no preço do produto.* [Ex.C]

[*RedVsup*] = *O sinal de R$50,00 do Zé foi abatido no preço do produto* [Ex.C]

(9.12) *Zé deu uma nota de R$5,00 para Ana.* [Ex.C]

[*Rel*] = *A nota de R$5,00 que o Zé deu para a Ana foi gasta adequadamente.* [Ex.C]

[*RedVsup*] = *?*A nota de R$5,00 do Zé para a Ana foi gasta adequadamente.* [Ex.C]

Como se nota, as construções (9.11) e (9.12), por serem *CVS*, autorizam a redução do *Vsup* e a formação do grupo nominal complexo, ao passo que (9.13), por ser uma construção com verbo pleno, não autoriza essa propriedade.

A segunda e a terceira propriedades dizem respeito à possibilidade de redução (elipse) seja do *Npred* seja do *Quantif*. Com os verbos-suporte *ter* e *estar prep*, associados a grandezas ou unidades de medida, o quantificador do nome predicativo não pode ser apagado.

(9.14) *O meu jardim tem uma área de 200m*^{2}. [Ex.C]

**O meu jardim tem uma área.* [Ex.C]

(9.15) *Esse gás está à pressão de 300 bares.* [Ex.C]

**Esse gás está à pressão.* [Ex.C]

No entanto, isso pode ocorrer com os quantificadores dos nomes associados ao verbo *dar*, como se verifica em:

(9.11) *Zé deu uma entrada de R$50.000,00 pela casa.* [Ex.C]

= *Zé deu uma entrada pela casa.* [Ex.C]

(9.12) *Zé deu um sinal de R$50,00 pelo produto.* [Ex.C]

= *Zé deu um sinal pelo produto.* [Ex.C]

Com relação à redução do *Npred*, ocorre o oposto: as construções com os verbos-suporte *ter* e *estar prep* permitem o apagamento do *Npred* enquanto as construções com o *Vsup dar* não o permitem.

(9.16) *Esta árvore tem 2m de altura.* [Ex.C]

(9.16a) = *Esta árvore tem 2m.* [Ex.C]

(9.11) *Zé deu uma entrada de R$50.000,00 pela casa.* [Ex.C]

(9.11a) =/Zé deu R$50.000,00 pela casa. [Ex.C]

(9.12) *Zé deu um sinal de R$50,00 pelo produto.* [Ex.C]

(9.12a) =/Zé deu R$50,00 pelo produto. [Ex.C]

As frases (9.11a) e (9.12a) não são consideradas agramaticais; pelo contrário, são naturalmente aceitáveis em Língua Portuguesa, porém elas não são semanticamente equivalentes às frases com o quantificador, pois a principal noção partitiva (de *entrada* ou de *sinal*) se perde. A interpretação de (9.11) e (9.12) implica que o *Quantif* (*R$50.000,00* e *R$50,00*) corresponde a apenas uma parte do valor; já a interpretação de (9.11a) e (9.12a) sugere que o *Quantif* seja o valor total pago pelas mercadorias.

A quarta propriedade sintática concerne à possibilidade de construção de frases interrogativas cuja resposta seja a frase transformada com o *Quantif de Npred*. Tanto as *CVS* quanto as construções com verbo pleno respondem adequadamente às perguntas em *De quanto?*, referindo-se ao quantificador:

(9.17) *De quanto é a nota que Rui deu para Ana?* R: *De R$5,00.* [Ex.C]

(9.18) *De quanto é a entrada que Zé deu pela casa?* R: *De R$50.000,00.* [Ex.C]

(9.19) *De quanto é o sinal que Zé deu pelo produto?* R: *De R$50,00* [Ex.C]

Quando a pergunta recai sobre o nome, em vez de recair sobre o *Quantif*, há uma diferença entre as *CVS* e as construções com verbo pleno: as construções com verbo pleno respondem satisfatoriamente às perguntas em *O quê?*; já as *CVS* não respondem satisfatoriamente, como demonstram os exemplos a seguir:

(9.20) *O que Rui deu para Ana?* R: *Uma nota de R$5,00.* [Ex.C]

(9.21) *?*O que Rui deu pela casa?* R: *Uma entrada de R$50.000,00.* [Ex.C]

(9.22) *?*O que Rui deu pelo produto?* R: *Um sinal de R$50,00* [Ex.C]

Por fim, a última propriedade identificada por Baptista e Ranchhod (1998) que se aplica às *CVS* com o verbo *ter* é a mobilidade do constituinte *Npred* dentro da frase com quantificador. A mesma propriedade se verifica nas construções com o verbo-suporte *dar*.

(9.23) *O meu jardim tem 200m*^{2} ***de área***. [Ex.C]

De área, *o meu jardim tem 200m*^{2}. [Ex.C]

O meu jardim, ***de área****, tem 200m*^{2}. [Ex.C]

O meu jardim tem, ***de área****, 200m*^{2}. [Ex.C]

(9.24) *Zé deu R$50.000,00* ***de entrada pela casa***. [Ex.C]

De entrada pela casa, *Zé deu R$50.000,00.* [Ex.C]

Zé, ***de entrada pela casa****, deu R$50.000,00.* [Ex.C]

Zé deu, ***de entrada pela casa****, R$50.000,00.* [Ex.C]

(9.25) *Zé deu R$50,00* ***de sinal pelo produto***. [Ex.C]

De sinal pelo produto, *Zé deu R$50,00.* [Ex.C]

Zé, ***de sinal pelo produto****, deu R$50,00.* [Ex.C]

Zé deu, ***de sinal pelo produto****, R$50,00.* [Ex.C]

Com base na aplicação dessas propriedades, consideraremos que nomes do tipo *entrada* e *sinal*, que podem ser restringidos por elementos quantificadores, são nomes predicativos e, portanto, suas construções devem ser tratadas como *CVS*. Os demais candidatos a nomes predicativos que autorizam a inserção de quantificadores, mas não atenderam a esses testes, foram excluídos do rol de *Npred*.

9.1.3 Determinantes

Os determinantes das *CVS* com o verbo *dar*, em geral, sofrem fracas restrições de seleção. Como regra, os determinantes selecionados pelos *Npred* são livres, podendo

admitir artigo definido, indefinido, pronome possessivo ou até a ausência de determinantes. Alguns *Npred*, no entanto, tais como as nominalizações em *-ada* dão preferência para o artigo indefinido (*deu uma sapatada, deu uma joelhada, deu uma caminhada*), mas isso não significa que o determinante seja fixo.

Por outro lado, há casos em que o determinante selecionado pelo *Npred* é fixo, ou seja, não é distribucionalmente livre; nesses casos, verifica-se ora o artigo definido ora a ausência de determinante (Ø).

A única diferença sintática que se verifica entre (9.26) e (9.27) diz respeito às restrições sobre o determinante:

(9.26) *A Ana deu uma entrada no hospital.* [Ex.C]

(9.27) *A Ana deu entrada no hospital.* [Ex.C]

A construção (9.26) indica que a "Ana entrou rapidamente no hospital para fazer uma visita a um paciente", por exemplo; já a construção (9.27) indica que "ela foi encaminhada para o hospital e é, necessariamente, a paciente". Nesse caso, considera-se que há dois nomes predicativos diferentes: *entrada1* e *entrada2*, que são homógrafos. As duas interpretações só podem ser depreendidas pela análise do determinante, que é livre em (9.26) e fixo em (9.27). Outra prova de que se trata de duas construções diferentes é o fato de que o nome *entrada*, em (9.26), pode se apresentar no diminutivo e ser acompanhado de modificador (*A Ana deu uma entradinha rápida no hospital*), o que não acontece com a construção expressa em (9.27).

Nesses casos, as duas construções são classificadas como *CVS*, mas indicando-se duas entradas lexicais para o nome *entrada*. O fato de que (9.27) exige determinante fixo (Ø) poderia nos conduzir a classificá-lo incorretamente como expressão cristalizada. Ressalte-se, porém, que essa construção não é uma *EC* porque não se verifica o caráter metafórico do nome *entrada*. Levando-se em conta a chave dicotômica apresentada na Figura 4, seria necessário que o determinante fosse fixo e o nome *entrada* tivesse sentido figurado para que a construção fosse classificada como *EC*. Como a interpretação figurada não se verifica em (9.27), então a construção é classificada como *CVS*.

Os dois exemplos apresentados dos *Npred entrada* distinguem entre duas *CVS* diferentes. Por outro lado, as restrições sobre os determinantes dos nomes predicativos também servem para distinguir entre *CVS* e construções fixas:

(9.28) *A Ana deu (um + o + Ø) fora no Rui.* [Ex.C]

(9.29) *A Ana deu (*um + o + *Ø) fora da sala.* [Ex.C]

O *Npred fora*, em (9.28), admite livremente qualquer determinante (definido, indefinido ou Ø), ao passo que a expressão fixa *dar o fora de Nloc*, em (9.29), exige o determinante definido *o*. Essa característica também se verifica em outras línguas. Em PE, por exemplo, há uma construção parecida com o nome predicativo *salto*:

(9.30) *O Rui deu (um + alguns + três + o maior) salto(s) (bem alto).* [Ex.C]

(9.31) *O Rui deu (*um + o + *Ø) salto para o Brasil.* [Ex.C]

Em termos de significado, (9.30) significa literalmente "saltar, pular", ao passo que (9.31) significa "ir para *Nloc*, mudar-se definitivamente para *Nloc*". Em termos de sintaxe, a diferença entre (9.30) e (9.31), assim como entre (9.28) e (9.29) se revela pelas restrições sobre a seleção do determinante, que é livre em (9.28) e (9.30) e é fixo em (9.29) e (9.31), além das diferenças na distribuição argumental.

Apenas 15 construções com o *Vsup dar* exigem determinante fixo (Ø), tais como:

(9.32) *O médico deu (Ø + *uma + *sua + *alguma) alta ao paciente.* [Ex.C]

(9.33) *O Rui deu (Ø + *uma + *a + *alguma) ciência no documento.*[43] [Ex.C]

Nas construções com verbo-suporte, os determinantes selecionados pelo *Npred* são, geralmente, livres. Os raros casos de determinantes fixos foram assinalados na matriz (ver Apêndice A). Assim, marcamos, não apenas a obrigatoriedade do determinante fixo (dado pela etiqueta *Det=fixo*), mas também as possibilidades de ocorrência dos determinantes livres, a partir da verificação em *corpus*. A informação de determinante obrigatório só aparece na marcação da coluna *Det=fixo*; se ele é fixo, significa que é obrigatório. Consideramos os seguintes tipos de determinantes: determinante zero (*Det=:E*), artigo definido (*Det=:Def*), artigo definido seguido de modificador (*Det=:Def+Mod*), artigo indefinido (*Det=:Indef*), artigo indefinido seguido de modificador (*Det=:Indef+Mod*) e pronome possessivo (*Det=:Poss0*).

No Léxico-Gramática, analisamos apenas os determinantes que são selecionados pelo nome predicativo; não fizemos análise dos determinantes que acompanham os argumentos. Na variante brasileira, sabe-se que os nomes próprios podem ser ou não acompanhados de artigos definidos, dependendo do dialeto. Não analisaremos os

43 Ressalte-se novamente que a inaceitabilidade é verificada no âmbito da frase simples. Existem casos em que o determinante indefinido introduz os *Npred alta* e *ciência*, mas desde que acompanhado de *Modif* (*e.g. O médico não me deu uma alta propriamente dita* [Ex.R] ou *Finalmente o médico deu uma alta provisória pra ela* [Ex.R]). O possessivo *sua* também poderia introduzir o nome predicativo *alta*, mas não é correferente ao sujeito (*e.g. O médico já deu sua alta, espero que fique boa logo, querida. - disse Esme* [Ex.R]). Esses determinantes são considerados inaceitáveis porque essas construções não são imediatamente interpretáveis, mas, em outros contextos, esses determinantes poderiam ser aceitáveis.

determinantes que introduzem os argumentos, haja vista que isso não é propriedade específica das *CVS*, mas funciona da mesma forma em predicados verbais e predicados nominais.

9.1.4 Preposições

Para a descrição das *CVS* com o *Vsup dar* em PB, foram consideradas as seguintes preposições: *a*, *até*, *com*, *contra*, *de*, *em*, *para*, *por* e *sobre*. Cada tipo de preposição entra em um tipo particular de estrutura sintático-semântica de *CVS*. A partir da combinação entre o tipo semântico dos argumentos, seus papéis semânticos e o número de argumentos, identificaram-se alguns padrões para a preferência de certas preposições. A Tabela 4 mostra a distribuição dos padrões sintático-semânticos dos dados, de acordo com cada tipo de preposição.

O número de *CVS* que pertence a cada padrão não corresponde ao número total de *CVS* descritas neste livro porque há 488 *CVS* com apenas um argumento e, portanto, não possuem preposição em sua estrutura. De outras 77 *CVS* da matriz não foi possível depreender qualquer padrão, levando-se em conta a combinação dos papéis semânticos dos argumentos, a estrutura sintática da construção e o tipo de preposição selecionada.

Estrutura	**nº**
*N*0(AGENT-GEN) *V Npred* ***em*** *N*1(OBJECT-GEN) **Ex.:** *Ana deu uma varrida na casa*	398
*N*0(AGENT-X) *V Npred (**a+para**) N*1(PATIENT) **Ex.:** *Rui deu retaguarda à Ana*	131
*N*0(AGENT-X) *V Npred* ***em*** *N*1(PATIENT) **Ex.:** *Rui deu um tombo na Ana*	101
*N*0(AGENT-CAUSE) *V Npred* ***em*** *N*1(OBJECT-GEN) **Ex.:** *Ana deu uma variada no figurino*	55
*N*0(AGENT-SPEAKER) *V Npred* ***em*** *N*1(ADDRESSEE) **Ex.:** *Rui deu um sermão na Ana*	52
*N*0(EXPERIENCER-VOL) *V Npred (**a+para**) N*1(PATIENT) ***Prep*** *N*2(OBJECT-F) **Ex.:** *Rui deu um estímulo para Ana se concentrar*	47
*N*0(AGENT-GEN) *V Npred* ***Preploc*** *N*1(LOCATIVE-X) **Ex.:** *Rui deu uma ida à praia*	39
*N*0(AGENT-SPEAKER) *V Npred (**a+para**) N*1(ADDRESSEE) ***Prep*** *N*2(MESSAGE) **Ex.:** *Rui deu um aviso à Ana para fechar a loja*	30

*N*0(AGENT-SPEAKER) *V Npred* (***a+para***) *N*1(ADDRESSEE) (***de+sobre***) *N*2(TOPIC) **Ex.:** *Rui deu a notícia da inflação à Ana*	21
*N*0(AGENT-GEN) *V Npred* (***Prep***) *N*1(PATIENT) (***com+de***) *N*2(OBJECT-GEN) **Ex.:** *Rui deu uma borrifada de água na planta*	12
*N*0(EXPERIENCER-VOL) *V Npred* (***a+para***) *N*1(PATIENT) ***de*** *N*2(TAG) **Ex.:** *Rui deu o apelido de "rei" ao Zé*	12
*N*0(AGENT-GEN) *V Npred* ***em*** *N*1(OBJECT-CL) **Ex.:** *Ana deu uma depilada na perna*	12
*N*0(OBJECT-GEN) *V Npred* ***em*** *N*1(OBJECT-GEN) **Ex.:** *A esponja deu uma absorvida na água*	9
*N*0(AGENT-GEN) *V Npred* ***com*** *N*1(CO-AGENT) *N*0(OBJECT-GEN) *V Npred* ***com*** *N*1(CO-OBJECT) **Ex.:** *Rui deu uma trepada com a Ana*	8
TOTAL	927

Tabela 4 – Padrões sintático-semântico das *CVS*, com base nas preposições
Fonte: Elaboração própria

Como se nota pela Tabela 4, a grande maioria das construções com o *Vsup dar* exige a preposição *em*, em predicados com dois argumentos. Essa é uma das principais diferenças entre PB e PE, em relação às *CVS* com *dar*. Em PE, a preposição característica dessas construções é *a*, como será demonstrado na subseção 9.1.4.3.

Por ordem de frequência, o segundo tipo de preposição mais recorrente nas *CVS* com *dar* é a preposição *a* ou *para*, introduzindo complementos dativos. A análise particular de cada tipo de preposição será feita nas próximas subseções.

9.1.4.1 Especificidades das preposições *a* e *para*

Quando introduzem um complemento dativo, as preposições *a* e *para* podem comutar entre si, como se demonstra em:

(9.34) *O jogador deu um autógrafo (à + para a) criança.* [Ex.C]

Quando o complemento é dativo e introduzido por *a* ou *para*, em geral, a construção admite também a transformação do complemento dativo *Prep N* em um pronome oblíquo (*-lhe*), a que chamamos *pronominalização* ou *cliticização*:

(9.34) *O jogador deu um autógrafo (à + para a) criança.* [Ex.C]

≡ *O jogador deu-lhe um autógrafo.* [Ex.C]

Nas *CVS* com outros verbos-suporte, muitas vezes, o complemento introduzido pela preposição *a* não é um complemento dativo, como ocorre com a *CVS fazer uma viagem* (BARROS, 2014).

(9.35) *O Zé fez uma viagem ao campo.* [Ex.C]

**O Zé fez-lhe uma viagem.* [Ex.C]

Quando o complemento introduzido por *a* ou *para* não é um complemento dativo, ele não aceita a pronominalização. Nas *CVS* com o verbo *dar* foram identificados apenas três casos em que o complemento introduzido por *a* ou *para* não é dativo. São eles:

(9.36) *A trinca formada por Clooney, Pitt e Matt Damon também deu uma escapada para Mônaco.* [Ex.R]

(9.37) *Na hora da exibição do primeiro capítulo, o ator deu uma escapulida para o outro cômodo da festa.* [Ex.R]

(9.38) *A nossa apresentadora deu uma fugida para o lugar onde ficam os comentaristas do carnaval.* [Ex.R]

As três construções exigem pelo menos um complemento locativo, o qual pode ser expresso pelo locativo de origem (introduzido pela preposição *de*) ou pelo locativo de destino (introduzido pela preposição *para*), como ocorre nos exemplos (9.36), (9.37) e (9.38).

Para finalizar a análise das preposições *a* e *para*, considerem-se os casos de construções com três argumentos em que o primeiro complemento (ou segundo argumento) é dativo e o segundo complemento (ou terceiro argumento) é uma frase completiva:

(9.39) *A máfia deu uma determinação aos civis para fecharem o comércio.* [Ex.C]

Essa frase é do tipo *Onnn*, selecionando: (i) *a máfia* para a posição de *N*0, (ii) *os civis*, para a posição de *N*1, introduzido pela preposição *a*, e (iii) *fecharem o comércio*, para a posição de *N*2, introduzido pela preposição *para*. Conforme foi apontado anteriormente, na subseção 9.1.1, os argumentos *N*1 e *N*2 podem se fundir em um só, já que *N*1 é correferente ao sujeito de *N*2, porém, levando-se em conta o Princípio da Máxima Projeção Argumental, considerar-se-ão dois complementos diferentes.

Os complementos dativos e locativos podem ser introduzidos tanto pela preposição *a* quanto por *para*, mas os complementos frásicos (*QueF*) só podem ser introduzidos pela preposição *para*, nas *CVS* com *dar* que tenham três argumentos.

9.1.4.2 Especificidades das preposições *com* e *contra*

Inicialmente é preciso distinguir entre os casos em que a preposição *com* introduz um complemento obrigatório e os casos em que introduz um complemento comitativo. Os complementos comitativos (ou de companhia) não são considerados complementos essenciais da construção e, portanto, não fazem parte da frase de base. Barros (2014, p.70) sugere dois testes para identificar um complemento comitativo: (i) se a frase puder ser parafraseada por outras frases em que o sujeito e o complemento introduzido por *com* apareçam colapsados numa forma de sujeito plural (*e.g. Zé fez um passeio com Ana* ≡ *Zé e Ana fizeram um passeio* ≡ *Eles fizeram um passeio*); e (ii) se a preposição *com* puder ser substituída pela expressão *na companhia de*.

Consideramos que o primeiro critério não seja válido porque não distingue complementos comitativos e complementos obrigatórios que são simétricos, conforme a própria autora (BARROS, 2014) já havia notado. No caso das *CVS* com o *Vsup dar*, identificaram-se cinco *Npred* que exigem um complemento essencial introduzido por *com*, o qual pode aparecer colapsado com o sujeito numa forma plural de sujeito, mas ainda assim é considerado um complemento obrigatório.

(9.40) *O depoente ainda deu uma discutida com o motorista do Vectra.* [Ex.R]

≡ *O depoente e o motorista do Vectra ainda deram uma discutida.* [Ex.C]

(9.41) *Encontrei uma amiga que me contou um fuxico de um vereador do sertão, segundo ela, deu uma namorada com um vereador do sertão.* [Ex.R]

≡ *Ela e o vereador do sertão deram uma namorada.* [Ex.C]

(9.42) *Ele deu uma rapidinha com sua melhor amiga.* [Ex.R]

≡ *Ele e sua melhor amiga deram uma rapidinha.* [Ex.C]

(9.43) *Gordinha deu uma trepada com amigão.* [Ex.R]

≡ *Gordinha e amigão deram uma trepada.* [Ex.C]

(9.44) *O batman já deu uma com a mulher gato.* [Ex.R]

≡ *O batman e a mulher gato já deram uma.* [Ex.C]

Em todos esses casos, o complemento introduzido por *com* admite a simetria e, por isso, pode ser colapsado com o sujeito para formar sujeito composto, portanto, neste

trabalho, o primeiro critério apontado por Barros (2014) não será utilizado, já que não se trata de uma propriedade definitória dos complementos comitativos.

Já o segundo critério (se puder substituir a preposição *com* pela expressão *na companhia de*) será utilizado para identificar os complementos comitativos e excluí-los das construções de base. Adote-se esse critério para analisar as seguintes frases:

(9.45) *O Rui deu uma discutida com a Ana.* [Ex.C]

=/ O Rui deu uma discutida na companhia da Ana. [Ex.C]

(9.46) *O Rui deu um passeio com a Ana.* [Ex.C]

≡ *O Rui deu um passeio na companhia da Ana.* [Ex.C]

Na construção (9.45), o complemento *com a Ana* não pode ser substituído por *na companhia da Ana*, pois alteraria o sentido da construção, portanto esse complemento é considerado obrigatório em (9.45), mas é um complemento comitativo não-obrigatório em (9.46), já que a preposição *com* pode ser substituída pela expressão *na companhia de* sem alterar o significado da construção.

Definidos os argumentos obrigatórios introduzidos por *com*, passemos à análise de alguns casos particulares. Foram identificadas quatro construções que selecionam, a rigor, a preposição *em*, mas podem alternativamente aceitar a preposição *com*, introduzindo o segundo argumento. São elas[44]:

(9.47) *Ferrón, então, deu uma peitada (em + com) Yotún, que caiu no gramado.* [Ex.R]

(9.48) *O atacante Paolo Guerrero deu uma trombada (no + com o) árbitro.* [Ex.R]

(9.49) *Não diria tombo, mas o Tarantino deu uma topada (em + com) Kill Bill.* [Ex.R]

(9.50) *O programa CQC da BAND deu uma sacaneada (em + com) Marcelinho Carioca.* [Ex.R]

No mesmo sentido, a preposição *contra* também pode comutar com a preposição *em*[45], em todos os casos em que *contra* pode ser selecionada:

(9.51) *Um dos criminosos deu um disparo (contra a + na) parede.* [Ex.R]

44 Estes exemplos foram extraídos da *web* originalmente com a preposição *em*, mas consideramos igualmente aceitáveis essas construções com a preposição *com*, por isso indicamos a comutação entre *em* e *com* através da notação "*(em + com)*".

45 Estes exemplos também foram extraídos da *web* com a preposição *contra*, mas consideramos igualmente aceitáveis as construções com *em*.

(9.52) *Damas deu um tiro (contra os + nos) policiais, mas não acertou.* [Ex.R]

Conforme aponta Chacoto (2005), somente os nomes com polaridade negativa (carga semântica negativa) aceitam a preposição *contra*, como ocorre em (9.51) e (9.52).

9.1.4.3 Especificidades da preposição *em*

Conforme atestado na Tabela 4, a maior parte das *CVS* com *dar* em PB exige a preposição *em*. A estrutura sintática básica dessas construções é [*N*0 *V Npred em N*1], em que o *N*0 é instanciado predominantemente por um AGENT-GEN e a posição de *N*1 pode ser instanciada por um PATIENT ou um OBJECT-GEN, na maioria dos casos.

Grande parte dos nomes terminados em *-ada*, em PB, selecionam a preposição *em*. Em trabalho anterior (RASSI *et al.*, 2015), analisamos as *CVS* com *dar* que fazem conversão com *levar*. Naquele estudo, identificamos que a preposição típica das construções das classes *dar-levar* em PB é *em*. Em poucas construções, admite-se alternativamente a preposição *a*.

(9.53) *O Rui deu um sermão (na + à) Ana.* [Ex.C]

(9.54) *O Rui deu uma surra (no + ao) João.* [Ex.C]

(9.55) *O Rui deu um beijo (na + à) Ana.* [Ex.C]

As construções mais naturais e com maior frequência no *corpus* se fazem, de fato, com a preposição *em*.

Por outro lado, em PE, a preposição típica das *CVS* que fazem conversão com *levar* é *a*. Apenas nos nomes da classe DL33 (ver Rassi *et al.* (2015)), que se constroem com um *Npc* na posição de *N*1, admite-se a preposição *em*. Nessa classe, as preposições *a* e *em* não estão em alternância, aceitando-se apenas *em*.

(9.56) *O Rui deu um tapa (na + *à) perna da Ana.* [Ex.C]

(9.57) *O Rui deu um tabefe (no + *ao) rosto da Ana.* [Ex.C]

A permuta da preposição exclusivamente na classe DL33 decorre de uma reestruturação do complemento parte do corpo, ou seja, nas construções em que *N*1 pode ser preenchido por *Npc*, a preposição típica é *em* por ser o resultado da reestruturação da forma [à Nhum em Npc]. Apresentam-se a seguir dois exemplos de reestruturação do *Npc*:

(9.58) *O Rui deu um tapa (*na + à) Ana na perna.* [Ex.C]

≡ *O Rui deu um tapa (na + *à) perna da Ana.* [Ex.C]

(9.59) *O Rui deu um tabefe (*na + à) Ana no rosto.* [Ex.C]

≡ *O Rui deu um tabefe (no + *ao) rosto da Ana.* [Ex.C]

Quando o complemento *N*1 é um nome não humano (*N-hum*), ocorre a alternância entre as preposições *a* e *em*. No entanto, quando o complemento *N*1 é obrigatoriamente do tipo humano (*Nhum*), admite-se exclusivamente a preposição *a* em PE e prioritariamente a preposição *em* em PB.

9.1.4.4 Especificidades da preposição *de*

A preposição *de* introduz complementos dos *Npred* com *dar* em quatro situações bem distintas: (i) quando introduz complemento locativo de origem (*e.g. A Ana deu uma fugida de casa*); (ii) quando introduz o segundo complemento, que geralmente é um INSTRUMENT, e comuta com a preposição *com* (*e.g. O Rui deu uma demão de tinta na parede*); (iii) quando introduz o complemento identificado como TAG (*e.g. O Rui deu o apelido de "Xuxa" à Ana*); e (iv) quando introduz o complemento MESSAGE, que é geralmente uma frase completiva (*QueF*) e pode comutar com a preposição *sobre* (*e.g. O Rui deu à Ana o anúncio de que fora despejado*).

A primeira situação abarca apenas sete casos, mas não será analisada nesta seção. Apesar de o complemento locativo ser introduzido pela preposição *de*, todos os casos de preposições locativas (*de, para, por* e *em*) serão retomados na subseção 9.1.4.6.

A segunda situação, em que a *Prep de* introduz o complemento INSTRUMENT, abarca dez *Npred* (*e.g. borrifada*, *demão* e *esguicho*). Esses nomes predicativos exigem três argumentos: AGENT-GEN, na posição de *N*0, PATIENT, na posição de *N*1 e INSTRUMENT, na posição de *N*2. Essa estrutura sintática se refere à frase de base. O argumento INSTRUMENT se refere a um objeto concreto utilizado para executar a ação, por exemplo, o *perfume*, a *tinta* e a água, respectivamente para *borrifada*, *demão* e *esguicho*.

Nas frases de superfície, é possível que os complementos *N*1 e *N*2 mudem de posição relativa, sem que isso altere o significado da construção.

(9.60) *A Ana deu uma borrifada de perfume no pescoço do Rui.* [Ex.C]

≡ *A Ana deu uma borrifada no pescoço do Rui com perfume.* [Ex.C]

Além de permitir a alternância na ordem dos constituintes, esses *Npred* também autorizam a comutação entre a preposição *de* e a preposição *com*, conforme se verifica em (9.60).

A terceira situação, em que a preposição *de* introduz o complemento identificado como TAG, é bastante controlada linguisticamente. Essas construções também exigem três argumentos: AGENT-GEN na posição de *N*0, PATIENT na posição de *N*1 e TAG na posição de *N*2. Conforme será explicado na subseção 9.2.5, o papel semântico TAG foi criado para se referir a argumentos que funcionam como uma etiqueta, um rótulo atribuído por alguém (*N*0 - AGENT-GEN) a alguém ou alguma coisa (*N*1 - PATIENT). Trata-se de construções como:

(9.61) *Poucos sabem, mas foi a amiga Elizabeth Taylor que deu o título de Rei do Pop a Michael.* [Ex.R]

Nessa construção, a expressão *Rei do Pop* é identificada como uma TAG atribuída a *Michael* por *Elizabeth Taylor*. O mesmo ocorre com outros *Npred*, tais como: *nome, sobrenome, rótulo, alcunha, apelido, designação etc.*

Como regra geral, ressalte-se que todo complemento de *Npred* que seleciona o *Vsup dar* é introduzido por preposição. A única exceção a essa regra, levando-se em conta os casos analisados, refere-se aos complementos TAG, que podem ser introduzidos pela *Prep de* ou essa preposição pode estar apagada, como se demonstra em:

(9.62) *Talvez para que José não parecesse tão estrangeiro aos egípcios, o Faraó lhe deu o nome Zafenate-Panéia* [Ex.R]

= *Talvez para que José não parecesse tão estrangeiro aos egípcios, o Faraó lhe deu o nome de Zafenate-Panéia* [Ex.C]

O argumento *Zafenate-Panéia*, que é anotado com o papel semântico TAG, pode ser introduzido pela preposição *de* ou sem preposição.

Por fim, a quarta e última situação engloba outras 21 *CVS*, que são constituídas por três argumentos: AGENT-SPEAKER na posição de *N*0, ADDRESSEE na posição de *N*1 e TOPIC na posição de *N*2.

(9.63) *Parabéns a torcida que deu uma demonstração a esta merda de diretoria de que quando ele mostra serviço todos comparecem e prestigiam.* [Ex.R]

Esse exemplo, retirado da *web*, apresenta todos os argumentos essenciais da construção: *N0=:a torcida*, *N1=:a esta merda de diretoria* e *N2=:de que [...] todos comparecem*. Nem sempre as frases de base apresentam todos os argumentos essenciais explícitos, mas nesse caso todos eles aparecem. Essas construções aceitam ainda a comutação com a preposição *sobre*. A diferença entre as preposições *de* e *sobre*, que introduzem *N2* é: quando a frase *QueF* é introduzida pela *Prep de*, é uma oração finita, com verbo conjugado; e quando a frase *QueF* é introduzida pela preposição *sobre*, a oração aparece no infinitivo ou outra forma reduzida não conjugada.

9.1.4.5 Especificidades da preposição *sobre*

Todos os exemplos mencionados no *caput* da subseção 9.1.4 ilustram casos de construções com dois argumentos. A preposição *sobre* só ocorre em construções com três argumentos, introduzindo o terceiro argumento (ou segundo complemento), conforme se verifica em:

(9.64) *O Rui deu sua opinião à Ana* ***sobre*** *os protestos.* [Ex.C]

Há 30 *Npred* que selecionam a preposição *sobre* introduzindo o segundo complemento, tais como, *comentário, declaração, anúncio, esclarecimento, notícia, opinião, palpite*, entre outros. Em muitas dessas construções, a preposição *sobre* pode comutar com as preposições *acerca de* ou *de*, na mesma posição sintática, como se demonstra em:

(9.65) *O Rui deu uma declaração à imprensa sobre o seu casamento.* [Ex.C]

= *O Rui deu uma declaração à imprensa acerca do seu casamento.* [Ex.C]

= *O Rui deu uma declaração à imprensa de que vai se casar.* [Ex.C]

As três frases são apenas transformações de uma mesma construção. A diferença entre elas é o tipo de complemento *N2=:QueF*, que pode ser preenchido por uma nominalização, quando introduzido por *sobre* ou *acerca de*, e é preenchido por uma frase completiva no modo indicativo, quando introduzido pela preposição *de*.

9.1.4.6 Especificidades das preposições locativas

São consideradas preposições locativas ([*Preploc*]) as diferentes preposições que introduzem argumentos locativos (*Nloc*). Essas preposições podem ser: *a*, *até*, *de*, *em*, *para* ou *por*. Seguem-se alguns exemplos que ilustram os tipos de preposições:

(9.66) *Ele ocupava a terceira posição quando deu uma escapada da pista.* [Ex.R]

(9.67) *A colega radialista Rose Alves deu uma esticada até o atlântico pra celebrar o mar.* [Ex.R]

(9.68) *Gaby e a equipe deu uma parada no Bar do Neno.* [Ex.R]

Esses três exemplos apresentam respectivamente construções com as preposições: *de*, que indica origem; *até*, que indica destino; e *em*, que indica um lugar estático. Como há diferentes tipos de preposições que podem funcionar como [*Preploc*], elas são indicadas na matriz como [*Preploc*], seguidas do tipo específico de preposição (*a*, *até*, *de*, *em*, *para* ou *por*), para que seja possível diferenciar os usos dessas preposições em construções locativas e em construções não-locativas.

Há 40 *CVS* que exigem uma *Preploc* introduzindo o segundo argumento (ou primeiro complemento). Essas *CVS* possuem estruturas sintáticas bastante regulares: são construções com dois argumentos, sendo que o primeiro é o sujeito, podendo ser um AGENT-GEN ou um OBJECT-GEN e o segundo, introduzido por [*Preploc*], é o complemento LOCATIVE-X. Todas elas foram classificadas na classe DLC2 e aquelas que possuem equivalência semântica com uma construção verbal estão descritas no ViPER nas classes 35LD ou 38LD, como as seguintes construções:

(9.69) *(O Pedro + A embarcação) navegou pelos oceanos.* [Ex.C]

(9.69a) ≡ *(O Pedro + A embarcação) deu uma navegada pelos oceanos.* [Ex.C]

(9.70) *O Pedro navega na internet (todos os dias).* [Ex.C]

(9.70a) ≡ *O Pedro dá uma navegada na internet (todos os dias).* [Ex.C]

O verbo *navegar*, assim como a construção nominal *dar uma navegada*, possui dois significados distintos: um literal e outro figurado, portanto, trata-se de duas construções diferentes. Existem dois nomes predicativos *navegada1* e *navegada2*, que aparecem na tábua como duas entradas lexicais diferentes. As duas construções verbais (9.69) e (9.70) foram inseridas na classe 35LD do ViPER, e as duas construções nominais (9.69a) e (9.70a) foram classificadas na classe DLC2, neste trabalho.

9.2 Propriedades distribucionais

9.2.1 Nomes humanos (*Nhum*) e Nomes não-humanos (*N-hum*)

Consideram-se nomes humanos (*Nhum*) todos os argumentos que podem ser preenchidos por um nome próprio de pessoa, ou sintagmas nominais que designam membros de grupos socioprofissionais (*o aluno, o professor, o juiz, o atleta, o presidente etc.*) e os nomes humanos coletivos (*a comissão, o grupo, a turma, a equipe etc.*).

(9.71) *(Zé + O aluno + A equipe) deu uma descansada.* [Ex.C]

Devem-se considerar também as extensões da classe de *Nhum* abordadas por Gross (1975, p.48), ou seja, elementos concretos que podem ser interpretados metonimicamente como grupo de pessoas.

(9.72) *A bancada deu uma justificativa para a aprovação da lei.* [Ex.C]

(9.73) *A igreja deu uma explicação sobre os padres pedófilos.* [Ex.C]

Nomes como *igreja, bancada, mesa, plenário, empresa, governo, Brasil*, dentre outros, podem ter uma interpretação locativa em outros contextos, mas funcionam aqui como *Nhum*, pois podem ser substituídos por nomes próprios de pessoas.

Já a indicação de nome não-humano (*N-hum*) é definida por oposição à classe dos *Nhum*. Seguindo os postulados de Gross (1975, p.49):

> [...] não consideramos a notação *N-hum* como representando uma classe, ela é apenas um meio formal de precisar a distribuição dos *Nhum*. As classes de substantivos 'não humanos' têm sido habitualmente descritas numa base puramente semântica, isto é, em geral não operatória.[46]

Nesse sentido, marcamos a propriedade *N-hum* para aqueles casos em que a posição sintática não pode ser preenchida por um nome humano, como nos exemplos a seguir:

(9.74) *(*O Rui + A laranja) deu uma apodrecida.* [Ex.C]

(9.75) *(*O Rui + A lenha) deu um estalo.*[47] [Ex.C]

46 Tradução minha. Do original: "Nous ne considererons pas la notation *N-hum* comme représentant une classe, elle n'est pour nous qu'un moyen formel de préciser la distribution des *Nhum*. Les classes de substantifs 'non humains' ont habituellement été décrites sur des bases purement sémantiques, c'est-à-dire non opératoires en général" (GROSS, 1975, p.49).

47 Note-se que, em Português Europeu, o *Npred* estalo também pode significar tapa. Nesse sentido, o sujeito *Nhum* AGENT é autorizado (*e.g. Cavaco Silva deu um estalo à classe política* [Ex.R]).

(9.76) *(*O Rui + O leite) deu uma talhada.* [Ex.C]

Apesar de apresentarmos exemplos de *Nhum* e *N-hum* somente na posição sujeito, essas considerações valem para todos os argumentos em qualquer posição sintática. Na posição de complemento do verbo, por exemplo, indicamos (9.77) e (9.78) cujo *N*1 é marcado como *N-hum*.

(9.77) *O Zé deu uma ferventada no (leite + *Rui).* [Ex.C]

(9.78) *O Zé deu uma provada no (suco + *Rui).* [Ex.C]

Nesse sentido, a etiqueta de *N-hum* só é marcada, de forma restritiva, para nomes concretos, não-animados, que não tenham interpretação locativa, temporal ou eventiva.

Há nomes predicativos que se referem tipicamente a processos climáticos, mas que podem ser associados metaforicamente a ações ou processos humanos, tais como: *esfriada*, *esquentada*, *refrescada* e outros, como se observa em:

(9.79) *(O tempo + Ana) deu uma esfriada.* [Ex.C]

(9.80) *(A sala + Rui) deu uma refrescada.* [Ex.C]

Esses *Npred* que se referem tipicamente a atividades climáticas foram marcados na matriz com *Nhum* e *N-hum*, simultaneamente, na posição de sujeito, mas não foram desdobradas em duas ou mais entradas lexicais porque, independentemente do tipo de sujeito, não mudam suas propriedades distribucionais ou sua estrutura argumental.

Procedeu-se da mesma forma, marcando-se "+" para *Nhum* e *N-hum* na posição de *N*1, nos casos em que o complemento pode ser tanto um humano (PATIENT) quanto um não-humano (OBJECT).

(9.81) *O Rui deu um chute (no Zé + na cadeira).* [Ex.C]

Com relação aos nomes predicativos que expressam ações tipicamente de animais, adotamos a mesma decisão de outros autores (GROSS, 1981; GIRY-SCHNEIDER, 1987; RANCHHOD, 1990; BAPTISTA, 2005b). Seguindo a tradição do Léxico-Gramática, consideramos os atos/ações típicas de animais, bem como a expressão das vozes de animais com o preenchimento do sujeito de tipo humano (*N*0=:*Nhum*).

(9.82) *(O gato + A Ana) deu um berro.* [Ex.C]

(9.83) *(O lobo + O Rui) deu um uivo.* [Ex.C]

Reconhecemos que essas construções com os *Npred berro*, *bramido*, *cacarejo*, *coice*, *galope*, *gorjeio*, *latida*, *miada*, *mugido*, *oinc*, *pio*, *relincho*, *rugido*, *trote*, *uivo*, *urro* e *zurro* podem ter a posição sujeito preenchida tanto por nome de pessoa como por nome de animal. No entanto, assinalamos com "+" na matriz apenas a propriedade *Nhum*, por considerarmos que eles compartilham o traço [+animado] com os nomes humanos, portanto, podem ser considerados nomes humanos.

Essa decisão se justifica por um fenômeno já bastante conhecido na literatura, que se refere à possibilidade de os nomes de animais poderem ser metaforicamente investidos de qualidades tipicamente humanas (prosopopeia). Assim, a distinção entre humano e não-humano não se mantém nesses casos.

Convém ressaltar que alguns desdobramentos lexicais de *Npred* atribuídos originalmente a nomes de animais foram feitos com base em propriedades sintáticas das construções, e não nos traços semânticos do sujeito. Consideramos, por exemplo, duas entradas lexicais para o nome predicativo *trote*:

(9.84) *(O cavalo + O Rui) deu um trote.* [Ex.C]

(9.85) *(*O cavalo + O Rui) deu um trote na Ana.* [Ex.C]

O desdobramento desse *Npred* se justifica porque as duas construções possuem diferentes estruturas argumentais, e não por causa do tipo semântico do sujeito, já que ambas são marcadas com "+" em *Nhum* e "-" em *N-hum*. Essa restrição é muito fraca no que se refere a nomes de animais justamente porque frequentemente são atribuídas características humanas a eles.

Convém esclarecer que alguns nomes de instrumentos podem assumir a posição de sujeito (GUILLET; LECLÈRE, 1981, p.106), tanto nas construções com verbo pleno, como ilustra (9.86), quanto em construções com verbo-suporte, como ilustra (9.87).

(9.86) *O Rui derrubou a árvore com um machado.* [Ex.C]

= *O machado derrubou a árvore.* [Ex.C]

= *O machado do Rui derrubou a árvore.* [Ex.C]

(9.87) *O cinegrafista deu um close no ator com a câmera.* [Ex.C]

= *A câmera deu um close no ator.* [Ex.C]

= *A câmera do cinegrafista deu um close no ator.* [Ex.C]

As três frases são equivalentes semanticamente, tanto na construção com o verbo pleno *derrubar* (9.86), quanto na construção com o verbo-suporte *dar* e o nome predicativo *close* (9.87). Consideraremos como frase de base a construção cujo *N*0 é preenchido por *Nhum*, levando em conta o *Princípio da Máxima Projeção Argumental* e levando em conta que os nomes de instrumentos não são considerados complementos essenciais, mas sim circunstanciais, assim não participam da rede argumental da construção de base.

Nesses casos, o instrumento "sobe" para a posição de sujeito, fenômeno este que é conhecido como *elevação do objeto* (CASTELEIRO, 1981). As *CVS* que autorizam a elevação do objeto foram marcadas com *N*0=:*Nhum* porque considera-se que o instrumento na posição sujeito é uma transformação possível, mas não constitui a frase de base.

9.2.2 Nomes de parte do corpo (*Npc*)

Os nomes de partes do corpo estão descritos em uma coluna específica na matriz (ver Apêndice A) porque apresentam características muito singulares. Há um número considerável de construções que admitem, principalmente na posição de complemento (*N*1), tanto um *Npc* quanto um *Nhum* em frases de superfície[48].

(9.88) *E Mike Tyson deu um tapa na cara de Robert Downey Jr.* [Ex.R]

(9.88a) ≡ *E Mike Tyson deu um tapa em Robert Downey Jr.* [Ex.C]

Entre (9.88) e (9.88a), ocorre a reestruturação do grupo nominal *na cara de Robert Downey Jr.* para o grupo nominal *em Robert Downey Jr.*, omitindo-se o nome que especifica a parte do corpo. Guillet e Leclère (1981) analisaram esse tipo de reestruturação do grupo nominal em construções verbais. Adotamos a mesma análise para as construções nominais:

(9.89) *O patrão chicoteou as costas do escravo.* [Ex.C]

[*Reestr.*] *O patrão chicoteou o escravo nas costas.* [Ex.C]

≡ *O patrão deu uma chicotada nas costas do escravo.* [Ex.C]

[*Reestr.*] *O patrão deu uma chicotada no escravo (nas costas).* [Ex.C]

48 No exemplo (9.88), a primeira frase foi retirada da *web* e a segunda frase é uma transformação, cuja estrutura também pode ser encontrada na *web*, em frases como: *um guarda municipal deu um tapa em uma paciente*, *ela deu um tapa em Ailton Felix* e outras.

(9.90) *O Rui cutucou o ombro da Ana.* [Ex.C]

[*Reestr.*] *O Rui cutucou a Ana no ombro.* [Ex.C]

≡ *O Rui deu uma cutucada no ombro da Ana.* [Ex.C]

[*Reestr.*] *O Rui deu uma cutucada na Ana (no ombro).* [Ex.C]

Apesar de considerarmos que a frase com complemento *em Nhum* é apenas uma reestruturação da frase de base com complemento do tipo *em Npc de Nhum*, assinalamos "+" para as duas posições de *Nhum* e *Npc* simultaneamente. Essa decisão foi tomada a fim de diferenciar entre os casos que aceitam a reestruturação do grupo nominal (admitindo tanto *Nhum* quanto *Npc*) e os casos que aceitam apenas um tipo de argumento (ou *Nhum* ou *Npc*).

Alguns nomes predicativos, diferentemente de (9.89) e (9.90), só aceitam *Npc* como argumento, seja na posição sujeito, seja na posição de complemento, isto é, não admitem o *Nhum* como argumento.

(9.91) *(O estômago do Rui + *O Rui) deu uma embrulhada.* [Ex.C]

Se, por um lado, há *Npred* que só admitem *Npc* como argumento, por outro lado, há *Npred* que só aceitam *Nhum* como argumento. Conforme aponta Baptista (2005b, p.76), certos nomes predicativos exprimem qualidades físicas de pessoas, porém só admitem *Nhum* na posição sujeito; não autorizam o uso de *Npc* nessa posição:

(9.92) *(O Rui + ?*A barriga do Rui) deu uma emagrecida.* [Ex.C]

(9.93) *(O Rui + ?*A garganta do Rui) deu uma engasgada.* [Ex.C]

Os *Npc* na posição de complemento são bastante frequentes como argumentos de *Npred* constituídos com o verbo-suporte *dar*, porém, na posição sujeito, são apenas 14 *CVS* que exigem exclusivamente *N0=:Npc*, tais como, *anelada, atrofiada, cariada* e *lacrimejada*. Outros *Npred* que admitem *N0=:Npc* podem aceitar também outros tipos de argumentos. Observe-se, por exemplo:

(9.94) *(A Ana + O tecido + A pele da Ana) deu uma enrugada.* [Ex.C]

(9.95) *(*A Ana + O tecido + O cabelo da Ana) deu uma ondulada.* [Ex.C]

(9.96) *(*A Ana + *O móvel + O estômago da Ana) deu uma embrulhada.* [Ex.C]

O exemplo (9.94) autoriza qualquer tipo de argumento (*Nhum*, *N-hum* e *Npc*), na posição sujeito. O exemplo (9.95) autoriza *N-hum* ou *Npc* na mesma posição, mas não autoriza *Nhum*. Já o exemplo (9.96) autoriza apenas *Npc* na mesma posição, não admitindo *Nhum* ou *N-hum*.

Os *Npc* estabelecem, de forma sistemática, uma relação semântica de inalienabilidade com um *Nhum* (BOONS; GUILLET; LECLÈRE, 1976). Essa relação semântica particular está na origem de certos fenômenos sintáticos, tais como a interpretação correferencial (metonímica) dos *Npc* com o artigo definido na posição de complemento direto de certos verbos (BAPTISTA, 2005b, p.75). Esse fenômeno pode ser exemplificado por construções como:

(9.97) *Ezequiel Rodriguez também deu uma arrumada no cabelo.* [Ex.R]

Nessa construção, o complemento *no cabelo* possui necessariamente um artigo definido e se refere ao *cabelo de Ezequiel Rodriguez*. O complemento *no cabelo* só pode ter outro referente se houver, na frase, um adjunto restritivo que especifique de quem é o cabelo, como *Ezequiel Rodriguez também deu uma arrumada no cabelo da Ana*.

9.2.3 Nomes Locativos (*Nloc*)

Antes de iniciar a análise dos nomes locativos, é preciso definir em que situações os nomes locativos são considerados argumentos essenciais da predicação e em que situações os mesmos nomes locativos são considerados circunstanciais cênicos (não obrigatórios). Para tanto, considerem-se os seguintes exemplos:

(9.98) *O Rui vive na praia.* [Ex.C]

(9.99) *O Rui malha na praia.* [Ex.C]

Parece consensual entre linguistas e gramáticos que a construção (9.98) exige um complemento locativo (*na praia*), ao passo que a construção (9.99) admite, mas não exige, o elemento *na praia*. Em (9.99), a predicação estaria completa com apenas o argumento da posição sujeito (*Rui malha*), o que significa que o elemento *na praia*, em (9.99), não é obrigatório, por isso é chamado de locativo cênico ou circunstancial de lugar. Admitindo-se a diferença entre essas duas construções, pergunta-se: quais as propriedades sintáticas que nos permitem classificar *na praia*, em (9.98) como argumento (obrigatório) e, em (9.99), como circunstancial (não-obrigatório)?

Gross (1981, p.12) estabelecia algumas perguntas clássicas para identificar argumentos, como "*(prep) o quê?*" ou "*(prep) quem?*", e outras perguntas para identificar circunstanciais, como "*(prep) quando?*", "*(prep) onde?*", "*como?*" e "*por quê?*". O próprio autor admite, no entanto, que a pergunta "*Onde?*" pode trazer informações de diferentes tipos, conforme aponta nos exemplos:

(9.100) *Onde Max vai? R= Max vai à praia.* [Ex.C]

(9.101) *Onde Max caminha? R= Max caminha na praia.* [Ex.C]

O constituinte locativo de (9.100) expressa um destino, ao passo que o constituinte locativo de (9.101) expressa um elemento da cena de *caminhar*. Isso permite afirmar que o teste da pergunta com "*Onde?*" não é suficiente para distinguir entre os argumentos locativos e os circunstanciais cênicos.

Para identificar se o constituinte locativo é um argumento ou um circunstancial, pode-se verificar a aceitabilidade de outros testes. Os circunstanciais cênicos podem ser identificados, por exemplo, por meio de uma paráfrase com verbos-suporte de ocorrência, tais como: *acontecer*, *dar-se*, *decorrer*, *ocorrer*, *suceder* e *ter lugar*.

(9.102) *A Ana deu um escândalo na festa.* [Ex.C]

≡ *A Ana deu um escândalo # Isso (aconteceu + se deu + decorreu + ocorreu + sucedeu + teve lugar) na festa.* [Ex.C]

O fato de a construção (9.102) permitir a paráfrase com algum verbo-suporte de ocorrência indica que o complemento *na festa* é um circunstancial cênico, e não um argumento essencial da construção. Por outro lado, os casos em que o argumento locativo é obrigatório não autorizam essa paráfrase, como se observa em:

(9.102) *A Ana deu um escândalo na festa.* [Ex.C]

≡ *A Ana deu um escândalo # Isso (aconteceu + se deu + decorreu + ocorreu + sucedeu + teve lugar) na festa.* [Ex.C]

(9.103) *A Ana deu uma passada na festa.* [Ex.C]

**A Ana deu uma passada # Isso (aconteceu + se deu + decorreu + ocorreu + sucedeu + teve lugar) na festa.* [Ex.C]

Em (9.103), o constituinte locativo *na festa* é tão necessário para a construção da frase que ele não pode ser deslocado para fora da predicação; por isso é chamado de argumento.

Já em (9.102), o constituinte locativo *na festa* pode ser deslocado da predicação de *dar um escândalo* e ser inserido em outra sentença. A possibilidade de esse elemento ser retirado do escopo da predicação indica que ele é apenas um circunstancial, não um argumento. No caso da construção com o *Npred passada*, o locativo é considerado um dos argumentos obrigatórios da frase de base.

Conforme foi apontado na subseção 9.1.4, uma das preposições principais que introduzem argumentos locativos é a *Prep em*. Contudo, essa preposição também pode introduzir argumentos não-locativos, em particular aqueles que designam um espaço abstrato, mas não indicam um lugar, como espaço físico (GUILLET; LECLÈRE, 1981, p.106).

(9.104) *O banco deu um adiamento no prazo.* [Ex.C]

*?*Onde o banco deu um adiamento? R= No prazo.* [Ex.C]

(9.105) *A Ana deu um beijo no Rui.* [Ex.C]

*?*Onde a Ana deu um beijo? R= No Rui.* [Ex.C]

Os argumentos *no prazo* e *no Rui* não podem ser considerados argumentos locativos das frases (9.104) e (9.105) porque não respondem à pergunta "*Onde?*". Eles poderiam responder à pergunta "*Em quê?/ Em quem?*". Esses constituintes são considerados argumentos, porém não são locativos. Isso nos permite concluir que o teste da pergunta "*Onde?*" serve para identificar constituintes locativos, porém não serve para determinar se o constituinte locativo é um argumento ou um circunstancial.

Posta a diferença entre os argumentos e os circunstanciais, passemos à análise dos argumentos que só podem ser instanciados por nomes locativos (*Nloc*). Consideramos *nomes locativos*: (i) os substantivos próprios que são nomes de cidade, estados, países, ruas *etc.*, também chamados topônimos; e (ii) os substantivos comuns que têm como referência um espaço físico, um local.

Dependendo do tipo de preposição que introduz o argumento locativo, esse *Nloc* pode ser de diferentes tipos semânticos. Neste trabalho, consideramos quatro subtipos de *Nloc*:

i) os locativos estáticos (LOCATIVE-PLACE), que são introduzidos pela [*Preploc*] *em* e respondem à pergunta "*Onde?*";

ii) os locativos dinâmicos (LOCATIVE-PATH), que são introduzidos pela [*Preploc*] *por* e respondem à pergunta "*Por onde?*";

iii) os locativos de origem (LOCATIVE-SOURCE), que são introduzidos pela [*Preploc*] *de*, indicam a origem ou o estado inicial antes do movimento indicado pelo *Npred* e respondem à pergunta "*De onde?*"; e

iv) os locativos de destino (LOCATIVE-DEST ou LOCATIVE-DESTINATION), que são introduzidos pelas [*Preploc*] *até*, *a* ou *para*, indicam o destino ou a posição final do *N*0 após o movimento indicado pelo *Npred* e respondem à pergunta "*(Para + Até) onde?*".

Há ainda quatro casos de *Npred* marcados na matriz (ver Apêndice A) com o papel semântico LOCATIVE-X (*deslocada*, *escapada*, *escapulida* e *fugida*). O subtipo de elemento locativo não foi especificado porque esses *Npred* aceitam tanto um locativo de origem quanto um locativo de destino, na posição de *N*1.

Os papéis semânticos dos argumentos locativos serão analisados mais a fundo na subseção 9.2.5. Por ora, basta-nos indicar o tipo de argumento que se considera como *Nloc* e os tipos de preposições locativas que os *Npred* podem selecionar.

9.2.4 Frases completivas (*QueF*) na posição de argumento

As frases completivas podem, por vezes, preencher a posição de argumento de alguns *Npred*. São, portanto, representadas como *Ni=:QueF*. Especificamente nas construções com o *Vsup dar*, só existem completivas nas *CVS* com três argumentos. Essas frases completivas podem assumir ou a posição do primeiro argumento, como demonstra (9.106), ou do terceiro argumento, como demonstra (9.107).

(9.106) *O juiz ter assinado a liminar deu efetivação aos contratados.* [Ex.C]

(9.107) *A Ana deu uma autorização ao Rui para se casar novamente.* [Ex.C]

Conforme foi apontado na subseção 9.1.1, a ordem dos constituintes não é fixa. Em Português, é mais comum que o sujeito esteja na posição de primeiro argumento (*N*0) e os complementos estejam em posições pós-verbais. No caso das construções com três argumentos, o segundo e o terceiro não possuem posição fixa, portanto o segundo argumento pode passar para a posição do terceiro argumento e vice-versa. No entanto, convencionou-se considerar, para as *CVS* com três argumentos, que o segundo argumento (ou primeiro complemento) é aquele que sofre conversão (ver subseção 9.3.2). Nesse sentido, as frases completivas assumem a posição do terceiro argumento (ou segundo complemento).

Também é possível que frases completivas preencham simultaneamente as duas posições sintáticas de *N*0 e *N*2, tal como:

(9.108) *O juiz ter assinado a liminar deu brecha a outros interessados para entrarem com o pedido na justiça.* [Ex.C]

Essa construção apresenta uma *QueF* na posição sujeito (*N*0=:*O juiz ter assinado a liminar*) e outra *QueF* na posição de segundo complemento (*N*2=:*para entrarem com o pedido na justiça*). Ressalte-se que o sujeito de *N*2 é correferente ao primeiro complemento, que é um *Nhum* (*N*1=:*a outros interessados*).

Todos os exemplos citados apresentam frases completivas no infinitivo, mas há diferentes tipos de frases completivas. Os tipos de completivas analisados neste trabalho são: (i) completivas infinitivas (*e.g. O juiz ter assinado a liminar* <...>); (ii) completivas finitas (*e.g. Que o juiz tenha assinado a liminar* <...>); e (iii) completivas factivas (*e.g. O fato de o juiz ter assinado a liminar* <...>).

As completivas infinitivas, representadas na matriz por *N*0=:*V infw*, são, na verdade, uma redução da completiva finita. Essa operação é chamada de *desfinitização* por Casteleiro (1981, p.245-265) e Baptista (2005b, p.82-88).

As completivas finitas correspondem às frases conjugadas. Dependendo da natureza semântica dos nomes predicativos, seleciona-se o modo verbal da completiva (indicativo ou subjuntivo). Assim, alguns *Npred* só aceitam completivas no indicativo, como (9.109), e outros só aceitam completivas no subjuntivo, como (9.110).

(9.109) *O Rui deu uma demonstração à Ana de que (*seja + é) capaz.* [Ex.C]

(9.110) *O diretor deu uma advertência ao aluno para que ele não (falte + *falta) aula.* [Ex.C]

Esses dois subtipos de completivas finitas estão representados na matriz como: *Ni=:QueFind* para as frases no indicativo, e *Ni=:QueFsubj* para as frases no subjuntivo.

As completivas factivas, por sua vez, correspondem às frases introduzidas pela expressão *O fato de que*, como se observa em:

(9.111) *O fato de que Rui não tenha fechado o negócio deu uma estimativa do tamanho do problema.* [Ex.C]

= *O fato de Rui não ter fechado o negócio deu uma estimativa do tamanho do problema.* [Ex.C]

Como se nota na equivalência entre as duas frases, as completivas factivas podem entrar tanto em construções finitas quanto em construções infinitivas. As propriedades gerais das construções factivas já foram descritas por Casteleiro (1981, p.191) e retomadas na análise das *CVS* com *ser de*, por Baptista (2005b, p.102-108).

Não nos aprofundaremos na análise e descrição das construções completivas, haja vista que elas já foram longamente descritas na literatura. Para uma visão mais aprofundada sobre as propriedades das completivas e suas reestruturações, veja-se, entre outros, Raposo (1975), Casteleiro (1981), Oliveira (1981) e Baptista (2005b).

9.2.5 Papéis semânticos dos argumentos

Os papéis semânticos (ou papéis temáticos) correspondem, grosso modo, às informações de *Quem? Fez o Quê? A Quem? Como? Quando?* e *Onde?*, ou seja, são as relações semânticas que se estabelecem entre o predicado e seus argumentos. Apesar de haver, na literatura, diferentes conjuntos de papéis semânticos (FILLMORE; BACH; HARMS, 1968; GILDEA; JURAFSKY, 2002; PALMER; GILDEA; KINGSBURY, 2005; BICK, 2007), adotaremos um rol próprio, que é um subconjunto do conjunto proposto por Talhadas (2014). O autor anotou e analisou predicados verbais, a partir dos quais constituiu uma lista de 37 papéis semânticos. Daquele rol, observamos 19 que aparecem como papéis dos argumentos dos predicados nominais com o *Vsup dar*.

No trabalho de Talhadas (2014), alguns desses papéis semânticos foram subdivididos de forma a ficarem mais específicos. O papel AGENT, por exemplo, pode ocorrer como AGENT-GENERIC, AGENT-SPEAKER, AGENT-GIVER, AGENT-TAKER, AGENT-CREATOR e AGENT-CAUSE. No caso das *CVS* com o verbo *dar*, aparecem agentes dos tipos AGENT-GENERIC, AGENT-SPEAKER e AGENT-CAUSE.

Além dos 19 papéis semânticos pertencentes ao conjunto de Talhadas (2014), criamos mais um, intitulado TAG, que corresponde a uma etiqueta, um código, um rótulo atribuído a outro argumento. Em suma, trabalhamos com um rol final de 20 papéis semânticos, que são definidos a seguir:

- AGENT-GEN: é um agente genérico, portanto deve ser um humano, voluntário, volitivo, que pratica uma ação. A noção de ação é necessária ao predicado para que o argumento se configure como agente. (*e.g.* ***A Ana*** *deu um tiro no Rui*).
- AGENT-CAUSE: é muitas vezes o sujeito, que pode ser humano ou não-humano, voluntário ou não, de predicados que não expressam necessariamente uma ação concreta. O sujeito desses predicados pode ser tanto um agente quanto uma causa (*e.g. (**A Ana + A vida**) deu um privilégio ao Rui*).

- AGENT-SPEAKER: é um subtipo de agente, com a especificidade de que a ação é um ato declarativo (*e.g.* ***A Ana*** *deu uma declaração ao Rui de que voltaria cedo*).
- EXPERIENCER-GEN: é um experienciador genérico, ou seja, ele participa da cena, mas a cena não envolve uma ação. O experienciador se distingue do agente porque o primeiro é o participante de um processo, enquanto o segundo é o autor de uma ação (*e.g.* ***A Ana*** *deu uma acalmada*).
- EXPERIENCER-VOL: é um subtipo de experienciador, mas, neste caso, ele é volitivo. O predicado semântico não corresponde a uma ação, mas a um processo, e a pessoa que participa desse processo o faz de maneira voluntária (*e.g.* ***A Ana*** *deu abertura para o Rui falar*).
- OBJECT-GEN: refere-se a uma entidade não-humana (*N-hum*) e concreta. Pode aparecer como sujeito da construção, mas é mais comum ocorrer como complemento (*e.g. A Ana deu uma limpada* ***nos móveis***).
- OBJECT-CL: refere-se a uma parte do corpo humano. Assim como outros tipos de objetos, o OBJECT-CL pode ocorrer na posição de sujeito, mas é mais comum em posição de complemento (*e.g. A Ana deu uma depilada* ***nas pernas***).
- OBJECT-F: corresponde aos complementos frásicos principalmente em predicados de percepção mental (*e.g. A Ana deu sua anuência ao Rui para* ***ele se casar***).
- PATIENT: pode ser tanto humano (*Nhum*) quanto não-humano (*N-hum*) desde que seja um afetado pela ação ou pelo processo expresso pelo predicado. Também é designado em outros trabalhos como TEMA (*e.g. A Ana deu um tiro* ***no Rui***).
- CO-AGENT: é um tipo de argumento simétrico que pratica conjuntamente a mesma ação que o agente. Se o sujeito é agente, o complemento é o co-agente (*e.g. A Ana deu uma trepada com* ***o Rui***).
- CO-OBJECT: é um tipo de argumento simétrico que pratica conjuntamente a mesma ação que o objeto da posição sujeito. A diferença entre CO-AGENT e CO-OBJECT é que o primeiro exige argumentos humanos, enquanto o segundo exige argumentos não-humanos (*e.g. A água deu uma misturada com* ***o açúcar***).
- ADDRESSEE: é o argumento humano que corresponde ao objeto indireto em atos de comunicação, ou seja, é o ouvinte (*e.g. A Ana deu um conselho para* ***o Rui*** *não voltar à cidade*).
- MESSAGE: é o argumento frásico que corresponde ao conteúdo de uma comunicação, em geral em predicados com três argumentos (*e.g. A Ana deu o ultimato para o Rui* ***se redimir***).

- INSTRUMENT: é o objeto concreto ou instrumento usado para realizar uma ação (*e.g. A Ana deu uma borrifada de **perfume** no pescoço*).
- TAG: é uma etiqueta, um rótulo atribuído a algum argumento e se refere a um nome ou designação (*e.g. A Ana deu o título de "**astro**" ao Rui*).
- TOPIC: também é designado na literatura como SUBJECT-MATTER e se refere ao assunto (*e.g. A Ana deu um anúncio ao Rui sobre **sua partida***). Esse papel semântico ocorre, em geral, nos predicados semânticos que expressam declaração.
- LOCATIVE-SOURCE: refere-se a um local de origem (*e.g. A Ana deu uma saída **de casa***).
- LOCATIVE-DEST: refere-se a um local de destino (*e.g. A Ana deu uma fugida **para a praia***).
- LOCATIVE-PLACE: refere-se ao local estático onde o sujeito se encontra (*e.g. A Ana deu um mergulho **na piscina***).
- LOCATIVE-PATH: refere-se percurso que o sujeito percorre (*e.g. A Ana deu uma trafegada **pela cidade***).

9.3 Propriedades transformacionais

9.3.1 Simetria

A propriedade da simetria (BAPTISTA, 2005a) se verifica quando dois argumentos de um predicado desempenham, relativamente ao núcleo predicativo, o mesmo papel semântico. Nesse sentido, os argumentos podem trocar de posição relativa e/ou serem coordenados, sem que isso altere o significado das frases resultantes. Note-se, por exemplo:

(9.112) *O Rui se casou com a Ana.* [Ex.C]

(9.112a) ≡ *A Ana se casou com o Rui.* [Ex.C]

(9.112b) ≡ *O Rui e a Ana se casaram <um com o outro>.* [Ex.C]

(9.112c) ≡ *A Ana e o Rui se casaram <um com o outro>.* [Ex.C]

Para que o predicado aceite a simetria, deve ser possível inverter a posição do sujeito com a do complemento, sem que o significado da construção se altere. Nas sentenças acima, o sujeito de (9.112) passa a ser complemento de (9.112a), e o complemento de

(9.112) passa a ser sujeito de (9.112a). Em (9.112b) e (9.112c), as duas entidades (*Ana* e *Rui*) se coordenam e assumem a posição de sujeito. Independentemente da posição sintática desses argumentos, o predicado semântico dessas frases de base é o mesmo.

O conceito de simetria, como uma propriedade transformacional, foi introduzido por Borillo (1971) para os verbos em Francês. Em Português, Fonseca (1984) tratou dos verbos simétricos, mas considerando apenas os predicados verbais. Em seguida, Ranchhod (1990) e Baptista (2005a) analisaram essa propriedade nos predicados nominais com os verbos-suporte *estar prep* e *ser de*, respectivamente.

As construções com o verbo-suporte *dar* dificilmente aceitam essa transformação, mas há alguns casos que merecem consideração. A possibilidade de simetria entre os predicados com o *Vsup dar* é tão rara que Baptista (2005a, p.362) chega a considerar, em nota de rodapé:

> [...] à excepção das construções com verbo-suporte *dar* (VAZA, 1988; BAPTISTA, 1997), que, aparentemente, nunca admitem a simetria visto caracterizarem-se justamente pela natureza 'orientada' da relação que se estabelece entre os dois GN argumentos do nome predicativo. Aliás, é essa relação orientada que está na base de certas operações formais que invertem a direcção da relação e a que G. Gross (1989) chamou Conversão[49].

Algumas raras construções com o verbo-suporte *dar* em PB, no entanto, admitem essa propriedade. São construções com *V-n -da* que não são muito frequentes em PE. Em geral, se o verbo pleno que dá origem à nominalização admite a simetria, a construção com *Vsup=:dar* e nominalização em *V-n-da* também a admite.

(9.113) *O Rui deu uma trepada com a Ana.* [Ex.C]

≡ *A Ana deu uma trepada com o Rui.* [Ex.C]

≡ *(O Rui e a Ana + A Ana e o Rui) deram uma trepada.* [Ex.C] ou

(9.113) *O Rui deu uma discutida com a Ana.* [Ex.C]

≡ *A Ana deu uma discutida com o Rui.* [Ex.C]

≡ *(O Rui e a Ana + A Ana e o Rui) deram uma discutida.* [Ex.C]

A única construção da matriz (ver Apêndice A) que não é formada pela nominalização em *-da* e que admite a simetria é *dar aperto de mão*:

49 O texto consta na nota de rodapé nº 4, também disponível *on-line* por meio do endereço: http://ler.letras.up.pt/uploads/ficheiros/4546.pdf.

(9.115) *O Rui deu um aperto de mão na Ana.* [Ex.C]

≡ *A Ana deu um aperto de mão no Rui.* [Ex.C]

≡ *(O Rui e a Ana + A Ana e o Rui) deram um aperto de mão.* [Ex.C]

Vale esclarecer que a propriedade da simetria exige o traço da reciprocidade. Nesse sentido, é preciso que Rui tenha apertado a mão de Ana e que Ana tenha apertado a mão de Rui para que a construção seja simétrica. É esse traço obrigatório que distingue a construção *dar um aperto de mão* da construção *dar um beijo*, por exemplo, porque é possível formar:

(9.116) *O Rui deu um beijo na Ana, mas a Ana não deu um beijo no Rui.* [Ex.C] Mas não[50]:

(9.117) *?*O Rui deu um aperto de mão na Ana, mas a Ana não deu um aperto de mão no Rui.* [Ex.C]

Baptista (2005a, p.364) também ressalta que alguns advérbios simples e compostos "operam exclusivamente sobre predicados simétricos (nestes exemplos, a frase matriz está indicada entre parêntesis rectos): [*O Pedro e a Ana dão-se (E + um com o outro)*] *como o cão e o gato*". Em subseção anterior (6.1.4), foram discutidas também construções como (9.118), que não são constituídas por *Vsup*, mas também admitem a simetria.

(9.118) *O Rui (Ø + não) se dá (Ø + bem) com a Ana.* [Ex.C]

≡ *A Ana (Ø + não) se dá (Ø + bem) com o Rui.* [Ex.C]

≡ *A Ana e o Rui (O Rui e a Ana) (Ø + não) se dão (Ø + bem).* [Ex.C]

Essas construções foram classificadas como expressões fixas, já que exigem um complemento fixo (o advérbio *bem*), o qual pode ser apagado, e um complemento livre, mas com algumas restrições de seleção, além de possuírem determinante fixo (Ø) e sentido figurado. A simetria pode ser observada tanto em construções nominais como em construções verbais e adjetivais.

50 Assim como em vários outros casos de inaceitabilidade, esclareça-se que a interpretação de *dar um beijo* como predicado simétrico é possível, assim como a interpretação de *dar aperto de mão* como predicado não simétrico também pode ser possível, desde que o contexto seja muito bem explicado. Consideramos essas construções como inaceitáveis justamente porque exigem do leitor maior esforço cognitivo para aceitá-las e também porque necessitam de um contexto muito bem delimitado para que sejam interpretadas como tal.

9.3.2 Conversão

A conversão (G. GROSS, 1982, 1989) é uma operação formal (ou transformação) que estabelece uma relação não-orientada de equivalência sintática e semântica (parafrástica) entre duas frases elementares. O verbo suporte *standard*, de orientação ativa – no caso o verbo *dar* – é substituído por outro, um verbo-suporte converso, de orientação passiva – no caso o verbo *levar* ou o verbo *receber*, por exemplo.

(9.119) *Felipe Scarpelli deu um soco no atleta Sr. Bruno César* [Ex.R]

[*Conversão*] ≡ *O atleta Sr. Bruno César levou um soco do Felipe Scarpelli* [Ex.C]

(9.120) *Zezé Di Camargo deu uma explicação ao público sobre o que aconteceu na véspera.* [Ex.R]

[*Conversão*] ≡ *O público recebeu uma explicação do Zezé Di Camargo sobre o que aconteceu na véspera.* [Ex.C]

Sobre a maioria das construções com o verbo-suporte *dar* incide a operação da conversão, uma "operação sintática que executa uma permuta dos argumentos em torno do núcleo predicativo da frase sem alterar seu significado global, é semelhante à Passiva das construções verbais" (BAPTISTA, 2005b, p.184).

Assim como a construção ativa é considerada *standard*, também a nominalização da construção ativa é considerada *CVS standard*. A mudança de orientação do sentido ativo para passivo numa construção verbal dá origem a uma *construção passiva*. Já a mudança de orientação de ativo para passivo numa construção nominal (ou *CVS*) dá origem a uma *construção conversa*.

Esse tipo de transformação permite a inversão da ordem e a classificação sintática dos argumentos (sujeito e complemento), mas não altera seus papéis semânticos.

(9.121) *O Rui deu um beijo na Ana.* [Ex.C]

[*Conversão*] ≡ *A Ana recebeu um beijo do Rui.* [Ex.C]

Assim, numa transformação da construção verbal ativa para passiva, *Rui* é sempre o agente e *Ana* é sempre paciente, independente de quem esteja na posição de sujeito ou de complemento. Na relação entre a construção *standard* e a construção conversa, os papéis temáticos dos argumentos também não se alteram. Todas essas frases constituem uma classe de equivalência parafrástica (HARRIS, 1961; HARRIS, 1991). A Figura 5 apresenta todas essas equivalências.

A tradição gramatical costuma abordar apenas as relações diretas entre as construções verbais ativa e passiva ou entre a construção verbal e a construção nominal. Nesta pesquisa, consideramos tanto as relações diretas entre passiva e ativa e entre construção *standard* e conversa, quanto as relações indiretas que se estabelecem entre: (i) a construção verbal ativa (*O Zé beijou a Ana*) e a construção nominal passiva (*A Ana recebeu um beijo do Zé*), chamada de *nominalização passiva*; (ii) a construção nominal ativa (*O Zé deu um beijo na Ana*) e a construção verbal passiva (*A Ana foi beijada pelo Zé*), chamada de *nominalização ativa*; e (iii) a construção verbal passiva (*A Ana foi beijada pelo Zé*) e a construção nominal passiva (*A Ana recebeu um beijo do Zé*), chamada de *nominalização com conversão*.

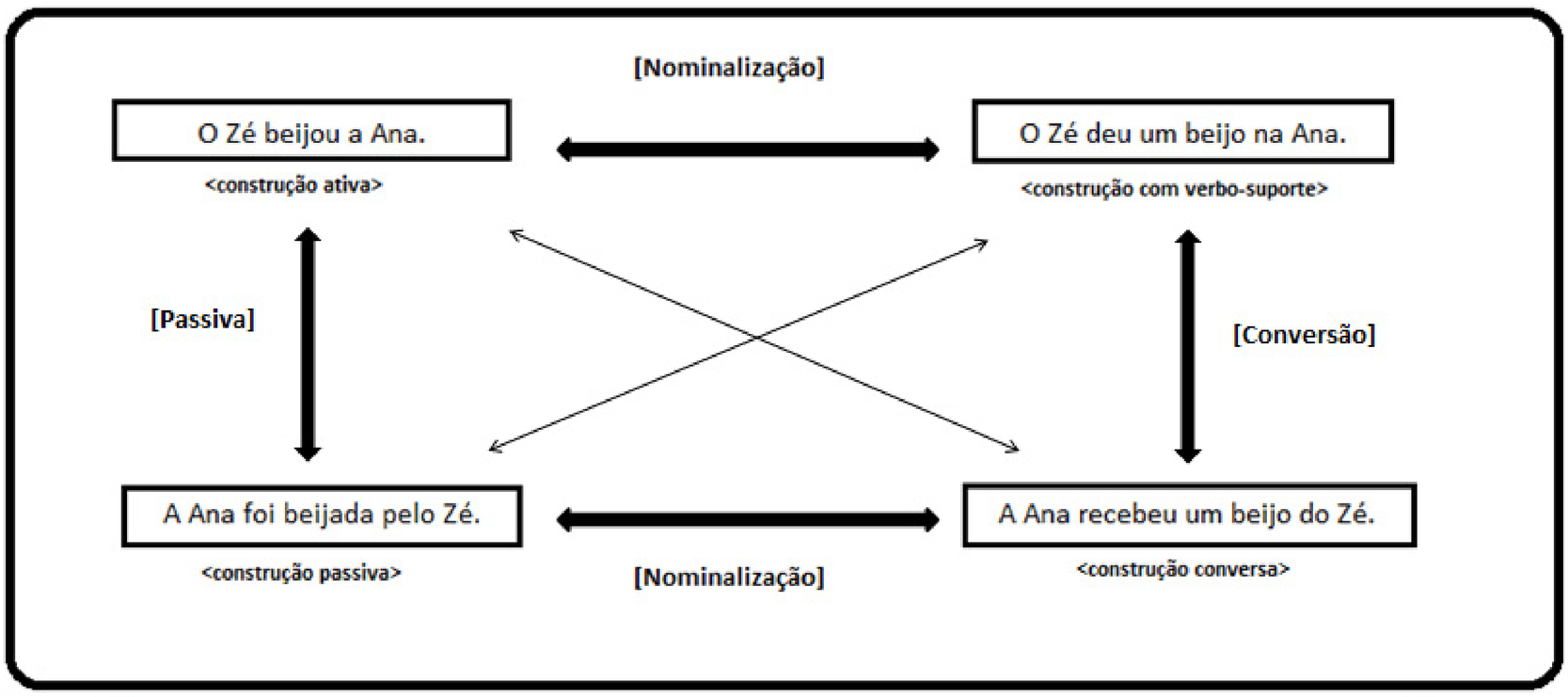

Figura 5 – Relações de equivalência parafrástica entre sentenças de orientação ativa e passiva
Fonte: Elaborado pela autora e já publicado em Rassi *et al.* (2015)

A transformação da conversão foi apontada inicialmente por Gaston Gross (1982, 1989), que identificou que um bom número de construções com verbo-suporte *dar* em Francês têm a possibilidade de derivar construções conversas. Posteriormente essa propriedade foi também analisada por Ranchhod (1990), Baptista (1997, 2005b) e Rassi *et al.* (2015) para o Português.

Assim como a passiva só se faz em construções transitivas diretas, a conversão também só é autorizada em construções que tenham, pelo menos, dois argumentos. Excluem-se da operação de conversão 488 construções que só possuem um argumento, como:

(9.122) *Ana* ***deu um suspiro***. [Ex.C]

Além das construções que interditam a conversão por possuírem apenas um argumento, destacam-se também as construções que têm como segundo argumento um nome locativo e não admitem conversão.

(9.123) *Ana **deu uma passada** na loja.* [Ex.C] [*Conversão*] **A loja (**recebeu** + **levou** + **teve**) uma **passada** por parte da Ana.* [Ex.C]

Uma propriedade específica das construções com o *Vsup dar*, em relação à conversão, é o fato de que um grande número de nomes predicativos em *-ada* ou *-ida* não autorizam a conversão. Há 383 construções (ver Apêndice A) que, mesmo selecionando dois ou três argumentos, não admitem a conversão, como em:

(9.124) *A Ana deu uma lida no texto.* [Ex.C]

[*Conversão*] **O texto (recebeu + levou + teve) uma lida por parte da Ana.* [Ex.C]

(9.125) *A Ana deu uma estudada no material.* [Ex.C]

[*Conversão*] **O material (recebeu + levou + teve) uma estudada por parte da Ana.* [Ex.C]

Em grande parte dessas construções, o segundo argumento (*N*1, na posição de complemento) é anotado com o papel temático OBJECT-GENERIC. Quando esse argumento (*N*1, na posição de complemento) é anotado como PATIENT normalmente aceita a conversão. Praticamente todos os nomes predicativos que indicam atos violentos admitem essa propriedade, geralmente com os verbos *levar* ou *tomar*, como comprovam os exemplos seguintes.

(9.126) *A Ana **deu uma cacetada** no Rui.* [Ex.C]

[*Conversão*] ≡ *O Rui **levou uma cacetada** da Ana.* [Ex.C]

(9.127) *A abelha **deu uma ferroada** no Rui.* [Ex.C]

[*Conversão*] ≡ *O Rui **tomou uma ferroada** da abelha.* [Ex.C]

Os exemplos citados consistem em frases de base, portanto todos os seus argumentos estão explícitos. Em frases de *corpus*, porém, é muito comum que o argumento AGENT, em construções conversas, esteja oculto, como se demonstra em:

(9.126a) *O Rui levou uma cacetada.* [Ex.C]

(9.127a) *O Rui tomou uma ferroada.* [Ex.C]

Já que a conversão funciona como uma espécie de passiva nominal e que as construções passivas permitem o apagamento do agente da passiva, assim também a construção conversa permite o apagamento do complemento AGENT.

9.3.3 Apassivação

A apassivação também é um processo transformacional que muitas das *CVS* com *dar* aceitam. Nas construções nominais, a transformação passiva funciona da mesma forma como nas construções verbais: a frase está na voz ativa quando o sujeito é o agente da ação (*e.g. O Rui deu um esclarecimento à Ana*) e está na voz passiva quando o sujeito da frase é o TEMA, PACIENTE ou OBJETO – neste caso, o nome predicativo (*e.g. O esclarecimento foi dado por Rui à Ana*).

Existem apenas duas diferenças sintáticas entre as construções verbais e as construções nominais no que concerne à transformação passiva: (i) o constituinte que preenche a posição sujeito na voz passiva é o objeto direto, em construções verbais, mas é o nome predicativo, em construções nominais; (ii) as construções verbais, para aceitarem a passiva, precisam ter dois argumentos (são construções transitivas diretas), enquanto as construções nominais podem ter apenas um argumento para aceitarem a passiva.

Essas duas diferenças estão intrinsecamente relacionadas. A possibilidade de predicados nominais com um argumento autorizarem a passiva tem a ver com o fato de que o nome predicativo ocupa a posição sintática de objeto direto, apesar de ser o predicador central da construção. Quase todas as construções com o verbo-suporte *dar* admitem passiva, incluindo construções com um, dois ou três argumentos, como se demonstra respectivamente em:

(9.128) *Tudo o que sabemos até agora é que o* ***espirro foi dado*** às 7:30h da manhã de hoje, quinta-feira. [Ex.R]

(9.129) *Outro* ***gole foi dado*** *no copo vermelho.* [Ex.R]

(9.130) *A* ***garantia foi dada*** *pela ministra Tereza Campello.* [Ex.R]

Nas frases de superfície, nem todos os argumentos essenciais da frase de base aparecem explícitos, mas considere-se que *dar espirro* seja um predicado do tipo *On* (com um argumento), *dar gole* seja um predicado do tipo *Onn* (com dois argumentos) e *dar garantia* seja um predicado do tipo *Onnn* (com três argumentos).

Foi identificado apenas um grupo de *CVS* que não admite a apassivação; são aquelas em que o nome predicativo é nome de doença:

(9.131) *Vacinei meu bebê a 6 meses e mesmo assim* ***ele deu catapora****.* [Ex.R]

[*Passiva*] **A* ***catapora foi dada*** *no meu bebê de 6 meses.* [Ex.C]

(9.132) *Meu marido **deu caxumba**, aos 32 anos.* [Ex.R]

[*Passiva*] **A **caxumba foi dada** no meu marido, aos 32 anos.* [Ex.C]

(9.133) *A Manu **deu febre** e levei no hospital.* [Ex.R]

[*Passiva*] **A **febre foi dada** na Manu.* [Ex.C]

O verbo-suporte *standard* mais característico dessas construções é o verbo *ter*, mas na língua em uso encontramos esses *Npred* também associados ao *Vsup dar*. O alto nível de fraqueza semântica do verbo-suporte implica um alto grau de arbitrariedade na sua seleção.

O fato de que essas construções não admitem passiva é mais uma prova em favor de considerar como verbo-suporte elementar dessas construções o *Vsup ter*, pois as construções com *ter* também não admitem apassivação.

9.3.4 Nominalização

A nominalização é considerada como uma transformação geral que consiste em "incorporar" uma proposição em outra; a proposição que é incorporada assume o papel de um substantivo, seja na posição de sujeito, seja na posição de complemento, da frase-matriz. Além disso, a nominalização se submete a uma série de transformações que reduzem uma proposição a um substantivo ou a um infinitivo (GIRY-SCHNEIDER, 1978, p.11-14), por ser a forma nominal do verbo. Essas nominalizações podem ser de três tipos: (i) a que transforma um verbo em substantivo (*He told the story* ≡ *His telling of the story*), exigindo ainda outro verbo que seja o cerne da predicação; ou (ii) a que transforma um adjetivo em substantivo (*John is eager to please* ≡ *John's eagerness to please*), que também funciona como sintagma nominal e exige um predicado; ou ainda (iii) a que transforma uma frase de base em outra frase de base (*He studies eclipses* ≡ *He makes studies of eclipses*).

Neste trabalho, trataremos apenas do terceiro tipo de nominalização, ao que Harris (1964) chamou de *construção com verbo-operador "U"*, e posteriormente Gross (1981) nomeou *construções com verbo-suporte*. Entre *He studies eclipses* e *He makes studies of eclipses*, existe equivalência sintática e semântica: na segunda frase, há formação de um grupo nominal (*studies of eclipses*); o sujeito é o mesmo nas duas frases; o lugar do verbo (*study*) é ocupado por um outro verbo (*makes*), que é chamado de *Vsup*.

Grande parte dos nomes predicativos que selecionam o *Vsup dar* são nominalizações de verbos, tais como: *dar giro* ≡ *girar*, *dar golpe* ≡ *golpear*, *dar grito* ≡ *gritar* e outros. Na

matriz, encontram-se 818 nomes predicativos que possuem relação morfológica com a raiz de um verbo pleno.

(9.134) *Ana deu um **tropeço**.* [Ex.C]

≡ *Ana **tropeçou**.* [Ex.C]

Não pretendemos aqui analisar diacronicamente se se trata de um nome deverbal (*tropeço*) ou de um verbo denominal (*tropeçar*). A abordagem etimológica dessas palavras não está no escopo desta pesquisa. Identificamos apenas se existe uma relação morfológica entre a raiz do nome (*tropeç-*) e a raiz do verbo (*tropeç-*).

Outros nomes predicativos possuem essa relação morfológica com um adjetivo, e não com verbo, como é o caso de *dar gratidão* ≡ *ser grato*.

(9.135) *Ana deu sua **gratidão** ao Rui.* [Ex.C]

≡ *Ana é **grata** ao Rui.* [Ex.C]

Apenas seis *Npred* estão associados à raiz exclusivamente de adjetivo, mas há 104 *Npred* que possuem a mesma raiz tanto de um verbo quanto de um adjetivo, como é o caso de:

(9.136) *O clima deu uma **refrescada**.* [Ex.C]

≡ *O clima **refrescou**.* [Ex.C]

≡ *O clima ficou **fresco**.* [Ex.C]

Como não é nosso objetivo identificar a origem dessas palavras (se têm origem verbal, adjetival ou nominal), limitar-nos-emos a indicar na matriz se tal *Npred* tem relação morfológica com essas outras classes gramaticais ou não.

Uma particularidade das nominalizações que selecionam o *Vsup dar* é a possibilidade de elas serem derivadas, não de um verbo nem de um adjetivo, mas de um outro substantivo. Conforme apontado em seções anteriores, há uma grande produtividade das nominalizações que se formam a partir de um nome de objeto concreto, que pode ser utilizado como instrumento (ou arma), acrescentando-lhe o sufixo *-ada*. É o caso de *bengala*, que dá origem ao *Npred bengalada*. Nomes como *bengalada* selecionam o *Vsup dar* e a construção toda passa a indicar "bater com o objeto x", em que "x" é ocupado pelo nome concreto *bengala*. Observe-se que muitos desses nomes não possuem raízes morfologicamente associadas a verbos ou adjetivos; apenas a substantivos.

(9.137) *O Rui deu uma **tijolada** no Zé.* [Ex.C]

*?*O Rui **tijolou** o Zé.* [Ex.C]

≡ *O Rui bateu no Zé com um **tijolo**.* [Ex.C]

Foram identificados 35 *Npred* em *-ada* que são nominalizações de substantivos concretos. Outros 263 *Npred* podem ter origem tanto em verbo como em adjetivo ou nome. Segue-se um exemplo:

(9.138) *O Rui deu uma **caprichada** no visual.* [Ex.C]

≡ *O Rui **caprichou** no visual.* [Ex.C]

≡ *O Rui foi **caprichoso** com o visual.* [Ex.C]

≡ *O Rui teve **capricho** com o visual.* [Ex.C]

Na matriz, as nominalizações estão descritas nas colunas 43 a 48 com as seguintes notações: (i) [*N-n*] indica que o *Npred* é associado morfologicamente a outro substantivo, seguida do nome correspondente; (ii) [*Adj-n*] indica que o *Npred* é associado morfologicamente a um adjetivo, seguido do adjetivo correspondente; e (iii) [*V-n*] indica que o *Npred* é associado morfologicamente a um verbo, seguido do verbo pleno correspondente.

Há 274 *Npred* que não são nominalizações, ou seja, não são derivados de algum verbo (*abraçar* → *abraço*) nem de adjetivo (*predileto* → *predileção*) e nem de outros nomes concretos (*colher* → *colherada*). Os nomes que não são deverbais, deadjetivais nem denominais são designados como *nomes autônomos*.

9.4 Desdobramentos lexicais

Alguns *Npred* foram duplicados na matriz (ver Apêndice A) por se tratarem de desdobramentos lexicais, ou seja, construções com nomes predicativos homógrafos, mas que possuem diferença semântica ou sintática. Segundo Laporte[51], qualquer distinção estritamente correlata com uma propriedade reprodutivelmente observável é formalizada como separação de entradas.

Foram adotados critérios formais para definir se há ou não desdobramento lexical. Decidiu-se por desdobrar as entradas nos seguintes casos:

51 Trecho proferido em conferência ministrada aos 27 dias do mês de agosto de 2015, na Universidade Federal de São Carlos, sobre o tema "Problemas de cientificidade na descrição da gramática e do léxico".

- quando os nomes predicativos homógrafos possuem distribuições argumentais diferentes, como se verifica nos exemplos a seguir:

(9.139) *Entrei no carro só de marra, ele ligou o carro e* ***deu o retorno****.* [Ex.R]

(9.140) *O megarresort não* ***deu o retorno*** *esperado a investidores.* [Ex.R]

(9.141) *A empresa não me* ***deu o retorno*** *de quando vou receber o produto.* [Ex.R]

Em (9.139), apresenta-se a construção *dar retorno* com apenas um argumento na posição sujeito AGENT-GENERIC (*ele*) e poderia ser parafraseado por *O carro fez o retorno*. Em (9.140), a construção *dar retono* exige dois argumentos: um na posição sujeito OBJECT-GENERIC (*o megarresort*) e outro na posição de complemento PATIENT (*investidores*). Já a construção identificada em (9.141) apresenta três argumentos essenciais: o sujeito AGENT-GENERIC (*empresa*), o complemento PATIENT (*me*) e o complemento TOPIC (*de quando vou receber o produto*).

- quando os nomes predicativos homógrafos, mesmo tendo o mesmo número de argumentos, possuem diferentes restrições de seleção, seja em relação aos determinantes, seja em relação ao tipo de argumento (*Nhum*, *N-hum*, *QueF* ou outros). Por exemplo, há duas construções desdobradas para o *Npred cumprimento*: um referente ao ato de *cumprimentar* e outro referente ao ato de *cumprir*.

(9.142) *Obama agradeceu aos eleitores e referiu que* ***deu os cumprimentos*** *aos candidatos republicanos.* [Ex.R]

(9.143) *A Polícia Civil* ***deu cumprimento*** *a dois mandados judiciais.* [Ex.R]

Os dois exemplos apresentam construções de *dar cumprimento*, ambos com dois argumentos. A diferença entre as construções está no tipo de argumento selecionado pelo *Npred*: em (9.142), o complemento *N*1 é obrigatoriamente do tipo *Nhum* enquanto (9.143) exige um *N-hum* para a posição de complemento *N*1.

- quando um *Npred* possui interpretação literal e o outro possui interpretação figurada (metafórica).

(9.144) *Peguei sabonete passei um pouco nas peças e delicadamente com as mãos* ***deu uma ensaboada****, depois enxaguei bem.* [Ex.R]

(9.145) *Lula* ***deu uma ensaboada*** *em Tarso Genro por causa da idéia de revisão na Lei da Anistia.* [Ex.R]

As duas construções possuem as mesmas propriedades sintáticas relativas ao número de argumentos e às restrições de seleção, mas há diferença semântica entre elas, por isso as entradas foram duplicadas. A construção (9.144) de *dar uma ensaboada* sugere uma interpretação literal, em que "um sujeito humano esfrega sabão em algo ou alguém", enquanto a construção (9.145) sugere uma interpretação metafórica, no sentido de "dar uma bronca".

Existem alguns nomes predicativos que poderiam ser analisados diacronicamente com um significado distinto do que está em uso hoje em dia. É o caso dos nomes como *pancada* e *porrada*, que diacronicamente têm origem em nomes concretos como *panca* e *porra*, porém essa associação ao nome do instrumento que lhe deu origem já se perdeu no processo de mudança linguística. Como este trabalho adota uma perspectiva sincrônica, esses *Npred* não foram desdobrados e a única entrada lexical que consta na matriz é referente ao significado atual dos nomes *pancada* e *porrada*.

O capítulo anterior apresentou as propriedades gerais das *CVS*. Este capítulo apresentou as propriedades específicas das *CVS* com o verbo *dar* em Português Brasileiro. Foram descritas: (i) propriedades formais, tais como o número de argumentos de cada *Npred*, as especificidades sobre os quantificadores, os determinantes que introduzem o *Npred* e o tipo de preposição que introduz os complementos obrigatórios do *Npred*; (ii) propriedades distribucionais, tais como os tipos semânticos dos argumentos (nome humano, nome não-humano, nome de parte do corpo, nome locativo, frase completiva), além dos papéis semânticos desses argumentos; e (iii) propriedades transformacionais, tais como a simetria, a conversão, a passiva e a nominalização.

Além da descrição dessas propriedades, este capítulo abordou os casos em que as entradas lexicais foram desdobradas na matriz. Muitos nomes predicativos homógrafos instauram diferentes predicados e, portanto, foram duplicados. No próximo capítulo, apresentaremos a classificação proposta para as *CVS* com *dar*.

10 Classificação das *CVS*

Neste capítulo, apresentamos os critérios formais para a classificação das *CVS* com *dar*. A classificação que consta na matriz (ver Apêndice A), especificamente na coluna 5 identificada como Classe-PB, é estritamente formal/estrutural com base no número de argumentos e no tipo semântico desses argumentos. Os critérios observados, bem como a definição das classes formais, são apresentados na seção 10.1. Essa categorização será chamada de *classificação sintática*.

Além da classificação sintática, foi proposta também uma clusterização semântica dessas *CVS*, com base não na estrutura sintática, mas nas semelhanças semânticas dos *Npred*. Essa segunda categorização é apresentada na seção 10.2 e será chamada de *clusterização semântica*, que é basicamente um agrupamento de *Npreds* que possuem traços semânticos semelhantes.

A última seção deste capítulo apresenta um tipo particular de verbo-suporte, chamado *verbo-suporte de ocorrência* (*Vsup-oc*). Esse termo foi cunhado por Gross (1986, p.107-109) para designar um pequeno conjunto de verbos cuja função principal é introduzir circunstanciais, principalmente de tempo e espaço, em predicados eventivos. Essa categoria de verbos-suporte será explorada na seção 10.3, já que o verbo *dar* pode entrar nesse tipo de construção.

10.1 Classificação sintática das *CVS*

A partir da análise das propriedades formais, distribucionais e transformacionais descritas nos capítulos anteriores, propusemos uma classificação sintática das construções com verbo-suporte, que leva em conta: (i) o número de argumentos do predicado, considerando-se as construções com sujeito e um ou dois complementos essenciais como argumentos; (ii) as restrições ao preenchimento lexical do sujeito e dos complementos, que podem ser humanos (*Nhum*) ou não-humanos (*N-hum*); (iii) a possibilidade de o nome predicativo admitir uma frase completiva como argumento (seja na posição sujeito seja como complemento); e (iv) especificidades sobre a natureza distribucional de alguns argumentos, que podem ser especificamente nomes de parte do corpo (*Npc*) ou nomes locativos (*Nloc*).

As construções foram organizadas sintaticamente em 15 classes. Os nomes atribuídos às classes são códigos convencionais e arbitrários, apesar de terem sido criados com base nos critérios expostos no parágrafo anterior. Todas as classes têm nomes iniciados por "D" porque correspondem às classes das *CVS* com o verbo *Dar*. As classes [DH1],

[DN1], [DR1] e [DPC1] englobam construções com um argumento. A classe [D3] abarca construções com três argumentos. As demais classes englobam construções com dois argumentos. A letras *H*, *N*, *R*, *PC*, *LC* e *QF* correspondem respectivamente aos argumentos de tipo *Humano*, *Não-humano*, *não-Restrito*, *Parte-do-Corpo*, *LoCativo* e *completiva* (*Que-F*). Nas classes de *CVS* com dois argumentos, a ordem das letras no nome indica a ordem dos argumentos na construção. Assim, por exemplo, a classe [DHN2] é formada por *CVS* com o *Vsup dar*, seleciona dois argumentos, sendo que o sujeito é Humano e o complemento é Não-humano.

A Tabela 5 apresenta as classes, juntamente com sua estrutura sintática típica, um exemplo característico da classe e o número de *CVS* pertencentes a cada uma. A seguir à tabela, apresentamos a descrição de cada classe.

Classe	**Estrutura**	**Exemplo**	**nº**
DH1	*Nhum dar Npred*	*Ana deu uma pirueta*	276
DN1	*N-hum dar Npred*	O ferro deu uma oxidada	106
DR1	*N± hum dar Npred*	*(Rui+O clima) deu uma refrescada*	94
DPC1	*Npc dar Npred*	*Meu estômago deu uma embrulhada*	14
DH2	*Nhum dar Npred Prep Nhum*	*Rui deu um castigo para a Ana*	272
DN2	*N-hum dar Npred Prep N-hum*	*A esponja deu uma absorvida na água*	6
DR2	*N± hum dar Npred Prep N± hum*	Ana deu uma entrelaçada no Rui Um fio deu uma entrelaçada no outro	15
DPC2	*Nhum dar Npred Prep Npc*	*Rui deu uma coçada na cabeça*	12
DHN2	*Nhum dar Npred Prep N-hum*	*Ana deu um ajuste na roupa*	254
DLC2	*Nhum dar Npred Prep Nloc*	*Rui deu uma passada na faculdade*	40
DQF2	*Nhum dar Npred Prep QueF*	*Rui deu um jeito de consertar o carro*	5
DHR2	Nhum dar Npred Prep N± hum	Ana deu destaque ao (Rui+problema)	253
D2LC	*Nloc dar Npred Prep N1*	*A casa deu abrigo ao (Rui+móvel)*	5
DRN2	*N± hum dar Npred Prep N-hum*	*(Ana+A lei) deu embasamento à teoria*	11
D3	*N0 dar Npred Prep N1 Prep N2*	*Ana deu um apelido de macaco ao Rui*	126
TOTAL			**1.489**

Tabela 5 – Classificação sintática das *CVS* com *dar*
Fonte: Elaboração própria

[**DH1**]: *CVS* com apenas um argumento obrigatoriamente de tipo humano e na posição sujeito (*N0=:Nhum*);

[**DN1**]: *CVS* com apenas um argumento obrigatoriamente de tipo não-humano e na posição sujeito (*N0=:N-hum*);

[**DR1**]: *CVS* com apenas um argumento que pode ser tanto humano quanto não-humano ou ainda admitir outros tipos sintático-semânticos (*Nloc*, *Npc*, *QueF*). O único argumento desses predicados é o sujeito da construção;

[**DPC1**]: *CVS* com apenas um argumento, que é obrigatoriamente um nome de parte do corpo na posição sujeito (*N0=:Npc*);

[**DH2**]: *CVS* com dois argumentos obrigatoriamente de tipo humano em ambas as posições argumentais (*N0=:Nhum* e *N1=:Nhum*);

[**DN2**]: *CVS* com dois argumentos obrigatoriamente de tipo não-humano em ambas as posições argumentais (*N0=:N-hum* e *N1=:N-hum*);

[**DR2**]: *CVS* com dois argumentos, sendo que ambos podem ser preenchidos por nome humano ou não-humano (*N0=:N± hum* e *N1=:N± hum*);

[**DPC2**]: *CVS* com dois argumentos, sendo que o primeiro, na posição sujeito, é obrigatoriamente de tipo humano, e o segundo, na posição de complemento, é obrigatoriamente um nome de parte do corpo (*N0=:Nhum* e *N1=:Npc*);

[**DHN2**]: *CVS* com dois argumentos, sendo que o primeiro, na posição sujeito, é humano e o segundo argumento, na posição de complemento, é não-humano (*N0=:Nhum* e *N1=:N-hum*)[52];

[**DLC2**]: *CVS* com dois argumentos, sendo que o primeiro, na posição de sujeito, pode ser de qualquer tipo (*Nhum*, *N-hum* ou outros), mas o segundo, na posição de complemento, é obrigatoriamente um nome locativo (*N0=:N± hum* e *N1=:Nloc*);

[**DQF2**]: *CVS* com dois argumentos, sendo que o primeiro, na posição de sujeito, é obrigatoriamente de tipo humano, e o segundo, na posição de complemento, é uma frase completiva (*N0=:Nhum* e *N1=:QueF*);

52 Em alguns casos, a posição de N_1 pode também ser preenchida por um *Npc*, mas não por *Nhum*, como se observa em *Ana deu uma esticada (*no Rui + na roupa + nas pernas)*.

[**DHR2**]: *CVS* com dois argumentos, sendo que o primeiro, na posição de sujeito, é obrigatoriamente de tipo humano, e o segundo, na posição de complemento, pode ser de diferentes tipos semânticos, podendo aceitar nomes humano, não-humano, parte do corpo, locativo *etc.* (*N0=:Nhum* e *N1=:N± hum*);

[**D2LC**]: *CVS* com dois argumentos, sendo que o primeiro, na posição sujeito, é um locativo, e o segundo, na posição de complemento, pode ser humano ou não-humano (*N0=:Nloc* e *N1=:N± hum*);

[**DRN2**]: *CVS* com dois argumentos, sendo que o primeiro, na posição de sujeito, pode ser de diferentes tipos semânticos, e o segundo, na posição de complemento, é de tipo não-humano (*N0=:N± hum* e *N1=:N-hum*);

[**D3**]: *CVS* com três argumentos, sem distinção entre as restrições de seleção desses argumentos.

10.2 Clusterização semântica das *CVS* a partir dos *Npred*

A clusterização semântica foi feita com base na análise dos traços semânticos comuns aos *Npred*. Com base nessa análise semântica dos traços, foi possível agrupar cerca de 1.000 *Npred*, distribuídos em cinco classes. Há cerca de outros 500 *Npred* que não entraram em nenhuma dessas classes por não apresentarem regularidades entre os traços semânticos que lhes permitissem ser agrupados. Na verdade, esses 500 *Npred* são nominalizações em *-ada*, derivadas de ações e/ou processos verbais que, apesar de possuírem semelhanças sintáticas, não possuem semelhança do ponto de vista de seus significados, tais como, *afunilada, agachada, agendada, agilizada, agitada, aglomerada* e vários outros.

As classes semânticas identificadas são: *Npred* relacionados a esportes, nomes de gêneros acadêmicos (tanto escritos quanto orais), nomes de documentos, *Npred* que designam atos violentos e nomes predicativos *dicendi* (de fala). Cada classe será explicada a seguir.

10.2.1 *Npred* de esportes

Os nomes predicativos que consistem em nomes de jogadas e/ou golpes em alguma modalidade esportiva merecem atenção especial porque muito raramente possuem um verbo ou um adjetivo morfologicamente associado, ou seja, quase todos os *Npred* de modalidades esportivas que constam na matriz (ver Apêndice A) são nomes autônomos,

ou porque são estrangeirismos e não existe em Português um verbo pleno equivalente, ou porque são termos exclusivos daquela modalidade esportiva.

(10.1) *Volante do Vasco* ***deu um carrinho*** *em D'Alessandro.* [Ex.R]

(10.2) *No Floripa Fight, Cacareco* ***deu um cruzado*** *em Riscado, colocou pra baixo, passou a guarda e aplicou um mata leão.* [Ex.R]

(10.3) *A holandesa Ranomi Kromowidjojo errou a chegada,* ***deu uma braçada*** *a mais, e ficou com a prata.* [Ex.R]

No *corpus*, encontramos referências às modalidades de luta (*dar uma gravata*, *dar um cruzado*), à esgrima (*dar uma estocada*), à natação (*dar uma braçada*), ao vôlei (*dar um ace*, *dar toco*, *dar dois toques*), ao basquete (*dar enterrada*), ao tênis (*dar um backhand*, *dar um forehand*), ao boliche (*dar um strike*), à corrida (*dar um sprint*) e várias referências ao futebol (*dar balão, bicicleta, bico, botinada, caneta, carrinho, chapéu, chute a gol, chuveirinho, drible, finta, goleada, olé, paradinha, passe, pedalada, rebote* e *tesoura*).

Alguns nomes predicativos, tais como *chute* (futebol), *bicuda* (futebol), *soco* (luta) e *murro* (luta) são nomes de atos de esportes, porém, como exprimem também atos violentos, não foram incluídos no rol de atos esportivos porque são muito frequentemente encontrados em textos de linguagem não especializada sem referência a nenhuma modalidade esportiva, como se verifica em:

(10.4) *A mulher de 106 anos que* ***deu um chute*** *na bunda de Lampião.* [Ex.R]

(10.5) *O Paulo Amaral* ***deu uma bicuda*** *em uma garrafa de café que alguém tinha deixado.* [Ex.R]

(10.6) *Douglas, do São Raimundo,* ***deu um soco*** *no rosto de Leonardo.* [Ex.R]

(10.7) *O homem simplesmente* ***deu um murro*** *no nariz do suposto ladrão.* [Ex.R]

Essas construções com *Npred* da classe de modalidades esportivas são tipicamente construções com verbo-suporte, apesar de não serem nominalizações de verbos. Para os autores que consideram a equivalência entre a *CVS* e uma construção com verbo pleno como relação necessária para a identificação das *CVS*, esses nomes de jogadas, golpes ou movimentos não seriam abrangidos.

10.2.2 *Npred* de gêneros acadêmicos

As *CVS* constituídas por um *Npred* da classe *gênero acadêmico* formam, juntamente com o *Vsup dar*, uma construção equivalente a um verbo de comunicação. (10.8) *O professor deu (um + uma) (apresentação + palestra + conferência + aula + curso + minicurso + workshop + simpósio) sobre Linguística para os alunos da graduação.* [Ex.C]

Essas construções possuem três argumentos essenciais: *N*0 (sujeito, humano, AGENT: *professor*), *N*1 (objeto indireto, complemento dativo, ADDRESSEE: *alunos*) e *N*2 (complemento indireto, TOPIC: *Linguística*). Nem sempre todos esses argumentos estão expressos na frase, mas, levando-se em conta o Princípio da Máxima Projeção, devemos considerar que todos eles são necessários à predicação de base.

Vale ressaltar que os nomes predicativos de gêneros acadêmicos, combinados com outros verbos-suporte, podem suprimir o argumento ADDRESSEE. Citem-se, por exemplo, as construções com o verbo *fazer*:

(10.9) *O professor fez (um + uma) (apresentação + palestra + conferência + aula + curso + minicurso + workshop + simpósio) sobre Linguística (<E> + para os alunos da graduação).* [Ex.C]

Barros (2014) considera, para as *CVS* com o *Vsup fazer*, apenas dois argumentos como essenciais: o AGENT e o TOPIC. Como nós levamos em consideração a máxima projeção dos argumentos, consideraremos três argumentos essenciais para os predicados de gêneros acadêmicos, independente do verbo-suporte com que ele se constrói.

10.2.3 *Npred* de documentos

Os nomes predicativos da classe de documentos (*alvará, atestado, declaração, autorização, licença, permissão, certidão, certificado*) merecem atenção especial porque nem sempre eles devem ser considerados nomes predicativos.

Esses nomes predicativos, além da sua construção predicativa, podem também designar objetos concretos, o que pode ser determinado em função da construção em que aparecem. Nesse sentido, é preciso distinguir entre as nominalizações propriamente ditas e os nomes que se "concretizam". Grefenstette e Teufel (1995) distinguem as nominalizações reais dos usos concretos das formas nominalizadas. O exemplo usado pelos autores é com o nome *proposta*, mas a mesma análise foi aplicada a todos os nomes da classe de documentos.

(10.10) *Ele fez uma proposta ao comitê.* [Ex.C]

(10.11) *Ele colocou a proposta na gaveta.* [Ex.C]

O primeiro exemplo é de fato uma construção com verbo-suporte, já que admite todos os testes definitórios das *CVS*. Assim, o nome *proposta* deve ser classificado como *Npred*. Já o segundo exemplo é tipicamente uma construção com o verbo pleno *colocar*, já que é um verbo distribucional e é ele que seleciona os argumentos (*N*0 = AGENT, *N*1 = OBJECT, *N*2 = LOCATIVE). Nesse segundo exemplo (10.11), o nome *proposta*, apesar de ser uma nominalização do verbo *propor*, "concretizou-se" e passou a designar um objeto concreto que se transfere de um local para outro. Construções como essa especificada no exemplo (10.11) não foram consideradas para efeitos de análise e descrição das *CVS*, porém o nome predicativo, quando seleciona o *Vsup dar*, entrou no rol de *Npred* a serem descritos.

10.2.4 *Npred* de atos violentos

Alguns nomes predicativos descritos na matriz designam *atos violentos* e já foram objeto de estudo por Baptista (2004). Esses *Npred* se fazem com o verbo *dar* na construção *standard* e, via de regra, o verbo *levar* na construção conversa. Além disso, todos eles selecionam nome humano AGENT para a posição de sujeito (*N*0=:*Nhum*) da construção *standard* e podem selecionar, para a posição de complemento, um *Nhum* ou um *Npc*.

Dada essa diferença, os *Npred* de atos violentos podem ser divididos em dois grupos: (i) os que exigem exclusivamente *Nhum* na posição de complemento; e (ii) os que aceitam tanto *Nhum* quanto *Npc* na posição de complemento. O primeiro grupo de *Npred* foi classificado na classe DL22 por Baptista (1997) para o PE e, posteriormente, por Rassi *et al.* (2015) para o PB; já os *Npred* do segundo grupo foram classificados pelos mesmos autores na classe DL3.

A classe DL22 é formada por *CVS* com o *Vsup dar* que fazem a conversão com *levar*. Essa classe só aceita *Nhum* (mas não *Npc*) tanto na posição sujeito quanto na posição de complemento. É o caso de:

(10.12) *Jennifer deu uma surra em Alessandra.* [Ex.R]

(10.12a) **Jennifer deu uma surra no braço da Alessandra.* [Ex.C]

(10.13) *O dia em que o Homem Aranha deu uma coça no Batman.* [Ex.R]

(10.13a) **... O Homem Aranha deu uma coça na perna do Batman.* [Ex.C]

Os nomes do tipo *surra* constituem um pequeno conjunto na classificação do PE. Nomes predicativos como *açoite, coça, pancada, porrada, sova* e *surra* pertencem à classe DL22 tanto em PE como em PB. Além desses, existem dois *Npred* em PE que não existem em PB (*tareia* e *trepa*), e outros três *Npred* que existem em PB, mas não se verificam em PE (*ataque*, *bote* e *pedalada*). Ressalte-se que o nome *pedalada* possui dois significados e por isso foi duplicado na matriz (ver Apêndice A). O nome *pedalada*, que integra a classe DL22, consiste no ato violento de "dar um tapa na cabeça de outrem", como se demonstra em:

(10.14) *O jogador disse, em tom de brincadeira, que* ***deu uma pedalada*** *no psicólogo que o ajuda a resolver os problemas.* [Ex.R]

Já o segundo tipo de nomes que designam atos violentos foi agrupado na classe DL3, que reúne os *Npred* que admitem, na construção *standard*, tanto *Nhum* quanto *Npc* na posição de complemento em atos violentos. Esses *Npred* são muito mais numerosos e podem ainda se subdividir em três tipos. Cada um deles será mencionado a seguir, mas uma análise mais refinada de cada membro dessas classes foi feita em Rassi *et al.* (2015).

- classe DL31: *Npred* derivados de nomes de instrumentos (*agulhada, alfinetada, bengalada, chicotada, almofadada, botinada, cadeirada, canivetada, estilingada, flechada, raquetada, travesseirada, tesourada*, dentre outros);
- classe DL32: *Npred* derivados de nomes parte do corpo (*cabeçada, cornada, cotovelada, dentada, focinhada, joelhada, palmada, patada, trombada, narigada, ombrada, peitada, pernada, pezada* e *unhada*);
- classe DL33: *Npred* não derivados de nomes de instrumentos ou *Npc* (*bofetada, coice, corte, golpe, mordida, pisada, empurrão*, dentre outros).

10.2.5 *Npred* de fala

Os verbos *dicendi*, também chamados de "verbos de fala", são os verbos que introduzem diferentes turnos no discurso relatado ou indicam aquilo que se diz. Não é comum encontrarmos a classificação do verbo *dar* como um verbo *dicendi*, porém o verbo *dar*, enquanto verbo-suporte, muitas vezes é selecionado por nomes predicativos que correspondem a ações de *dizer*. Nesse sentido, o verbo *dar* pode ser substituído por *falar*, *dizer*, *desejar* ou suas variantes.

Gaston Gross (1989) já havia indicado alguns verbos de fala no Francês que podem assumir o papel de verbo-suporte, tais como *intimer* 'intimar', *dicter* 'ditar', *infliger* 'infligir' e *signifier* 'significar'. Giry-Schneider (1994) acrescenta a essa lista mais um bom número de variantes do verbo *dire*. Em seguida, Baptista (2010) também analisou a estrutura dos verbos *dicendi*, que pode ser adaptada à análise dos nomes predicativos *dicendi*.

Os nomes predicativos que figuram na posição de *N*1 dos verbos de fala, por serem "verdadeiros pivôs sintáticos e semânticos da frase" (GIRY-SCHNEIDER, 1994, p.114), funcionam como complementos apropriados dos verbos *dicendi*. A autora afirma que algumas variantes de *dizer* podem funcionar como extensões dos verbos-suporte *dar* ou *fazer*:

(10.15) (10.15) *Max (deu + vociferou) uma informação ao Jean.* [Ex.C]

(10.16) *Max (fez + vociferou) uma profecia ao Jean.* [Ex.C]

Por essa razão, consideraremos o verbo *dar* como um tipo especial de verbo-suporte, que pode funcionar como um verbo-suporte de nomes de fala (ou nomes *dicendi*), quando acompanhado de nomes predicativos que expressam ações de *dizer*.

Alguns verbos de fala compartilham as mesmas propriedades dos verbos-suporte *dar* e *fazer*. Tomando-se como exemplo a frase de base *Max disse umas (palavras corteses + cortesias) à Lea*, adaptada de Giry-Schneider (1994), identificam-se as seguintes propriedades:

1. A correferência entre os sujeitos:

   *?*Max disse umas cortesias do Jean para a Lea.* [Ex.C]

   Max notou umas cortesias do Jean para a Lea. [Ex.C]

2. A nominalização adjetival

   Max disse para a Lea umas palavras que eram de certa cortesia. [Ex.C]

3. A formação de grupo nominal complexo a partir da redução do *Vsup*

   Max disse para a Lea umas palavras de cortesia. [Ex.C]

4. A redução do *N=:palavras*

   Max disse para a Lea umas cortesias. [Ex.C]

Além dessas propriedades, verifica-se também que o verbo *dizer* se funde com o verbo-suporte *dar* (ou *fazer*). Na frase *Max gritou (= disse gritando) uma informação para o Jean*, o verbo *dicendi gritar* equivale a *dizer de forma gritada*. Esses verbos, que são extensões do verbo *dizer*, fundem-se com os verbos-suporte *dar* e *fazer*, quando empregados com certos nomes (*Max gritou essa informação = Max deu essa informação gritando* ou *Max exclamou uma recomendação = Max fez uma recomendação exclamando*).

O verbo *dar*, em si mesmo, não deve ser classificado como um verbo-suporte de fala, mas os nomes predicativos que expressam atos *dicendi* podem ser classificados semanticamente como *Npred de fala*. Há 28 *Npred* de fala na matriz, tais como: *amém*, *bis*, *boa-noite*, *depoimento*, *parabéns* e vários outros. Após justificar a classificação de algumas *CVS* como um tipo especial de construção *dicendi*, dividimos as construções com o *Vsup dar* e *Npred* de fala em três grandes grupos, de acordo com o tipo de nome predicativo com o qual se constrói: (i) nomes predicativos que são também nomes operadores; (ii) interjeições que migraram para a classe dos nomes predicativos; e (iii) nomes predicativos que não são complementos apropriados.

10.2.5.1 Nomes predicativos enquanto nomes operadores

À época, Giry-Schneider (1994) tratou alguns nomes predicativos como nomes operadores (*Nop*), por serem substantivos que exigem como complemento nominal uma frase completiva. O conceito de nome operador (*Nop*) foi introduzido por Leclère (1971) e foi retomado por Giry-Schneider (1994), ao tratar dos verbos de fala. Nomes como *notícia*, *fato* e *ideia* são considerados nomes operadores porque exigem um complemento nominal interno, que geralmente é uma frase completiva, como: *a **notícia** da volta do Rui*, *o **fato** de Rui ter voltado*, *a **ideia** de Rui voltar*. Ao mesmo tempo, esses *Nop* funcionam como complementos dos verbos de fala, podendo introduzir a completiva ou simplesmente substituí-la.

(10.17) *O Pedro contou à Ana (um + uma) (notícia + fato + ideia).* [Ex.C]

(10.18) *O Pedro comentou com a Ana (a notícia da volta do Rui + o fato de Rui ter voltado + a ideia de Rui voltar).* [Ex.C]

(10.19) O Pedro comentou com a Ana (a volta do Rui + o Rui ter voltado + o Rui voltar). [Ex.C]

Giry-Schneider (1994) apresenta várias justificativas para esses nomes, que são também nomes predicativos, serem considerados nomes operadores. Uma das

justificativas é de ordem empírica atestando que a grande maioria dos verbos de fala aceita melhor os nomes operadores como complementos do que as próprias frases completivas. A autora conclui que o que foi apresentado por Leclère (1971) e Gross (1975) como sendo fatos isolados deve ser considerado como regra geral: a estrutura com complemento operador apropriado é mais geral que a forma simples com completiva ou complemento nominal.

Além do nome *notícia*, que é um complemento apropriado do verbo *dar* enquanto verbo-suporte de fala, podem-se referir também outros nomes, como: *aconselhamento*, *anuência*, *anúncio*, *afirmação*, *argumento*, *boa-nova*, *comando*, *comentário*, *comprovação*, *confirmação*, *conselho*, *declaração*, *determinação*, *esclarecimento*, *explicação*, *informação*, *instrução*, *justificativa*, *opinião*, *ordem*, *orientação*, *palpite*, *recomendação*, *sugestão* e outros.

O que nos permite classificar todos esses nomes como *Nop*, além de serem *Npred*, é o fato de que eles são predicados que exigem um complemento interno (complemento nominal na forma de uma completiva). Nas construções com verbos de fala variantes do *dizer*, o *Nop* pode introduzir uma completiva, substituir a completiva, ou ser suprimido, restando apenas a completiva:

(10.20) *O Pedro contou a notícia de que Rui voltara à Ana.* [Ex.C]

= *O Pedro contou a notícia à Ana.* [Ex.C]

= *O Pedro contou que Rui voltara à Ana.* [Ex.C]

Já nas construções com *Npred dicendi*, o *Nop* pode introduzir uma completiva ou substituí-la, mas não pode ser suprimido:

(10.21) *O Pedro deu a notícia de que Rui voltara à Ana.* [Ex.C]

= *O Pedro deu a notícia à Ana.* [Ex.C]

= **O Pedro deu que Rui voltara à Ana.* [Ex.C]

O *Nop*, que é também *Npred*, não pode ser suprimido nas construções com verbo-suporte justamente porque ele é o elemento predicador principal da *CVS*, uma vez que o verbo-suporte é semanticamente fraco para predicar.

10.2.5.2 Expressões interjetivas que migram para a classe dos *Npred*

O segundo grupo de nomes predicativos que também selecionam o verbo-suporte *dar* inclui palavras que são originalmente pertencentes à classe das interjeições (ou

expressões de rotina), porém podem ser interpretadas como substantivos nas *CVS*. Fazem parte desse grupo construções como:

(10.22) *Jimmy Carter ouviu Rosemary cantar "Feelings" e* ***deu bis****.* [Ex.R]

(10.23) *O presidente do secretariado, Sr. Lindani Mthethwa,* ***deu as boas-vindas*** *a todos os presentes.* [Ex.R]

(10.24) *Lady Gaga* ***deu um bom-dia*** *aos seus quase 4 milhões de seguidores.* [Ex.R]

As palavras *bis*, *boas-vindas* e *bom-dia* constituem o próprio dizer, a forma proferida pelo sujeito AGENT-SPEAKER na posição de *N*0. As palavras *bis*, *boas-vindas* e *bom-dia* pertencem originalmente à classe das interjeições, pois são exclamações curtas que possuem significado de uma frase completa. Essas expressões interjetivas poderiam ser parafraseadas, respectivamente, por:

(10.22a) *Jimmy Carter ouviu Rosemary cantar "Feelings" e* ***pediu que ela cantasse novamente****.* [Ex.C]

(10.23a) *O presidente do secretariado, Sr. Lindani Mthethwa,* ***desejou*** *a todos os presentes* ***que eles fossem bem-vindos****.* [Ex.C]

(10.24a) *Lady Gaga* ***exclamou*** *aos seus quase 4 milhões de seguidores* ***que eles tenham um bom-dia****.* [Ex.C]

A frase completiva que funciona como complemento dos verbos *dicendi pedir*, *desejar* e *exclamar* é sinônima das interjeições (*bis*, *boas-vindas* e *bom-dia*), que expressam o mesmo sentido da frase completiva. Elas possuem as mesmas características das frases com outros verbos *dicendi*, porém a única diferença é que seus complementos podem ser substituídos por interjeições.

Nas frases em que essas expressões selecionam o verbo-suporte *dar*, a interjeição passa a assumir o papel de nome predicativo, pois a construção exige a presença de um elemento predicador, que não pode ser o verbo, dada a sua natureza de "auxiliar". Assim, as interjeições "migram" para a classe dos nomes predicativos, apresentando as mesmas propriedades de outras construções com verbo-suporte:

1. Correferência entre sujeitos:

A plateia deu bis ao cantor. [Ex.C]

**A plateia deu o bis da Lady Gaga ao cantor.* [Ex.C]

A plateia aplaudiu o bis da Lady Gaga ao cantor. [Ex.C]

2. Restrições sobre os determinantes

*A plateia deu (seu + *meu + *teu) bis ao cantor.* [Ex.C]

3. Formação de grupo nominal a partir da redução do *Vsup*

O bis da plateia ao cantor foi atendido. [Ex.C]

4. Substituição do *Vsup dar* por outras variantes *dicendi*

A plateia (pediu + gritou) bis ao cantor. [Ex.C]

Também fazem parte desse grupo outros nomes como *amém*, *boa-noite*, *boa-tarde*, *boas-festas*, *ciao*, *foda-se*, *glória-a-Deus*, *graças-a-Deus*, *melhoras*, *morras*, *oi*, *olá*, *ok*, *tchau*, *urras*, *vivas*, *xeque-mate*[53] e outros. Todos eles foram inseridos na matriz (ver Apêndice A), por se tratar de construções com verbo-suporte.

10.3 O verbo-suporte de ocorrência

Esta seção apresenta um tipo particular de construção com verbo-suporte, a que Gross (1986, p.107-109) chamou de verbo-suporte de ocorrência (*Vsup-oc*). Os *Vsup-oc* instauram uma ocorrência, um acontecimento, ligando um evento a seu circunstancial. Essa categoria engloba verbos como: *acontecer*, *ter lugar*, *ocorrer*, *decorrer*, *suceder*, *etc*. Tipicamente, o verbo *dar-se*, em construções como (10.25), também é classificado como *Vsup-oc*.

(10.25) *A discussão se deu no programa "Última Palavra".* [Ex.R]

O verbo *dar-se* é considerado verbo-suporte de ocorrência porque o elemento predicativo da frase é, na verdade, o nome *discussão*, e essa construção se diferencia das demais *CVS* apresentadas nas seções anteriores porque o *Npred* não ocupa a posição sintática de objeto direto do verbo, mas sim, de sujeito. O verbo *dar-se* serve apenas para introduzir o elemento circunstancial posposto ao verbo (*no programa "Última Palavra"*). Em Português, essa construção pode sofrer uma transformação para:

(10.25a) *Deu uma discussão no programa "Última Palavra".* [Ex.C]

As duas construções expressam a ocorrência de um evento e, portanto, o nome predicativo *discussão* é um nome eventivo, mas há diferenças sintáticas e semânticas

53 Vale ressaltar que a construção *dar xeque-mate* pode ser usada em sentido literal, indicando que o jogador fez uma jogada no xadrez e, em seguida, proferiu a palavra *xeque-mate* ou também pode ser usada no sentido figurado, como no caso de *A direção do Gama considera que deu o xeque-mate no Clube dos 13*.

entre elas: (i) o verbo em (10.25) é pronominal (verbo *dar-se*), enquanto em (10.25a) é simples (verbo *dar*); (ii) em (10.25), o nome predicativo aparece em posição pré-verbal, enquanto em (10.25a) o nome predicativo aparece em posição pós-verbal; (iii) o constituinte *a discussão* funciona como tema, em (10.25), mas como rema, em (10.25a); e (iv) dada a diferença entre tema e rema, o determinante, em (10.25), é definido, ao passo que, em (10.25a), é indefinido.

A discussão sobre os conceitos de tema e rema foge ao escopo desta pesquisa, mas, em linhas gerais, considera-se que o tema é o elemento dado, já conhecido, enquanto o rema é o elemento novo, que está sendo introduzido no momento da enunciação. Para um estudo aprofundado sobre tema e rema, ver, dentre outros, Halliday (1967) e Prince (1981).

Em suma, há duas construções sintáticas diferentes com o *Vsup-oc dar*. Por um lado, destacam-se as construções como (10.26) e (10.27), que podem ser formalizadas como [*Vdar* (*E+Det*) *Ni*]; e, por outro lado, construções como (10.28) e (10.29), formadas por [*Ni Vdar-se*]:

(10.26) *Os seguranças tentaram separar e* ***deu uma confusão****.* [Ex.R]

(10.27) *E à meia noite* ***deu uma briga*** *lá no canto quebraram três, quatro bancos nas costas do meu irmão.* [Ex.R]

(10.28) *O secretário confirmou que* ***a confusão se deu*** *porque "eles [fiscais] estavam exigindo a pulseira".* [Ex.R]

(10.29) *De acordo com populares,* ***a briga se deu*** *por conta de intrigas.* [Ex.R]

Smarsaro e Rocha (2011) descreveram as construções impessoais com o verbo *dar* na estrutura [*dar x em y*]. Trata-se de construções do tipo *Deu infiltração na laje*, *Deu febre em Pedro* e *Deu mofo no guarda-roupa*. Seguindo a proposta do Léxico-Gramática, Smarsaro e Rocha (2011, p.52) analisam essas construções em termos de transformações sintáticas e consideram "o verbo *dar* como suporte, todavia sem a presença do sujeito". As autoras caracterizam as estruturas *Deu frutas no sítio do João*, *Deu piolho na cabeça da criança* e *Deu infiltração na laje* como estruturas de verbo-suporte que revelam uma variação com os verbos *ter* e *estar com*, por meio de transformações sintáticas, conforme se demonstra em:

(10.30) ***Deu*** *frutas no sítio do João.* [Ex.C]

(10.30a) *O sítio do João* ***tem*** *frutas.* [Ex.C]

(10.30b) *?O sítio do João* ***está com*** *frutas.* [Ex.C]

(10.30) ***Deu*** *piolho na cabeça da criança.* [Ex.C]

(10.31a) *A cabeça da criança* ***tem*** *piolho.* [Ex.C]

(10.31b) *A cabeça da criança* ***está com*** *piolho.* [Ex.C]

(10.30) ***Deu*** *infiltração na laje.* [Ex.C]

(10.32a) *A laje* ***tem*** *infiltração.* [Ex.C]

(10.32b) *A laje* ***está com*** *infiltração.* [Ex.C]

Para classificar essas construções como impessoais, é preciso retomar o conceito de impessoalidade. São chamadas *construções impessoais* aquelas que não admitem sujeito. Se considerarmos que o sujeito é o constituinte com o qual o verbo concorda em número, podemos testar a aceitabilidade dos constituintes pós-verbais na posição de sujeito plural.

(10.30c) ***Deram*** *frutas no sítio do João.* [Ex.C]

(10.31c) ***Deram*** *piolhos na cabeça das crianças.* [Ex.C]

(10.32c) ***Deram*** *infiltrações na laje* [Ex.C]

Todos os exemplos são construídos, mas essa possibilidade de concordância do verbo com o nome predicativo também se verifica em frases reais de *corpus* com outros nomes predicativos, por exemplo:

(10.33) *a gente foi pra ver filme mas* ***deram*** *confusões e a gente acabou não fazendo nada.* [Ex.R]

(10.34) *Infelizmente* ***deram*** *uns problemas no vídeo que seria para hoje e este só poderá ser enviado amanhã.* [Ex.R]

Dado o fato de que o verbo *dar* pode concordar em número com o nome predicativo que o acompanha, não o classificaremos como impessoal, mas admitiremos que o *Npred* seja o sujeito desse verbo-suporte de ocorrência.

Segundo Gross (1986, p.107-108), os *Vsup-oc* servem para introduzir valores adverbiais, principalmente de tempo e de lugar, aos nomes eventivos. No entanto, frases como (10.30) e (10.31) não podem ser classificadas como construções com o verbo-suporte

de ocorrência porque *frutas* e *piolho* não são nomes eventivos. Além disso, o verbo *dar* não pode ser substituído por outros *Vsup-oc*, tais como:

(10.30) ***Deu** frutas no sítio do João.* [Ex.C]

(10.30d) *****Ocorreram** frutas no sítio do João.* [Ex.C]

(10.30e) **As frutas no sítio do João **se deram** por causa da chuva.* [Ex.C]

(10.31) ***Deu** piolho na cabeça da criança.* [Ex.C]

(10.31d) *****Ocorreu** piolho na cabeça da criança.* [Ex.C]

(10.31e) **O piolho na cabeça da criança **se deu** por falta de higiene.* [Ex.C]

Por outro lado, a construção *Deu infiltração na laje* (10.32) poderia ser classificada como construção com o verbo-suporte de ocorrência *dar*, já que *infiltração* é um nome eventivo e o verbo *dar* pode comutar com outros *Vsup-oc*:

(10.32) ***Deu** uma infiltração na laje.* [Ex.C]

(10.32d) ***Ocorreu** uma infiltração na laje.* [Ex.C]

(10.32e) *A infiltração na laje **se deu** por falta de planejamento.* [Ex.C]

Conclui-se, portanto, que estão em jogo diferentes estruturas sintáticas e, portanto, devem ser analisadas isoladamente caso a caso.

Durante a análise das construções causativas (ver Capítulo 5), foi evidenciada a possibilidade de algumas construções causativas também derivarem construções com essa mesma estrutura, tais como:

(10.35) *(Rui + Aquela caveira + O fato de Zé se atrasar + Que o Zé tenha se atrasado)* ***deu** (aflição + emoção + medo) em Ana.* [Ex.C]

(10.35a) ***Deu** (uma aflição + uma emoção + um medo) em Ana.* [Ex.C]

(10.35b) *(O medo + A aflição + A emoção) da Ana **se deu** pela chegada do Rui.* [Ex.C]

Na transposição da frase causativa (10.35) para a frase (10.35a), houve o apagamento do sujeito causal. Já na transposição de (10.35) para (10.35b), o elemento causal passa da posição de sujeito para a posição de circunstancial.

Em uma outra abordagem, Franchi, Negrão e Viotti (1998, p.114) chamam os verbos dessas construções de *verbos apresentacionais* porque desempenham o papel de

"apresentar" o tópico ou assunto que se pretende introduzir na conversa. Partindo das noções de *função apresentacional* e *foco apresentacional*, Franchi, Negrão e Viotti (1998) usam o termo *verbo apresentacional* para se referirem a uma categoria de verbos que expressam um acontecimento, uma ocorrência, uma existência.

Os autores dividem os verbos apresentacionais em duas classes: uma de sentenças existenciais e outra de sentenças ergativas. As existenciais, como o próprio nome sugere, indicam a existência da coisa, fato ou tópico conversacional que se deseja introduzir. Sua função é declarar que o referente do *SN* existe e que ele será o novo tópico. Os autores citam verbos como *ter* (*e.g.* ***Tem*** *muita gente na festa*), *ser* (*e.g.* ***Era*** *uma vez, uma princesa*), *fazer* (***Faz*** *muito tempo que não nos vemos*), *haver* (***Há*** *muito dinheiro nos bancos*) e outros como exemplos de verbos apresentacionais existenciais. Não foram encontrados na literatura trabalhos que refiram o verbo *dar* como apresentacional, porém os exemplos de verbos apresentacionais existenciais podem ser adaptados para o verbo *dar*.

A análise da frase com *ter* (***Tem*** *muita gente na festa*) se aplica a construções com *dar* (***Deu*** *mil pessoas na festa*), que é interpretada como existencial (*e.g. Havia mil pessoas na festa*), mas não como ocorrência (*e.g. *(Ocorreram + Sucederam + Aconteceram) mil pessoas na festa*).

Com relação aos verbos apresentacionais em sentenças ergativas, ocorre a posposição do sujeito ao verbo (FRANCHI; NEGRÃO; VIOTTI, 1998, p.114), sendo o sujeito um argumento não controlador, de propriedades típicas de objeto direto dos verbos transitivos. Em outras palavras, a função de objeto direto se confunde com a do sujeito de um verbo intransitivo. Os autores citam como exemplo de construção ergativa com verbos apresentacionais:

(10.36) ***Chega*** *uma hora em que a mente alcança um plano mais alto.* [Ex.C]

(10.37) *Começaram a* ***aparecer*** *os problemas no casamento.* [Ex.C]

A análise dessas construções com os verbos apresentacionais ergativos *chegar* e *aparecer* também pode ser adaptada para a análise de construções impessoais com o verbo *dar*, em frases como:

(10.36a) ***Deu*** *a hora em que a mente alcança um plano mais alto.* [Ex.C]

(10.37a) *Começou a* ***dar*** *problemas no casamento.* [Ex.C]

Laura (2013, p.50) analisa essas construções e conclui que "esses verbos não predicam nem o sujeito nem o argumento interno, ou seja, não lhes atribuem traços

procedentes de suas propriedades intencionais". Sua função se resume a apresentar o novo tópico discursivo, que é representado pelo sintagma nominal precedido pelo verbo apresentacional.

As construções com o verbo-suporte de ocorrência *dar* (ou chamado verbo apresentacional) também podem ser classificadas nessas duas categorias: (i) com valor existencial (*e.g. Deu infiltração na laje*); e (ii) em sentenças ergativas, com valor temporal (*e.g. Deu duas horas*). Para atestar essa subdivisão, recorremos a comparações com outros verbos existenciais (como *haver*) e temporais (como *ser*).

(10.38) ***Deu*** *infiltração na laje.* [Ex.R]

(10.38a) *(**Há + *É**) infiltração na laje.* [Ex.C]

(10.39) ***Deu*** *uma hora.* [Ex.R]

(10.39a) *(***Há + É**) uma hora.* [Ex.C]

O verbo-suporte de ocorrência *dar* também introduz os nomes predicativos que indicam fenômenos meteorológicos. Ruwet (1986) tratou dos verbos meteorológicos em diferentes línguas, considerando verbos como *chover*, *nevar*, *trovejar*, *esfriar*, dentre outros. Muitos verbos meteorológicos podem ser nominalizados – *chuva*, *neve*, *trovão*, *frio* – e constituir uma locução juntamente com o verbo *dar* ou o verbo *fazer*.

(10.40) ***Deu*** *uma chuva forte no Rio de Janeiro.* [Ex.C]

(10.41) ***Deu*** *um trovão ontem à noite.* [Ex.C]

Nas construções nominais, o nome meteorológico é o predicador da frase, assim o verbo *dar* é classificado como verbo-suporte de ocorrência. Nomes como *trovoada*, *trovejada*, *trovão*, *raio*, *tempestade*, *enchente*, *névoa*, *ventania*, *calor etc.* são predicativos e entram em construções com *Vsup-oc*, porém não estão descritos na matriz do L-G porque não são *CVS* comuns.

Este capítulo encerra a Parte II do livro, apresentando uma classificação das *CVS* com o verbo *dar* em Português Brasileiro. A classificação proposta identificou 15 classes sintáticas, dependendo principalmente do número de argumentos e do tipo semântico desses argumentos. Além da classificação sintática, que é formal e pode ser reproduzida para outros verbos, este capítulo apresentou uma proposta de agrupamento semântico das *CVS* com base nos significados dos nomes predicativos. Por fim, apresentamos também um tipo particular de *CVS*, cujo verbo é chamado de verbo-suporte de ocorrência.

As propriedades particulares desses verbos-suporte de ocorrência também foram abordadas neste capítulo.

A partir do próximo capítulo, inicia-se a Parte III deste livro, que trata do processamento automático das *CVS* em um *parser*.

Parte III

Processamento computacional dos predicados nominais

11 O processamento das construções com verbo-suporte

Neste capítulo, apresentaremos brevemente o estado da arte do processamento automático de construções com verbo-suporte, tanto em Português como em outras línguas. A falta de uma abordagem satisfatória para o processamento desse tipo de expressão multipalavra justifica a proposta deste trabalho, de analisar sintaticamente as relações que se estabelecem entre os constituintes de uma *CVS*. Ainda neste capítulo, apresentaremos a arquitetura geral da STRING (MAMEDE *et al.*, 2012), que é uma cadeia de processamento do Português, usada para testar a análise sintática automática das *CVS*.

11.1 Considerações preliminares

Os predicados nominais do tipo *construções com verbo-suporte*, por vezes chamados na Linguística Computacional ou no PLN de *expressões multipalavras*, *predicados complexos* ou *construções com verbo leve*, já foram objeto de processamento automático, principalmente na área da tradução automática (BUTT, 1994; WANG; IKEDA, 2008; VINCZE, 2012). Nesses trabalhos, porém, as *CVS* são analisadas como um bloco único cujos constituintes são relativamente fixos, tal como um subtipo de expressão multipalavra.

Não foi encontrado na literatura nenhum trabalho que refira o processamento automático das *CVS* em *parser* (analisador sintático automático) e que considere as complexas relações sintáticas que se estabelecem entre seus constituintes. Um dos objetivos desta pesquisa é justamente propor a análise sintática automática das *CVS* que leve em consideração sua estrutura sintática complexa. As *CVS* seguem as mesmas regras de formação e transformação da gramática geral da língua, mas possuem propriedades específicas e admitem sistematicamente várias transformações sintáticas, ainda que algumas delas sejam lexicalmente determinadas.

Ao contrário de outras abordagens que visam ao processamento das *CVS* para uma aplicação específica (como a tradução automática), este trabalho propõe uma análise sintática automática geral, a ser utilizada em qualquer aplicação computacional, ou seja, independente de domínio ou área.

Os *parsers* disponíveis para o Português, de que se tenha conhecimento, não possuem ainda informações sobre predicados nominais formados por verbo-suporte e nome predicativo. O *parser* PALAVRAS (BICK, 2000) é considerado um dos melhores analisadores

sintáticos para o Português do Brasil (MAZIERO; PARDO; NUNES, 2007), com precisão de 95% na etiquetagem de categorias morfossintáticas, mas esse *parser* não distingue construções verbais (com verbo pleno) de construções nominais (com verbo-suporte).

Conforme apontado e amplamente defendido na Parte II, é necessário analisar as *CVS* como predicados nominais em que o predicador é o nome, e não o verbo. Essa diferença deve ser levada em conta porque a estrutura sintática, a estrutura de constituintes, a estrutura de dependências e os papéis temáticos dos argumentos das *CVS* são diferentes das construções verbais com verbo pleno.

Nesta parte do livro, apresentamos a metodologia e os resultados da integração dos dados linguísticos do Léxico-Gramática (L-G) em uma cadeia de processamento automático do Português. Para tanto, (i) os dados descritos na Parte II foram formalizados em uma matriz do L-G; (ii) as regras de extração de dependências das *CVS* foram geradas automaticamente a partir das informações constantes na matriz do L-G; (iii) um *corpus* de referência foi produzido e anotado manualmente com *CVS* para servir como *golden standard* para a avaliação da performance do sistema; (iv) uma amostra do *corpus* foi processada na cadeia de processamento do Português, e ela corresponde à mesma amostra anotada manualmente; (v) o resultado do sistema foi avaliado com base nos resultados da anotação manual do *corpus* de referência; (vi) os principais erros encontrados na saída do sistema, após o processamento automático foram identificados e corrigidos; e (vii) após a análise de erros, foi feita uma segunda avaliação da performance do sistema.

Vale ressaltar que a análise, a descrição e a classificação das *CVS* em Português Brasileiro foram feitas em três trabalhos diferentes, no âmbito de três doutoramentos: Barros (2014) descreveu e classificou 1.815 predicados nominais formados pelo verbo-suporte *fazer*; Santos (2015) descreveu e classificou 2.273 predicados nominais com o verbo-suporte *ter*; e, nesta pesquisa, apresentamos 1.489 predicados nominais com o verbo-suporte *dar*. Todos esses dados foram sistematicamente analisados, descritos e formalizados em matrizes do Léxico-Gramática (L-G) das construções nominais do Português. Apesar de haver descrição das *CVS* com esses três verbos (*fazer*, *ter* e *dar*), esta pesquisa apresenta exclusivamente os resultados da inserção das construções nominais com o verbo-suporte *dar* na cadeia de processamento do Português STRING (MAMEDE *et al.*, 2012).

Os dados das construções com o verbo-suporte *dar* foram inseridos na STRING, concretamente como um dos componentes da gramática do Português, implementada no XIP (*Xerox Incremental Parser*), que é o analisador sintático automático utilizado

pela cadeia (MOKHTAR; CHANOD; ROUX, 2002). O XIP utiliza regras para extração de dependências entre as cabeças dos constituintes sintáticos (*chunks*); o objetivo é que o *parser* extraia automaticamente uma dependência (SUPPORT) entre o *Vsup* e o *Npred*. Essa dependência depois é utilizada para analisar adequadamente as relações entre o *Npred* e seus argumentos.

Em trabalho anterior (RASSI *et al.*, 2014), 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.

11.2 Cadeia de processamento automático do Português - STRING

A STRING é uma cadeia híbrida de processamento de língua natural para o Português, baseada tanto em métodos de base estatística como em processamento por regras. A cadeia possui uma estrutura modular, que executa as principais tarefas básicas de PLN, como a tokenização, a segmentação do texto em frases, a etiquetagem morfossintática dos *tokens* com suas respectivas partes-do-discurso (POS *tagging*), a desambiguação morfossintática (por regras e estatística), a análise sintática superficial (*chunking*) e a análise sintática profunda (*deep parsing*). A arquitetura geral da STRING é apresentada na Figura 6:

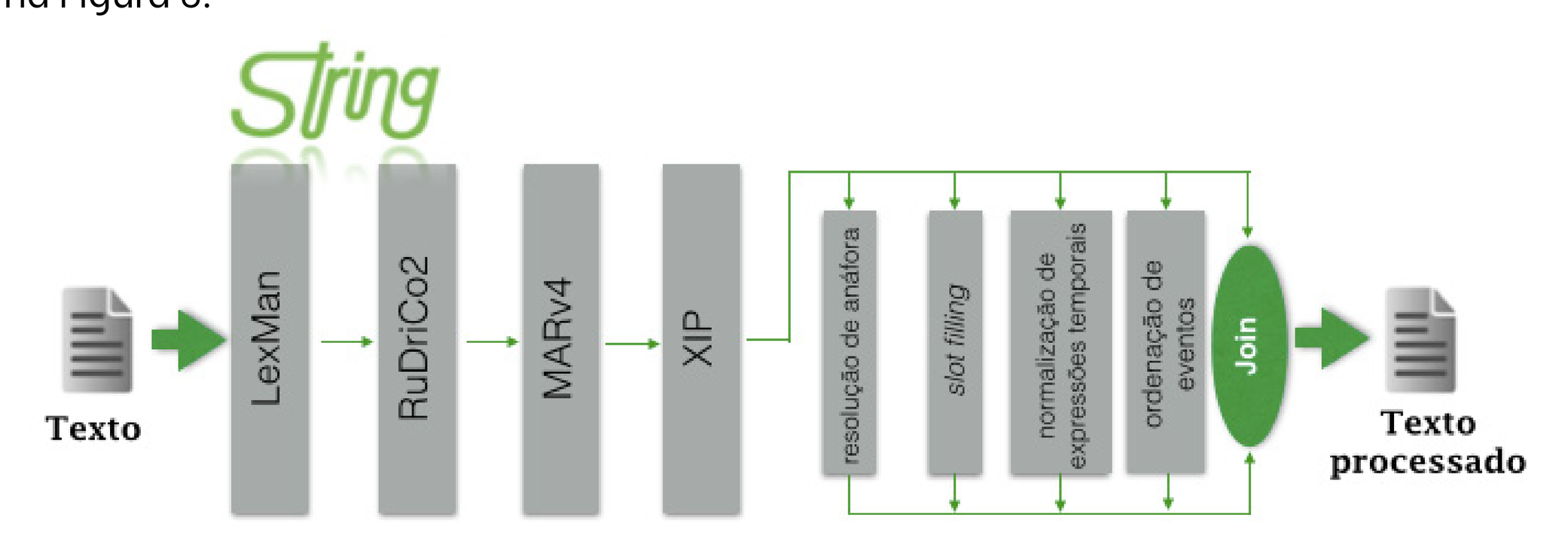

Figura 6 – Arquitetura geral da STRING
Fonte: Adaptado de https://string.l2f.inesc-id.pt/w/index.php/Architecture.

Cada módulo da cadeia executa uma tarefa específica[54]. O primeiro módulo (LexMan) recebe o texto para processar e executa as seguintes tarefas: (i) tokeniza o texto, ou seja, define os segmentos que formam o texto (palavras, pontuação *etc.*); (ii) etiqueta morfossintaticamente os segmentos, ou seja, recebe o resultado da tokenização como entrada e associa todas as possíveis etiquetas de partes-do-discurso (POS) a cada segmento; e (iii) segmenta o texto em sentenças, ou seja, indica o início e o fim de cada sentença. Este módulo também executa tarefas de reconhecimento de números e datas e pode resolver alguns problemas de ambiguidade morfossintática.

O segundo módulo (RuDriCo2) é um desambiguador morfológico baseado em regras e executa parte da desambiguação das etiquetas atribuídas no módulo anterior. O principal objetivo do RuDriCo2 é ajustar os resultados produzidos pelo analisador morfológico de acordo com as necessidades de cada *parser*. Para tanto, ele faz alterações de segmentação dos dados de entrada, por exemplo, para juntar segmentos, no caso de palavras compostas; assim, ele pode modificar a segmentação feita no LexMan.

O terceiro módulo (MARv4) também é um desambiguador morfossintático, porém é baseado em modelos estatísticos. Esse módulo recebe o resultado da etiquetação feita pelo RuDriCo2, no módulo anterior, e seleciona a melhor etiqueta de POS para cada segmento. Para tanto, o MARv4 emprega o modelo estatístico conhecido como *Hidden Markov Model* (*HMM*) e o algoritmo Viterbi. Uma cadeia de Markov é um tipo especial de autômato ponderado em que a sequência de entrada determina quais estados o autômato vai percorrer.

Finalmente, o último módulo a ser aplicado é o XIP (*Xerox Incremental Parser*) (MOKHTAR; CHANOD; ROUX, 2002), que é responsável pela análise sintática da STRING. Esse analisador também permite a introdução de informação lexical, sintática e semântica, acrescentando-a à saída dos módulos anteriores. O XIP é baseado em autômato de estados finitos e executa a análise sintática do texto através da aplicação sequencial de gramáticas locais, de regras de desambiguação morfossintática e do cálculo de *chunks* e de dependências. O XIP em si é composto pelos seguintes módulos:

- Léxicos: os léxicos do XIP permitem adicionar informações aos diferentes *tokens*.
- Gramáticas locais: o XIP permite escrever regras para buscar padrões, considerando contextos tanto à direita quanto à esquerda de um determinado padrão. Essas regras servem, por exemplo, para definir as entidades que são formadas por mais de uma palavra (*token*). Trata-se de expressões relativamente produtivas que

54 As informações constantes neste capítulo sobre a arquitetura e o funcionamento da STRING foram retiradas da página principal do projeto: https://string.l2f.inesc-id.pt/w/index.php/Main_Page. Acesso em: 14 maio 2015.

podem ser assim agrupadas para formar expressões linguísticas mais complexas e tratadas como unidades lexicais.

- Módulo de *chunking*: as regras de *chunking* do XIP executam uma análise superficial, ou seja, uma análise sintática básica do texto. Para cada tipo de sintagma, por exemplo, sintagma nominal (NP), sintagma preposicional (PP), sintagma verbal (VP), *etc.*, uma sequência de categorias gramaticais é agrupada em estruturas sintáticas elementares. Esses segmentos são chamados de *chunks*.
- Módulo de extração de dependências: no XIP, as dependências são relações sintáticas entre os diferentes *chunks*, as cabeças dos *chunks* ou os elementos dentro dos *chunks*.

As dependências permitem representar um conhecimento mais profundo e mais rico sobre a informação e o conteúdo do texto. As principais dependências correspondem às chamadas funções sintáticas profundas, tais como, sujeito (SUBJ), objeto direto (CDIR), modificador (MOD) *etc.* Outras dependências são apenas relações auxiliares, utilizadas para calcular as dependências sintáticas mais profundas. Por exemplo, a dependência CLINK simplesmente liga cada argumento de uma coordenação à sua conjunção coordenativa.

Depois do módulo do *parser*, a STRING ainda possui também módulos para o reconhecimento de entidades nomeadas, resolução de anáforas, normalização de expressões temporais, ordenação temporal de eventos, etiquetagem de papéis semânticos, dentre outros.

No âmbito da arquitetura geral da STRING, o presente trabalho se insere no módulo XIP porque consiste no processamento sintático automático das *CVS*. Nossa proposta consiste num conjunto de regras produzido automaticamente a partir das matrizes do L-G, que permitem ao sistema extrair dependência que liga o nome predicativo ao verbo-suporte, chamada de SUPPORT.

11.3 Estratégia para implementação dos dados na STRING

Conforme apresentado na seção anterior, a STRING processa os textos de forma modular, uma etapa de cada vez. Numa das etapas finais do módulo de *parsing*, o sistema extrai os eventos (ou predicados) e seus participantes (ou argumentos). No momento, a STRING extrai a estrutura de EVENT para todos os verbos plenos e muitos nomes predicativos.

No caso dos verbos, o sistema associa os eventos aos seus argumentos e circunstanciais. Assim, para uma frase como (11.1), o *parser* extrai a estrutura do evento com as seguintes dependências:

(11.1) *Max costuma ler o jornal no café às sextas-feiras.* [Ex.C]

EVENT_AGENT(ler,Max)
EVENT_OBJECT(ler,jornal)
EVENT_LOC-PLACE(ler,café)
EVENT_TIME-FREQUENCY(ler,a_as sextas-feiras)

No caso das *CVS*, é necessário que o sistema identifique o nome predicativo (e não o verbo-suporte) como EVENT, ou seja, como predicador, já que o *Vsup* serve apenas como uma auxiliar do nome. Assim, a dependência SUPPORT tem que ser extraída primeiro, como ocorre no exemplo:

(11.2) *Max deu um beijo na Ana.* [Ex.C]

SUPPORT_STANDARD(beijo,deu)

Para isso, é necessário fornecer ao sistema as informações de que o verbo *dar* é o verbo-suporte *standard* básico do nome predicativo *beijo*. Também é necessário informar ao sistema que, nessa construção, o *Npred* funciona como complemento direto (CDIR) do verbo-suporte e que os complementos dativos podem ser introduzidos pela preposição *em*, em Português Brasileiro.

As regras a seguir ilustram, de uma forma simplificada, o funcionamento do sistema modular baseado em regras:

```
if (CDIR(#1[lemma:dar],#2[lemma:beijo]) & ~SUPPORT(#2,#?))

    SUPPORT[vsup-standard](#1,#2)

if (SUPPORT(#1,?))

    EVENT[OTHER=+](#1).

if (SUPPORT[vsup-standard](#1[lemma:beijo],#2)

    & EVENT[other](#1) & SUBJ(#2,#3)) EVENT[agent-generic=+](#1,#3).
```

```
if (SUPPORT[vsup-standard](#1[lemma:beijo],#2) & EVENT[other](#1) &
    ^MOD(#2,#3) & PREPD(#3,?[lemma:em]) )
    COMPL(#1,#3),
    EVENT[patient=+](#1,#3).

if (SUPPORT[vsup-standard](#1[lemma:beijo],#2) & EVENT[other](#1) &
    (^MOD[dat](#2,#3) || ^CLITIC(#2,#3[dat]) ) )
    CINDIR[dat=~](#1,#3), EVENT[patient=+](#1,#3).
```

Essas regras são lidas da seguinte forma: primeiramente, a dependência SUPPORT é extraída com a *feature* _vsup-*standard* quando o nome *beijo* é reconhecido como complemento direto do verbo *dar*, desde que nenhum outro *Vsup* tenha sido extraído anteriormente; com base nessa dependência, uma outra dependência de EVENT é extraída para o nome predicativo; em seguida, o sujeito do *Vsup standard* é anotado com o papel semântico de agent (ou agent-generic na terminologia da STRING); depois disso, o sintagma preposicional (MOD) que modifica o *Vsup* e que é introduzido pela preposição *em* é convertido em um complemento (COMPL) do nome predicativo e é anotado com o papel semântico de *patient*; um procedimento semelhante é usado para o complemento dativo, quando reduzido a uma forma pronominal dativa (*lhe*), mas, nesse caso, em vez de COMPL, é usada a dependência CINDIR, por ser um complemento indireto. Todas essas regras são automaticamente produzidas para cada nome predicativo, a partir das informações constantes no Léxico-Gramática.

Para o L-G das construções com *dar*, com 1.489 nomes predicativos e 18 variantes verbais, considerando-se as variantes de *Vsup standard* e de *Vsup* converso, a STRING produziu automaticamente 5.100 regras.

A dependência SUPPORT e os eventos relacionados aos argumentos do *Npred* são representados a seguir:

```
SUPPORT_VSUP-STANDARD(beijo,deu)
EVENT_AGENT(beijo,Max)
EVENT_PATIENT(beijo,Ana)
```

Note-se que, na construção conversa, a estrutura do evento permanece a mesma, mas a dependência SUPPORT é diferente:

(11.3) *Ana ganhou um beijo do Max.* [Ex.C]

SUPPORT_VSUP-CONVERSE(beijo,ganhou)
EVENT_AGENT(beijo,Max)
EVENT_PATIENT(beijo,Ana)

A construção conversa executa a inversão das funções sintáticas dos argumentos, mantendo seus papéis semânticos. A detecção da construção conversa faz gerar um conjunto de regras que também inverte os papéis semânticos associados às posições sintáticas dos nomes predicativos.

Este capítulo, que introduz a terceira parte do livro, retomou os aspectos fundamentais do estado da arte sobre o processamento automático de construções com verbo-suporte. Grande parte dos trabalhos disponíveis na literatura tratam as *CVS* como quaisquer outros tipos de expressões multipalavras (assim como palavras compostas, locuções verbais e expressões fixas). Discutimos a problemática relacionada a essa perspectiva e defendemos a análise das *CVS* como estruturas complexas, formadas por constituintes sintaticamente decomponíveis e analisáveis, e que seguem as mesmas regras gerais de formação de constituintes e sofrem o mesmo tipo de transformações sintáticas aplicáveis a frases com idêntica estrutura interna, tal como sucede nas construções com verbos plenos. Por outras palavras, salvo as propriedades características das *CVS*, estas seguem as mesmas regras da gramática geral da língua.

Ainda neste capítulo, fornecemos noções gerais sobre o *parser* utilizado neste trabalho e apresentamos, em linhas gerais, a proposta para a extração de dependências na STRING, que é a cadeia de processamento do Português, onde as *CVS* foram integradas, onde as regras de extração de dependências foram geradas e onde o *corpus* de referência foi processado.

Em suma, este capítulo tratou da arquitetura da STRING, do funcionamento de seus módulos para o processamento do Português, da estrutura do *parser* XIP e da estratégia para o processamento automático das *CVS*. Depois dessa breve explanação sobre o sistema, apresentaremos, no próximo capítulo, os procedimentos para a constituição e anotação do *corpus* de referência que será processado na STRING.

12 A produção do *corpus* de referência

O *corpus* de referência[55] utilizado é uma amostra do *corpus* PLN.Br Full com frases extraídas automaticamente e anotadas manualmente. Essa amostra abrange candidatos a *CVS* formados por pares dos verbos-suporte *dar*, *fazer* e *ter* e um *Npred*. As matrizes do Léxico-Gramática dos predicados nominais com esses três *Vsup* foram intersectadas com grafos de referência do Unitex, a fim de procurar sistematicamente no *corpus* todas as combinações de um dos verbos-suporte com um dos nomes predicativos, considerando-se apenas as combinações previstas na matriz, a fim de limitar a "explosão combinatória" que usar todos os *Vsup* e *Npred* produziria.

Por meio dessa metodologia, foram identificadas no *corpus* 121.198 frases com pares (*Vsup, Npred*) candidatos a *CVS*, isto é, frases em que ocorre simultaneamente um potencial *Vsup* e um potencial *Npred*, dadas as combinações possíveis indicadas nas matrizes do L-G.

12.1 Extração automática das frases

Das 3 matrizes do L-G (uma para cada verbo-suporte elementar: *dar*, *fazer* e *ter*), foram utilizadas as informações de verbo-suporte elementar, as variantes de verbo-suporte e de verbo-suporte converso, além da coluna dos nomes predicativos. Essas informações sobre a coocorrência dos verbos com os nomes de cada construção foram intersectadas com um grafo de referência do Unitex (ver subseção 2.2.1). Para tanto, criamos o grafo ilustrado na figura 7

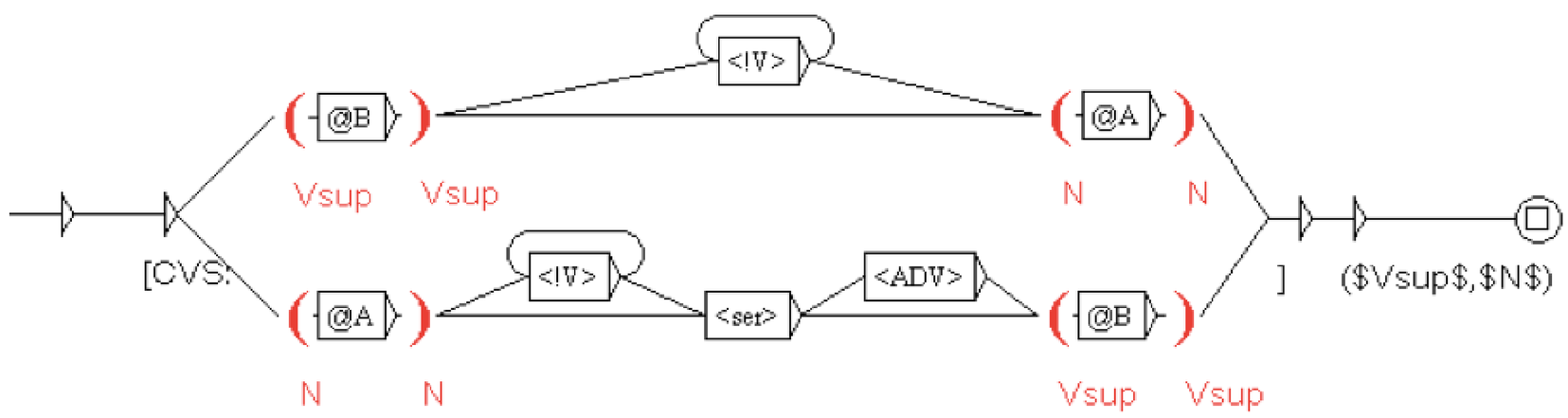

Figura 7 – Grafo de referência para identificação das *CVS*
Fonte: Elaborado pela autora e já publicado em Rassi, Baptista e Vale (2015, p.212)

Nesse grafo, atribuímos uma etiqueta "[CVS: ...]" a ser inserida em todas as frases do *corpus* em que existir um candidato a *CVS*. No penúltimo estado, inserimos a etiqueta "($VSup$,N)", a ser apresentada na saída (*output*), instanciando o verbo-suporte e o nome predicativo de tal construção, separados por vírgula e entre parênteses. As

55 Ressalte-se que partes do conteúdo deste capítulo já foram publicadas em Rassi, Baptista e Vale (2015).

variáveis “@A” e “@B” são preenchidas por *Npred* e *Vsup*, respectivamente, a partir da informação constante na matriz do L-G. A caixa que contém “<!V>” indica a negação da categoria verbo, ou seja, naquela posição pode haver qualquer palavra que não seja um verbo. Nessa caixa, uma das setas que saem indica um *loop*, ou seja, a possibilidade de ocorrer uma sequência indefinidamente longa de elementos que não sejam verbos. A caixa “<ser>” corresponde a qualquer forma conjugada do verbo *ser*. A caixa “<ADV>” corresponde a qualquer advérbio. Os símbolos “N” (nome) e “Vsup” (verbo-suporte) em vermelho identificam as variáveis usadas no *output*.

O grafo possui dois caminhos. A partir da etiqueta “[CVS:”, que delimita a *CVS*, o primeiro caminho se lê: instancia-se o verbo-suporte, seguido imediatamente de um nome predicativo ou seguido de qualquer sequência de palavras, desde que não sejam verbos; em seguida instancia-se o nome predicativo e fecha-se a etiqueta *CVS* com “]”.

No segundo caminho, que prevê a inversão da ordem de verbo e nome, nos casos de construções passivas, lê-se: instancia-se o nome predicativo, que pode ser seguido imediatamente de qualquer forma do verbo *ser*[56] ou pode haver entre eles uma sequência indefinidamente longa de palavras que não sejam verbos; o verbo *ser* pode ainda ser seguido imediatamente do nome predicativo ou haver, entre eles, algum advérbio; no fim, fecha-se a etiqueta “CVS” com “]”.

Ao intersectar o grafo de referência com as matrizes do L-G, o Unitex produz um subgrafo de resultado para cada linha da matriz que foi instanciada por uma combinação de *Vsup* e *Npred*. Assim, o número de subgrafos corresponde ao número de linhas das matrizes. Nesses grafos, definem-se os padrões combinatórios a serem procurados e, posteriormente, o sistema gera, para cada linha, o correspondente transdutor que permite procurar esses padrões nos textos.

A Figura 8 mostra um exemplo de um dos subgrafos de resultados gerado automaticamente pelo Unitex.

Neste subgrafo, representam-se as combinações determinadas pelo nome predicativo *esclarecimento*, que pode selecionar o verbo-suporte elementar *dar* ou suas variantes estilísticas e/ou aspectuais *conceder*, *fazer* e *prestar*, na construção *standard*, e selecionar os verbos-suporte conversos *receber* e *ter*, na construção conversa. No caminho inferior do grafo, apresentam-se as construções passivas correspondentes[57].

56 Os verbos *estar*, *ficar* e outros auxiliares das passivas de estado não foram considerados neste momento.

57 A possibilidade de o verbo *ter* integrar uma construção passiva (no segundo caminho do subgrafo) deve-se ao caráter automático do processo de geração do grafo. Uma vez que este é usado para reconhecimento (e não para geração), isso não constitui um problema para os objetivos deste trabalho.

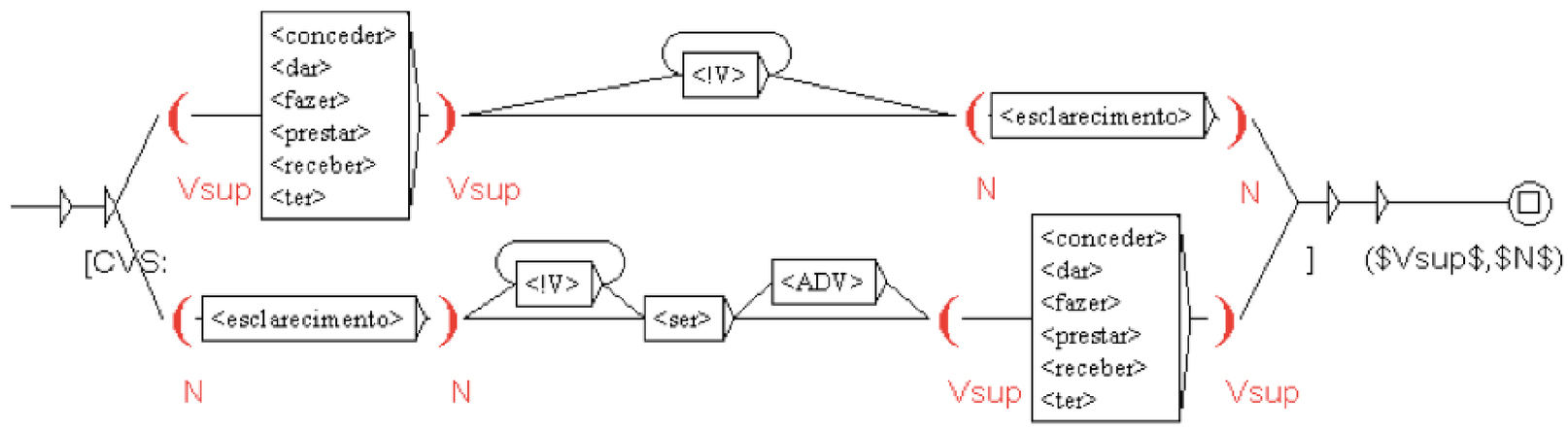

Figura 8 – Subgrafo 0398 da matriz dos *Npred* com *dar*
Fonte: Elaborado pela autora e já publicado em Rassi, Baptista e Vale (2015, p.213)

Além de todos os subgrafos de resultados, o Unitex também gera automaticamente um grafo único de resultados, que integra todos os subgrafos. Esse grafo de resultados foi aplicado ao *corpus* PLN.Br Full a fim de buscar as combinações de *Vsup* e *Npred* nos textos reais. A Fig. 9 mostra um trecho da concordância apresentada pelo Unitex, quando da procura pelo grafo de resultados das construções com o verbo *dar*.

A soma dos esforços de cada um, atletas e comissão técnica, será fundamental para atingirmos o tão almejado pódio, para [CVS: darmos o salto] (darmos,salto) a que nos propusemos.
Três cabeças-de-chave [CVS: deram adeus] (deram,adeus) ao torneio.
Karl Malone [CVS: fez oito pontos] (fez,pontos) no período seguinte, quando o Jazz ampliou a sua vantagem para 52 a 41.
Ao eleger Luiza Erundina em 1988, o "bom povo" [CVS: deu prova] (deu,prova) de "alta politização e alto nível".
O primeiro esboço que Duchamp fez da obra [CVS: tinha um título] (tinha,título) mais longo ainda: entre "Étant Donnés" e "la chute" aparecia o nome de Maria.
O catálogo da exposição de Duchamp no Palazzo Grassi de Veneza, encerrada em julho último, [CVS: dá as primeiras pistas] (dá,pistas).

Figura 9 – Extrato da concordância exibida pelo Unitex
Fonte: Elaborado pela autora

Somando-se os resultados da aplicação das três matrizes, o Unitex retornou 177.287 frases. Algumas dessas frases eram duplicadas porque, muitas vezes, o mesmo nome predicativo constava em duas ou três matrizes com as mesmas variantes de *Vsup*; portanto, excluindo-se as duplicadas, restaram 131.734 frases.

Algumas delas, que haviam sido mal segmentadas, foram excluídas da amostra. Também foram excluídos muitos pares com as formas *desse(s)* e *deste(s)*, por se tratarem, na sua grande maioria, da combinação da preposição *de* com o pronome demonstrativo *esse(s)/este(s)*, ambíguas com as respectivas flexões do verbo *dar*. Por fim, foram excluídas também todas as formas conjugadas de *poder*, *dever*, *olhar*, *saber* e *parecer*, deixando apenas as formas de substantivo singular e plural dessas palavras: *poder(es)*, *dever(es)*, *olhar(es)*, *saber(es)* e *parecer(es)*.

Portanto, o universo com o qual estamos trabalhando é formado por 92% dos dados iniciais (131.734 frases extraídas pelo Unitex), pois contém as 121.198 frases restantes após essas operações de filtragem.

Em seguida, todas as formas dos verbos e dos nomes foram lematizadas e foram calculadas as ocorrências de cada par de *Vsup* e *Npred*. As 121.198 frases são representadas por 4.668 *types* de pares (*Vsup,Npred*), variando de uma ocorrência até 3.519 ocorrências do mesmo par (*e.g. fazer parte*).

12.2 A seleção da amostra

O universo total do nosso objeto de estudo consiste em 121.198 frases, que é exaustivamente grande para uma anotação manual. Desse universo, foi selecionada uma amostra aleatória estratificada, a ser anotada, e que é proporcional ao número de ocorrências de cada par (*Vsup,Npred*) e, ao mesmo tempo, que leva em consideração a diversidade das combinações. A amostra foi estratificada em três grandes blocos:

- A amostra global recuperou pelo menos um caso de todos os pares que tenham, pelo menos, 21 ocorrências, somando-se um total de 2.646 frases.
- A amostra intermédia corresponde aos pares (*Vsup,Npred*), que tenham, no mínimo duas ocorrências, e, no máximo, 20 ocorrências.
- A última porção da amostragem é constituída pelos *hapax legomena*, ou seja, os pares que possuem uma única ocorrência no *corpus*.

A Tabela 6 apresenta a distribuição dos pares candidatos a *CVS* em cada bloco da amostra.

Amostra	**nº frases**	**% frases**	**nº *types***	**% *types***	**Cobertura do *corpus***
Amostra global (**freq.** > **21**)	2.646	2,18%	1.130	24,2%	84,85%
Amostra intermédia (**2** 6 **freq.** 6 **20**)	–	–	2.537	54,3%	14,30%
hapax legomena (**freq. = 1**)	–	–	1.001	21,5%	0,85%
TOTAL	**121.198**	**100%**	**4.668**	**100%**	**100%**

Tabela 6 – Distribuição da amostra
Fonte: Elaborado pela autora

Ressalte-se que o número (e porcentagem) de frases aparece apenas para a amostra global porque corresponde às frases que foram, de fato, anotadas. A amostra intermédia e os *hapax legomena* não foram considerados nesta etapa. A Tabela 6 indica também o número de *types* em cada bloco e sua porcentagem em relação ao total, além da porcentagem do *corpus* que é abrangido em cada bloco.

Como se pode perceber, a amostra global abrange 2.646 frases, ou seja, 2,18% do conjunto total (121.198 frases). Dos 4.668 *types* de pares, a amostra global conta com 1.130 *types*, o que corresponde a 24,2% dos pares diferentes e cobre 84,85% das sentenças do *corpus*. A amostra intermédia considera outros 2.537 *types*, que equivalem a 54,3% dos *types* do universo, mas cobrem apenas 14,30% das sentenças do *corpus*. O último bloco da amostra compreende os *hapax legomena*, contendo 1.001 *hapax*, que correspondem a 21,5% do total de *types* e cobrem 0,85% do *corpus*.

Conforme apontado anteriormente, apenas o primeiro subconjunto de dados, que corresponde à amostra global, foi anotado. Nesse sentido, o conjunto de frases que constituem o *corpus* de referência das *CVS* em Português, ou seja, o conjunto anotado que pode ser utilizado para avaliação de diferentes métodos de identificação, classificação ou processamento de *CVS*, foi estabelecido com base somente na amostra de quase 25% do total de *types* (*Vsup*, *Npred*), o que cobre 84,85% do *corpus* total.

A próxima seção descreve a ferramenta utilizada para a anotação, bem como o processo de anotação em si.

12.3 O processo de anotação

Esta seção apresenta o material, os métodos e os procedimentos utilizados no processo de anotação das 2.646 frases com candidatos a *CVS*: (i) a subseção 12.3.1 apresenta mais detalhadamente a ferramenta de anotação chamada CorpusAnnotator (SUISSAS, 2014), já indicada anteriormente na subseção 2.2.3; (ii) a subseção 12.3.2 descreve o grupo de anotadores e o manual de anotação utilizado; e (iii) a subseção 12.3.3 trata dos procedimentos para a normalização (regularização das etiquetas) das anotações.

12.3.1 A ferramenta de anotação

Para facilitar a tarefa dos anotadores, Suíssas (2014) criou o programa CorpusAnnotator. Class, desenvolvida em Java (JRE). Para o programa funcionar, são necessários dois arquivos com extensão .txt: (i) um arquivo com todas as frases a serem anotadas; e (ii) um arquivo de parametrização com todas as formas de singular e plural dos nomes predicativos, já que o *Npred* é o alvo a ser anotado.

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.*
- **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 (com verbo pleno, expressão fixa, ou outros)
  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]

Essa última etiqueta permite distinguir as *CVS* e também as construções com *VopC* de outras expressões formadas pelo mesmo par de elementos lexicais, mas com propriedades sintáticas e semânticas distintas.

Na amostra, há apenas um par candidato a *CVS*, o qual está indicado entre parênteses no início da frase a ser anotada. Além dessas etiquetas (SVC-STANDARD, SVCCONVERSE, VOPC ou OTHER), o programa também apresenta por *default* as opções:

- **Nenhuma das anteriores/Não sei**: que deve ser assinalado quando o anotador não souber qual etiqueta atribuir;
- **Tenho dúvidas**: que deve ser marcada, mesmo que tenha selecionado alguma das etiquetas anteriores (SVC-STANDARD, SVC-CONVERSE, VOPC ou OTHER), para que o anotador possa retornar, posteriormente, somente às frases em que ficou com dúvida, marcando a opção "Mostrar apenas as frases com dúvidas". Assim, permite que outra pessoa possa vir a decidir, mais tarde, qual é a melhor solução nesse caso.
- **Erro na frase**: que deve ser selecionada caso a frase esteja truncada, incompleta ou não apresente contexto suficiente para determinar qualquer etiqueta.

12.3.2 Procedimentos de anotação

O conjunto total de dados (2.646 frases) foi anotado por cinco anotadores linguistas especialistas em *CVS*, todos falantes nativos do Português, sendo quatro brasileiros e

um português. Todos os anotadores receberam um conjunto de orientações (*guidelines*), com explicações sobre a tarefa e o funcionamento do programa.

Nas *guidelines* constam apenas informações sobre como baixar o programa, como utilizar a ferramenta e como salvar o arquivo de saída dos dados produzidos. Não foram dadas aos anotadores informações adicionais sobre o conteúdo/assunto dos dados a serem anotados. Essas orientações encontram-se no Apêndice D.

12.3.3 Normalização da anotação

Os arquivos resultantes da anotação de cada anotador foram intersectados e cada anotação consta em uma coluna de uma tabela única. As quatro categorias definidas previamente (SVC-STANDARD, SVC-CONVERSE, VOPC e OTHER) poderiam ainda ser acrescidas de outras etiquetas: "?" para indicar "dúvida" e "#" para indicar "erro na frase" (ou apenas uma ou ambas)[58].

Todos os casos de dúvida, mesmo quando outra etiqueta fora também atribuída, foram analisados individualmente e serão discutidos adiante. Porém, para calcular a concordância entre anotadores, desconsideramos as etiquetas "?", "#" e "?#" nos casos em que outra categoria tenha sido também atribuída. Dessa forma, restaram apenas cinco etiquetas: as quatro referentes às quatro categorias e uma quinta etiqueta para as frases em que o anotador não atribuiu nenhuma das quatro etiquetas anteriores, marcando apenas "?". Essas etiquetas foram então convertidas em dígitos:

- SVC-STANDARD, representada por 1;
- SVC-CONVERSE, representada por 2;
- VOPC, representada por 3;
- OTHER, representada por 4;
- ? , representada por 5;

As frases que apresentavam erro ("#") foram descartadas da análise. Consideram-se erros na frase os casos em que a frase não foi segmentada ou identificada corretamente, como ocorre em (12.1), ou não apresentava contexto suficiente para atribuirmos uma etiqueta (12.2), ou ainda os casos em que os candidatos a *Vsup* e a *Npred* não são formas de verbo ou formas de nome, respectivamente, tais como (12.3), (12.4), (12.5) e (12.6).

58 O arquivo poderia conter até 19 etiquetas diferentes. São elas: SVC-STANDARD, SVCSTANDARD?, SVC-STANDARD#, SVC-STANDARD?#, SVC-CONVERSE, SVC-CONVERSE?, SVC-CONVERSE#, SVC-CONVERSE?#, VOPC, VOPC?, VOPC# , VOPC?#, OTHER, OTHER?, OTHER#, OTHER?#, #, ?#, ou "?".

(12.1) *(**ter**,**usufruto**) Waldomiro Diniz **tinha** uma utilidade.* [Ex.R]

(12.2) *(**ter**,**ponto**) Bicho **tem ponto** nos fundos do Planalto.* [Ex.R]

(12.3) *(**fazer**,**filho**) Carlos Alberto H. **Farias** Magioli **Filho**.* [Ex.R]

(12.4) *(**ter**,**passado**) Dou graças a Deus **ter passado** 28 anos no Exército seguindo a Convenção de Genebra (que veta tortura a prisioneiros).* [Ex.R]

(12.5) *(**ter**,**acordo**) Os dois deverão explicar aos senadores os novos **termos** do **acordo** fechado com o FMI.* [Ex.R]

(12.6) *(**dar**,**mostra**) A análise dos **dados** históricos **mostra** que existem maiores oscilações dos preços das ações do que dos dados fundamentais da economia.* [Ex.R]

Ao todo, 75 linhas foram excluídas por conterem erros na frase.

12.4 Cálculo da concordância entre anotadores

As anotações foram tabuladas em cinco colunas e, em seguida, comparadas com auxílio à ferramenta ReCal 0.1 Alpha for 3+ Coders[59], para calcular a concordância entre três ou mais anotadores. A Figura 10 apresenta a saída da ferramenta.

Average Pairwise Percent Agreement

Average pairwise percent agr.	Pairwise pct. agr. cols 1 & 5	Pairwise pct. agr. cols 1 & 4	Pairwise pct. agr. cols 1 & 3	Pairwise pct. agr. cols 1 & 2	Pairwise pct. agr. cols 2 & 5	Pairwise pct. agr. cols 2 & 4	Pairwise pct. agr. cols 2 & 3	Pairwise pct. agr. cols 3 & 5	Pairwise pct. agr. cols 3 & 4	Pairwise pct. agr. cols 4 & 5
80.805%	80.397%	84.286%	84.714%	83.353%	73.318%	80.552%	81.408%	78.413%	84.053%	77.557%

Fleiss' Kappa

Fleiss' Kappa	Observed Agreement	Expected Agreement
0.602	0.808	0.518

Average Pairwise Cohen's Kappa

Average pairwise CK	Pairwise CK cols 1 & 5	Pairwise CK cols 1 & 4	Pairwise CK cols 1 & 3	Pairwise CK cols 1 & 2	Pairwise CK cols 2 & 5	Pairwise CK cols 2 & 4	Pairwise CK cols 2 & 3	Pairwise CK cols 3 & 5	Pairwise CK cols 3 & 4	Pairwise CK cols 4 & 5
0.604	0.585	0.655	0.67	0.664	0.479	0.609	0.63	0.554	0.657	0.53

Figura 10 – Concordância entre anotadores usando o Recall
Fonte: *Output* da ferramenta ReCal 0.1 Alpha for3+ Coders

Os três quadros apresentados na Figura 10 correspondem respectivamente a: (i) a concordância entre cada par de anotadores; (ii) o coeficiente kappa de Fleiss (FLEISS, 1971, p.378-382), que é uma medida estatística para avaliar a confiabilidade da concordância entre um número fixo de avaliadores, considerando o grau de concordância efetivamente observado e comparando-o com o grau de concordância esperado; e (iii) o

59 Disponível em: http://dfreelon.org/recal/recal3.php#result1.

coeficiente kappa de Cohen (COHEN, 1960), que também é uma medida estatística que avalia a concordância entre pares de avaliadores, também chamada de *inter-annotator agreement* ou *inter-rater agreement*.

A média geral das concordâncias por par de anotadores é de 80,8%. As concordâncias individuais par a par serão analisadas mais adiante. Verifica-se que o coeficiente Kappa de Fleiss é de 0,602, tendo-se observado uma concordância de 0,808 contra uma concordância esperada de 0,518. O coeficiente Kappa de Cohen avalia a concordância entre pares de anotadores, comparando com a concordância esperada caso a anotação tivesse sido aleatória, cuja média neste caso foi de 0,604.

Como se pode perceber, os valores do coeficiente Kappa de Fleiss (0,602) e de Cohen (0,604) são bastante semelhantes. De acordo com a tabela de interpretação[60] do coeficiente Kappa de Fleiss (LANDIS; KOCH, 1977), a concordância entre os anotadores (0,602) é considerada no intervalo entre "moderada" e "substancial". Esse valor pode ser interpretado como um indicativo do grau de dificuldade da tarefa para humanos.

Levando-se em consideração apenas a concordância entre os pares de anotadores, a média geral de concordância entre todos os anotadores é de 80,8%. As concordâncias individuais, calculadas pela ferramenta podem ser comparadas por meio da matriz de confusão (Tabela 7):

Annotator	1	2	3	4	5
1	–	83.35%	**84.71%**	84.28%	80.39%
2	–	–	81.40%	80.55%	**73.31%**
3	–	–	–	84.05%	78.41%
4	–	–	–	–	77.55%
5	–	–	–	–	–

Tabela 7 – Concordância entre pares de anotadores
Fonte: Elaboração própria

Como se pode perceber pela Tabela 7, a concordância par a par de anotadores varia entre 73.31% e 84.71%, sendo que o anotador 1 atingiu os melhores resultados e o anotador 5 apresenta resultados mais baixos.

60 A tabela de interpretação do coeficiente Kappa de Fleiss pode ser consultada em: http://en.wikipedia.org/ wiki/Fleiss'_kappa.

12.5 Seleção da subamostra

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 três ou mais anotadores atribuíram a mesma etiqueta. A Tabela 8 dá conta do número de frases por grau de concordância entre anotadores.

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%**

Tabela 8 – Número de frases por grau de concordância entre anotadores na amostra global
Fonte: Elaboração própria

Selecionamos, do conjunto total da amostra, todas as frases que tinham o verbo-suporte elementar *dar* ou alguma variante do *Vsup standard* (*aplicar*, *atribuir*, *bater*, *conceder*, *estar com*, *fazer*, *impor*, *ministrar*, *pedir*, *prestar*, *ser de*, *ter* e *tirar*) ou do *Vsup* converso (*levar*, *pegar*, *receber*, *ter* e *tomar*) e um nome predicativo, e que tinham sido anotadas com a mesma etiqueta por, pelo menos, três anotadores. Nesse sentido, foram selecionadas 580 frases para compor uma subamostra e serem, posteriormente, analisadas na STRING. A Tabela 9 apresenta o número de frases dessa subamostra, distribuídas por grau de concordância entre os anotadores.

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%**

Tabela 9 – Número de frases por grau de concordância entre anotadores na subamostra
Fonte: Elaboração própria

Na Tabela 9, não indicamos as frases com concordância entre dois anotadores porque essas sentenças não foram selecionadas para a subamostra. Como se pode perceber, o número de frases por grau de concordância entre anotadores da subamostra é proporcional ao número de frases por grau de concordância na amostra global (ver Tabela 8).

Para calcular a correlação entre o número de frases por grau de concordância entre anotadores na amostra global e na subamostra, calculamos o coeficiente de Pearson, que é uma medida do grau de relação linear entre duas variáveis quantitativas. Esse coeficiente pode variar entre os valores -1 e 1, a partir da seguinte fórmula:

$$\rho(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$$

Considerando-se que X corresponde à concordância de anotações de 3, 4 e 5 anotadores, na amostra global (326, 627 e 1574), e que Y corresponde à concordância de anotação dos 3, 4 e 5 anotadores, na subamostra (95, 137 e 337), o coeficiente de Pearson foi calculado em 0,99754588, o que indica um alto grau de correlação entre a amostra global e a subamostra.

Pode-se analisar também a distribuição das frases por categoria (ou etiqueta) atribuída pela maioria dos anotadores. A Tabela 10 mostra a distribuição de frases da subamostra, discriminando as porcentagens por categoria.

	3 anot.	**4 anot.**	**5 anot.**	**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 erros	–	–	–	11	1,89%
TOTAL	**95**	**137**	**337**	**580**	**100%**

Tabela 10 – Distribuição da subamostra por categoria
Fonte: Elaboração própria

Como se nota, a categoria de verbo-suporte *standard* é a mais consensual entre os anotadores, correspondendo a quase 60%, seguida da categoria de verbo-suporte converso, com cerca de 25%. Os demais casos somam 14,29%.

Este capítulo abordou os procedimentos experimentais para a constituição e a anotação do *corpus* de referência. Foram descritos os seguintes procedimentos: (i) a extração automática de cerca de 121 mil frases que continham pares de *Vsup* e *Npred* candidatos a *CVS*; (ii) a seleção de uma amostra aleatória estratificada, contendo 2.646 frases para serem manualmente anotadas; (iii) o processo de anotação, a ferramenta computacional utilizada, o grupo de anotadores, os procedimentos adotados e o processo de normalização das anotações; (iv) o cálculo do grau de concordância entre anotadores, cuja média foi de 80,8%; e (v) a seleção de uma subamostra contendo apenas os candidatos a *CVS* construídas com o verbo *dar* e suas variantes e excluindo-se as *CVS* com *fazer* e *ter*, presentes na amostra inicial.

A amostra inicial contava com 2.646 frases, incluindo os *Vsup dar*, *fazer* e *ter* e suas variantes estilísticas e/ou aspectuais. Já a subamostra conta com 580 frases anotadas manual e independentemente e normalizadas. Essa subamostra constitui um *corpus* de referência das *CVS* com *dar* em Português Brasileiro. Esse *corpus* servirá como *golden standard* para a avaliação do processamento automático das *CVS* pela STRING[61].

No próximo capítulo, explicaremos como foi feita a validação formal e linguística da matriz do L-G e a conversão automática desses dados em regras que serão, posteriormente, utilizadas pela STRING.

61 O *corpus* está disponível para a comunidade científica através do *link* https://sites.google.com/site/amandaprassi/recursos e pode servir de base à avaliação de outros sistemas.

13 Integração dos dados na STRING

Seguindo a estratégia delineada em trabalho anterior (RASSI *et al.*, 2014), a integração das *CVS* no *parser* da STRING se deu em quatro etapas sucessivas:

(i) As 1.489 *CVS* foram descritas e formalizadas em matriz do L-G. Esse processo foi detalhadamente descrito na Parte II.

(ii) O léxico usado pelo *parser* foi enriquecido com os nomes predicativos que são específicos do PB. Como a STRING foi projetada para processar a variante europeia do Português, havia 40 *Npred* com sufixo *-ada* que não constavam no léxico utilizado pela STRING, tais como, *dar uma cadeirada*, *dar uma mãozada*, *dar uma esquentada etc.*, que é um recurso derivacional muito produtivo na variante brasileira do Português. Assim, foi preciso incluí-los no léxico da STRING.

(iii) Um programa de validação foi criado e aplicado à matriz de dados, a fim de verificar a consistência formal e linguística desses dados. Esse processo foi explicado no capítulo 14.

(iv) Depois de validados, os dados do L-G foram convertidos em regras de dependência que são usadas pelo *parser* XIP, usando um conversor automático; essas regras também foram apresentadas no capítulo 14.

Este capítulo apresenta, agora, na seção 13.1, os resultados da primeira avaliação da *performance* do sistema; na seção 13.2, faz-se uma análise de erros observados com os resultados da primeira avaliação; na seção 13.3, são descritos os casos particulares de verdadeiros-positivos que merecem atenção; e, na seção 13.4, procede-se à segunda avaliação da performance do sistema, após a revisão e correção de alguns erros identificados na etapa de análise dos erros.

13.1 Resultados da primeira avaliação

A primeira avaliação corresponde ao processamento das 580 frases da subamostra, em comparação com a saída do sistema com o *golden standard*, isto é, com a anotação consensual da maioria dos anotadores (ver Capítulo 12). Os valores de verdadeiro-positivo (TP), falso-positivo (FP), falso-negativo (FN) e verdadeiro-negativo (TN) foram definidos com base nos valores da anotação manual da subamostra. A avaliação foi feita considerando-se as métricas de avaliação: Precisão, Cobertura, Acurácia e Medida-F, que serão explicadas adiante. A Tabela 11 apresenta os resultados da primeira avaliação da performance da STRING.

TP	FP	FN	TN	Precision	Recal	Accuracy	F-measure
350	91	114	25	79%	75%	65%	77%

Tabela 11 – Resultado da primeira avaliação da performance da STRING
Fonte: Elaboração própria

Das 580 frases analisadas, a STRING extraiu a dependência SUPPORT de 441 frases, das quais 350 foram corretamente identificadas (TP) e 91 foram incorretamente extraídas (FP). Além disso, não extraiu 25 casos que, de fato, não deveriam ter sido identificados (TN), e deixou de extrair outros 114 casos que deveriam ter sido extraídos (FN). As métricas de avaliação (Precisão, Cobertura, Acurácia e Medida-F) foram calculadas com base nos valores de TP, FP, FN e TN, usando as seguintes fórmulas:

$$(P) Precisão = (TPTP + FP)$$

$$(R) Cobertura = (TPTP + FN)$$

$$(A) Acurácia = \frac{(TP + TN)}{(TP + FP + FN + TN)}$$

$$(F) Medida - F = 2(P * P + *RR)$$

Os resultados da saída do sistema podem ser observados também com relação a cada categoria (STANDARD, CONVERSE, VOPC e OTHER). A Tabela 12 apresenta, por categoria, os valores da anotação manual comparados aos da análise automática.

	Anotação	% anotação	Análise automática	% análise automática
STANDARD	347	59,82%	246	42,41%
CONVERSE	150	25,86%	195	33,62%
OTHER	69	11,89%	139	23,96%
VOPC	3	0,51%	0	0%
TOTAL	569	98,08%	580	100%

Tabela 12 – Comparação, por categoria, entre a anotação e a análise automática
Fonte: Elaboração própria

O total de frases da anotação (569) não corresponde ao total de frases do *corpus* de referência (580) porque, conforme apontado na Tabela 10, havia 11 frases que os anotadores anotaram com "erro na frase" e não atribuíram a elas nenhuma etiqueta.

A STRING extraiu a dependência SUPPORT[vsup-standard] de apenas 246 frases (consideradas *True Positives*) das 347 que os humanos anotaram como STANDARD. Já a dependência SUPPORT[vsup-converse] foi extraída para 195 frases, comparativamente às 150 anotadas pelos humanos. A STRING não identificou os três casos de VOPC anotados pelos humanos porque essa informação não estava codificada na matriz do L-G das *CVS* com *dar*. Outras 139 frases cuja dependência SUPPORT não foi extraída devem-se a dois problemas: algumas delas decorrem de erros nos dados do L-G e outras decorrem de diferentes tipos de ambiguidades não resolvidos ou de falhas em alguma etapa do processamento da cadeia.

A seguir, apresenta-se a análise dos principais erros identificados no processo. Após proceder à análise dos erros, foram adotadas quatro medidas: (i) correção dos erros que eram decorrentes dos dados do Léxico-Gramática; (ii) revisão sistemática da anotação manual, que não estava consistente em alguns pontos; (iii) inserção de algumas formas nos dicionários usados pelo XIP; e (iv) criação de "regras de limpeza" para os casos em que a STRING extraía dependências duplicadas. Após todas essas medidas, o *corpus* foi processado novamente. Os resultados da segunda avaliação serão apresentados na seção 13.4.

13.2 Análise dos erros

13.2.1 Análise dos falsos-positivos: erros de classificação do tipo de dependência

Foram observados 91 casos de falsos-positivos, ou seja, 91 dependências que a STRING extraiu indevidamente, seja porque (i) extraiu alguma dependência e não deveria ter extraído dependência alguma; ou (ii) extraiu um tipo de dependência diferente da que deveria ter sido extraída.

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 constituir construções *standard* quanto construções conversas[62].

62 Apesar de o objeto de investigação desta pesquisa não ser as *CVS* com o verbo *ter*, essas construções colocam um problema à descrição e processamento das *CVS* com o verbo *dar*, uma vez que o *Vsup ter* pode também funcionar como variante estilística de *dar* (em construções *standard*) ou de *receber* (em construções conversas).

(13.1) *A Ana (**deu** + **teve**) sua **participação** no jornal nacional.* [Ex.C]

(13.2) *O jornal nacional (**recebeu** + **teve**) a **participação** da Ana.* [Ex.C]

A primeira frase é tipicamente uma construção *standard*, ao passo que a segunda frase é tipicamente uma construção conversa. Ambas podem ser formadas pelo verbo *ter* e o mesmo nome predicativo *participação*. Nesses casos, decidiu-se adotar sistematicamente a dependência SUPPORT[vsup-standard], em detrimento da SUPPORT[vsup-converse], nos casos de ambiguidade de classificação, em que o verbo pode funcionar tanto como suporte *standard* quanto 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. Os exemplos apresentados a seguir foram etiquetados pela STRING como SUPPORT[vsup-standard], e foram anotados pela maioria ou unanimidade dos anotadores como converso.

(13.3) *(**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.* [Ex.R]

SUPPORT[vsup-standard](participação,terá)[63]

(13.4) *(**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.* [Ex.R]

SUPPORT[vsup-standard](prazo,tem)

(13.5) *(**ter**, **voto**) O PMDB conta com 5 integrantes, mas **terá** um **voto** a menos se Juvêncio estiver na presidência.* [Ex.R]

SUPPORT[vsup-standard](voto,terá)

(13.6) *(**ter**, **prejuízo**) Pará deve **ter prejuízo** com jogo do Brasil.* [Ex.R]

SUPPORT[vsup-standard](prejuízo,ter)

63 As dependências que foram extraídas incorretamente serão indicadas em vermelho e as que foram extraídas corretamente serão indicadas em verde, para facilitar a visualização dos TPs, FPs, TNs e FNs.

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 *standard*. Esses nomes predicativos, associados ao *Vsup ter*, ao mesmo tempo em que podem formar construções *standard* (13.7), também podem ser o resultado de uma conversão da construção com o verbo-suporte *standard dar* (13.8):

(13.7) *A Ana **tem** uma **orientação** católica.* [Ex.C]

(13.8) *O professor **deu** uma **orientação** ao aluno.* [Ex.C]

[*Conversão*] ≡ *O aluno **teve** uma **orientação** por parte do professor.* [Ex.C]

Como se pode perceber, (13.7) e (13.8) são construções *standard* diferentes; trata-se de dois nomes predicativos diferentes: um se refere a "uma formação religiosa e/ou ideológica", e o outro *Npred* se refere ao "ato de orientar, instruir, ensinar". Na matriz do L-G das construções com *dar*, está alistada apenas a construção *dar orientação*, cuja conversão se faz com *ter orientação*. O exemplo (13.7) 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*.

Foram identificados outros erros ou falhas nos dados do L-G, por exemplo:

(13.9) *(**ter**, **informação**) A página **tem informações** sobre o clube, fotos e os nomes dos membros.* [Ex.R]

SUPPORT[vsup-converse](informações,tem)

A STRING identificou a frase (13.9) como SUPPORT[vsup-converse], ao passo que os humanos a etiquetaram como *standard*. O problema se deve ao fato de que, na matriz do L-G, o verbo *ter* somente estava codificado como uma variante do verbo-suporte converso *receber*, mas não como uma variante do verbo-suporte *standard dar*. Esses casos foram corrigidos, de forma que o verbo *ter* também possa ser analisado como uma variante do *Vsup standard dar*.

Há ainda outras frases que também foram etiquetadas pela STRING como SUPPORT [vsup-standard], e que foram anotadas pela maioria ou unanimidade dos anotadores como construções com verbo-operador causativo (VopC):

(13.10) *(**dar**, **sorte**) "Colocar roupa branca e pular sete ondas **dão sorte** porque são rituais para atrair coisas boas e, se você acredita, funcionam.".* [Ex.R]

SUPPORT[vsup-standard](sorte,dão)

(13.11) (***dar***, ***prejuízo***) *Fraude on-line* ***dá prejuízo*** *de R$ 100 mi.* [Ex.R]

SUPPORT[vsup-standard](prejuízo,dá)

(13.12) (***dar***, ***voto***) *Motivo: administra o orçamento para construção de casas populares, que é polpudo e* ***dá votos****.* [Ex.R]

SUPPORT[vsup-standard](votos,dá)

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:

(13.13) *A Ana* ***deu sorte*** *na loteria.* [Ex.C]

(13.14) *A empresa da Ana* ***deu prejuízo*** *durante todo o ano.* [Ex.C]

(13.15) *A Ana* ***deu seu voto*** *para o candidato da oposição.* [Ex.C]

Nesses casos, a identificação dos papéis semânticos dos argumentos poderia auxiliar na desambiguação das etiquetas a serem atribuídas às *CVS*. Em (13.10), se o sujeito (*colocar roupa branca e pular sete ondas*) fosse corretamente etiquetado com o papel semântico de CAUSA, isso seria um indicativo de que a construção é causativa, e não *CVS*. Já o exemplo (13.11) 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 (13.12), o sujeito de *dá votos* é a palavra *orçamento*, o que significa que o orçamento é a causa que faz com que alguém obtenha votos, portanto a construção também é causativa.

Apesar de as restrições distribucionais ao preenchimento das posições argumentais e dos respectivos papéis semânticos terem sido codificadas na matriz do L-G, essas informações não foram usadas no processo de extração da dependência SUPPORT, já que é necessário primeiro identificar a *CVS* para validar essas restrições e porque tal validação é muito complexa, não podendo ainda ser realizada com precisão suficiente.

Há ainda outras frases que também foram incorretamente etiquetadas pela STRING como SUPPORT[vsup-standard], mas que correspondem a erros de processamento sintático. Esperava-se que o sistema reconhecesse uma relação entre um verbo e um nome, e ele reconheceu a relação entre outro verbo e/ou outro nome, como em:

(13.16) *(**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.* [Ex.R]

SUPPORT[vsup-standard](aulas,dá)

SUPPORT[vsup-standard](interpretação,dá)

(13.17) *(**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.* [Ex.R]

SUPPORT[vsup-standard](vantagem,teria)

SUPPORT[vsup-standard](rentabilidade,teria)

Os pares-alvos, cujas dependências deveriam ter sido extraídas, são (*dar, aula*) e (*ter, vantagem*), respectivamente em (13.16) e (13.17), mas o sistema extraiu também, além desses pares-alvos, uma dependência entre *dá* e *interpretação* (13.16) e entre *teria* e *rentabilidade* (13.17). Nesses casos, há uma situação de coordenação de grupos nominais e de extração da dependência de CDIR. A cadeia analisou (13.16) como uma coordenação entre *aulas* e *interpretação*, e não entre *voz* e *interpretação*, considerando-se a coordenação entre dois complementos diretos do verbo *dar*: *dá aulas de voz **e** dá interpretação*. Da mesma forma, a cadeia analisou (13.17) 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*. Assim, as sequências emparelham, inadequadamente, com as condições das regras de extração da dependência SUPPORT, produzindo esses falsos positivos.

No exemplo seguinte, a STRING também analisou incorretamente o verbo *volta* como um nome e coordenou-o com *saidinha*, pelo que também extraiu incorretamente a dependência SUPPORT[vsup-standard] entre *dar* e *volta*:

(13.18) (13.18) *(**dar**, **saída**) À noite, feito criança no mato, ele **dá** uma **saidinha** e volta com duas rãs e um sapo.* [Ex.R]

SUPPORT[vsup-standard](volta,dá)

Esperava-se que a STRING extraísse a dependência entre o *Vsup dá* e o nome predicativo *saidinha*, que é uma derivação de *saída*, mas o sistema, além de não extrair

a dependência desejada, ainda atribuiu incorretamente a POS "nome" (N) ao verbo *voltar*. Nesse sentido, extraiu a dependência entre *dá* e *volta*.

Falhas no processamento das frases, em algum ponto da cadeia STRING, correspondem a 20 erros ao todo. Incluem-se também nessa classe: (i) a identificação de expressões fixas como *CVS*; (ii) a identificação incorreta de nomes próprios ou nomes compostos como se fossem nomes predicativos comuns; e (iii) a classificação indevida de construções com o verbo-operador de ligação *ter* como *CVS standard*. Vejamos cada uma dessas situações pormenorizadamente.

13.2.1.1 A identificação de expressões fixas como *CVS*

O primeiro erro de processamento diz respeito à não-identificação das construções fixas (dependência FIXED) (BAPTISTA; MAMEDE; MARKOV, 2014), devido à ambiguidade que resulta de serem formadas pelos mesmos elementos lexicais que constituem as *CVS*. Essas construções fixas foram, várias vezes, extraídas em conjunto com a dependência SUPPORT[vsup-standard]:

(13.18) *(**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.* [Ex.R]

FIXED(dando,volta,cima)

SUPPORT[vsup-standard](volta,dando)

A expressão *dar a volta por cima* foi analisada pela STRING de duas formas diferentes ao mesmo tempo, e foram extraídas duas dependências diferentes para os mesmos constituintes: (i) como SUPPORT[vsup-standard], semelhante ao predicado *dar uma volta*; e (ii) como uma construção fixa, representado pela dependência FIXED.

Para corrigir esses erros, foi criada uma regra de "limpeza" que dá preferência à extração da dependência FIXED e exclui a dependência SUPPORT[64]. No final do processamento, deverá permanecer apenas a segunda dependência (como expressão fixa), que é a análise correta.

64 Essa regra de limpeza não leva em conta a aridade do conjunto de argumentos da dependência FIXED. Nesse caso, *dar a volta por cima* produz uma dependência FIXED com três argumentos: FIXED(dar,volta,cima), pelo que a decisão de manter a análise de expressão fixa e descartar a análise de *CVS* é mais segura do que no caso em que a aridade de FIXED é apenas dois. Neste último caso, a ambiguidade é mais delicada de resolver e a decisão tomada de dar prioridade às expressões fixas pode produzir resultados menos adequados.

Outras expressões fixas foram analisadas incorretamente tanto pelo sistema STRING quanto pela maioria dos anotadores. Trata-se de frases fixas que possuem algum(ns) constituinte(s) que pode(m) ser analisado(s) como nome predicativo:

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

SUPPORT[vsup-standard](tiro,dando)

(13.21) *(**dar**, **partida**) Nas semanas seguintes, o garoto correu como nunca, suou como nunca, **deu** tudo de si nas **partidas**.* [Ex.R]

SUPPORT[vsup-standard](partidas,deu)

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

SUPPORT[vsup-standard](passo,deu)

O primeiro exemplo (*dar tiro no pé*) é uma expressão fixa, mas possui os constituintes *dar* e *tiro*, que também podem corresponder a uma *CVS*, como em *O policial deu um tiro no bandido*.

O segundo exemplo fora extraído como candidato a *CVS* porque tem o nome *partidas* como adjunto da expressão fixa *dar tudo de si*. Os elementos *dar* e *partida* poderiam formar uma *CVS* em outro contexto sintático, como no caso da frase *O Zé deu partida no carro*.

Por fim, o terceiro exemplo (*dar um passo à frente*) é uma expressão fixa constituída pelos elementos *dar* e *passo*, o que permite a interpretação dessa combinação como uma *CVS*, como ocorre, por exemplo, em *A Ana deu dois passos para o lado*.

Tanto a anotação manual quanto a classificação automática desses casos foram revistas para a segunda avaliação, a fim de manter a consistência da descrição linguística. As expressões *dar tiro no pé*, *dar tudo de si* e *dar um passo à frente* foram inseridas no léxico das expressões fixas da STRING a fim de que apenas a dependência FIXED seja extraída.

13.2.1.2 A identificação de nomes compostos como nomes predicativos simples

O segundo problema de processamento/tokenização diz respeito à identificação de palavras compostas. Nos dados do L-G, constavam apenas as entradas de nomes predicativos simples, como o *Npred golpe*, por exemplo. Esse nome predicativo simples constava na matriz do L-G, mas o sistema não reconhecia as formas *golpe mortal*, *golpe militar* e *golpe de Estado*, que são substantivos predicativos compostos. Nesse sentido, foi necessário acrescentar novas entradas de palavras compostas na matriz para que a tokenização fosse feita corretamente.

Por outro lado, há substantivos compostos (*e.g. declaração de imposto de renda*) que foram analisados como *Npred* simples e, assim, a dependência de SUPPORT[vsupconverse] foi extraída incorretamente:

(13.23) *(**receber**, **declaração**) A falta de informações claras sobre quais agências do Banco do Brasil e da Caixa Econômica Federal abririam durante o final de semana para **receber** a **declaração** do IR (Imposto de Renda) dificultou um pouco a vida do contribuinte que se dispôs a acertar as contas com o "leão" (Receita federal) no final de semana.* [Ex.R]

SUPPORT[vsup-converse](declaração,receber)

A construção *receber uma declaração* é tipicamente uma *CVS* conversa, mas, no caso supracitado, o nome *declaração* não é um nome predicativo, e sim uma parte do nome composto *declaração de imposto de renda*[65], que não é de forma óbvia um nome predicativo, mas antes designa aqui um objeto concreto (é um documento que se chama *declaração de imposto de renda*). Como não se trata de um nome predicativo, não deveria haver dependência SUPPORT entre esse nome e o verbo *receber*. Para solucionar problemas desse tipo, foi necessário inserir no léxico da STRING o nome *declaração de imposto de renda* como um nome composto, e não uma combinação de palavras simples.

13.2.1.3 A classificação indevida de construções com o verbo-operador de ligação *ter* como *CVS standard*

Outro problema de processamento também bastante recorrente se refere à classificação de construções com verbo-operador de ligação como se fossem construções com verbo-suporte.

65 Não tratamos aqui os problemas de formalização resultantes do emprego, neste caso, da abreviatura *IR*, que é também um nome composto, e o fenômeno de sobrecomposição que dá origem à *declaração de imposto de renda*. Sobre composição nominal, veja-se, por exemplo, Baptista (1994).

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

SUPPORT[vsup-standard](nomes,tiveram)

(13.25) *(**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.* [Ex.R]

SUPPORT[vsup-standard](nome,tiver)

Esses dois exemplos são casos típicos de construções com verbo-operador de ligação. Gross (1981, p.32) distinguiu entre duas situações com verbos-operadores: (i) o verbo-operador acrescenta à frase de base um argumento de sentido facilmente identificável: o causativo; e (ii) o operador modifica pouco o sentido da frase; não introduz nenhum argumento novo porque o seu sujeito se liga a um complemento nominal da frase[66].

O primeiro tipo de operador se refere ao *verbo-operador causativo* (ver Capítulo 5) e o segundo tipo é chamado de *verbo-operador de ligação*, pois serve apenas para ligar um argumento que introduz outro argumento que já existia nessa frase de base.

(13.26) *Zé tem # A Ana está sob controle do Zé.* [Ex.C]

= *Zé tem a Ana sob seu controle.* [Ex.C]

Nessa construção, o verbo *ter* funciona como um verbo-operador de ligação (*VopL*), que liga o argumento topicalizado (*Zé*) ao predicado de base (*A Ana está sob controle do Zé*). O argumento *Zé* não é novo; ele já existia na frase de base como complemento do nome *controle*. Ressalte-se ainda que, na construção com verbo-operador de ligação, o pronome possessivo tem que ser correferente ao sujeito da construção, interditando frases como:

(14.26a) = *Zé tem a Ana sob (*meu + *teu + seu) controle.* [Ex.C]

A mesma explicação pode ser feita para a análise das frases (13.24) e (13.25), que poderiam ser analisadas como se indica em (14.24a) e (14.25a):

(14.24a) *Deputados tiveram # Os nomes dos deputados foram encontrados* < ... >. [Ex.C]

= *Os deputados tiveram seus nomes encontrados* < ... >. [Ex.C]

66 Tradução minha. Do original: "(i) l'opérateur apporte un élément de sens facilement identifiable: le causatif; il introduit un argument; (ii) l'opérateur modifie peu le sens de la phrase; il n'introduit pas de nouvel argument, car son sujet se lie à un complément de nom de la phrase" (GROSS, 1981, p.32).

(14.25a) *Se o participante tiver # Se o nome do participante for confirmado <...>.* [Ex.C]

= *Se o participante tiver seu nome confirmado <...>.* [Ex.C]

Há ainda outros casos que também foram classificados pelos anotadores como *VopL* e para os quais a STRING extraiu a dependência SUPPORT[vsup-standard]. 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 (13.27), e outra de sentido ativo, em (13.28), como se observa em:

(13.27) *(**ter**, **notícia**) Zé **teve** uma **notícia** ruim <quando Ana lhe contou sobre a morte do pai>.* [Ex.C]

(13.28) *(**ter**, **notícia**) Zé **tem** uma **notícia** ruim <para dar à Ana>.* [Ex.C]

Os tempos verbais, em rigor o aspecto "pontual" do pretérito perfeito, em (13.27), e o aspecto "durativo" do presente (ou do imperfeito), em (13.28), permitem distinguir esses dois empregos. O exemplo (13.27) é menos controverso, sendo claramente considerado como uma *CVS* conversa, já que tem como contraparte a construção *standard*:

(14.27a) *A Ana deu uma notícia ruim ao Zé.* [Ex.C]

O estatuto dessa construção conversa não gerou dúvidas entre os anotadores. Em contrapartida, o mesmo par (*ter*, *notícia*), em (13.28), 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 (13.28) 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*). De acordo com Santos (2015, p.49), 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 (13.28), também teria o estatuto de um verbo-operador de ligação.

> Estando o sujeito do Vop de ligação já presente na frase de base, o complemento preposicional, no qual esse elemento aparece, é reduzido a um pronome possessivo, com o objetivo de evitar a sua repetição. A função do *Vop* de ligação, neste caso, é a de colocar em posição de tópico o elemento que, na frase de base, se encontra em uma posição sintaticamente menos proeminente. (SANTOS, 2015, p.49).

O mesmo fenômeno pode ser observado em vários outros *Npred* associados ao verbo *ter*. Os exemplos de (13.29) a (13.34), retirados do *corpus* de referência, exemplificam o problema, contrapondo as construções com *Vsup* às frases com *VopL*.

(13.29) *(**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".* [Ex.R]

SUPPORT[vsup-standard](notícia,tenha)

(13.30) *(**ter**, **notícia**) O "The Wall Street Journal" **tem** boas **notícias** para todos vocês, ratos de sofá.* [Ex.R]

SUPPORT[vsup-standard](notícias,tem)

(13.31) *(**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.* [Ex.R]

SUPPORT[vsup-standard](informações,tenha)

(13.32) *(**ter**, **informação**) A página **tem informações** sobre o clube, fotos e os nomes dos membros.* [Ex.R]

SUPPORT[vsup-standard](informações,tem)

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

SUPPORT[vsup-standard](solução,teria)

(13.33) *(**ter**, **solução**) Quem ousaria dizer que **tem** a **solução** para o caso?* [Ex.R]

SUPPORT[vsup-standard](solução,tem)

Para todos esses casos, a STRING extraiu a dependência SUPPORT[vsup-standard]. Os exemplos (13.29), (13.31) e (13.33), no entanto, são *CVS* conversas enquanto os exemplos (13.30), (13.32) e (13.34) não deveriam ter sido extraídos, já que correspondem a construções com *VopL*.

(14.30a) *O jornal **deu** uma notícia a todos.* [Ex.C]

(14.30b) *O jornal **tem** uma notícia **para dar** a todos.* [Ex.C]

(14.32a) *A página **deu** informações sobre o clube aos usuários.* [Ex.C]

(14.32b) *A página **tem** informações sobre o clube **para dar** aos usuários.* [Ex.C]

(14.34a) *Alguém **deu** uma solução para o caso.* [Ex.C]

(14.34b) *Alguém **tem** uma solução **para dar** para o caso.* [Ex.C]

Esse problema, porém, não foi corrigido. Além da presença do complemento introduzido por *para*, não identificamos nenhum outro critério estrutural que diferencie essas duas construções. A mesma ambiguidade entre *CVS standard* e *VopL* permanece na segunda avaliação.

13.2.2 Análise dos falsos-negativos: problemas de não-identificação da dependência SUPPORT

Conforme explicitado na Tabela 11, não foi extraída a dependência SUPPORT em 114 frases. Os principais problemas de não-identificação da dependência pretendida são: (i) a extensão da frase, no caso de frases muito longas; (ii) a distância entre o verbo-suporte e o nome predicativo ou entre a *CVS* e seu sujeito; ou (iii) falha no processamento em alguma etapa anterior da cadeia. Para explicar o primeiro problema, tome-se como exemplo:

(13.35) *(**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.* [Ex.R]

SUBJ_POST(dá,mostras)

Essa frase está incompleta, pois não apresenta a oração principal. O predicado nominal *dar mostras* encontra-se numa oração relativa restritiva, mas o pronome relativo *que* se encontra separado do verbo-suporte por uma oração encaixada (oração reduzida de particípio). Dada a complexidade da frase, a sua incompletude sintática e a interação das regras da gramática, o *parser* analisa *mostras* como sujeito (SUBJ) e não como complemento direto (CDIR) de *dar*. Por essa razão, a dependência SUPPORT não foi extraída. Note-se que, numa frase mais simples, em que a dependência CDIR foi extraída corretamente, o sistema também extrai adequadamente a dependência SUPPORT:

(13.35a) *A revolução democrática **dá mostras** de que deseja ir fundo na moralização.* [Ex.C]

SUPPORT[vsup-standard](mostras,dá)

CDIR_POST(dá,mostras)

Em outros casos, o sistema não extraiu adequadamente a dependência SUPPORT, devido a erro de processamento de anáforas, que ocorre na cadeia em uma etapa anterior à de extração da dependência SUPPORT:

(13.36) *(**dar**, **declaração**) A **declaração** não tem valor legal, já que não foi **dada** em um depoimento formal.* [Ex.R]

SUBJ_PRE_ANAPHO(dada,valor legal)

O *Npred declaração* é o sujeito de uma construção passiva que se encontra subordinada à oração principal. No entanto, a STRING identificou o nome *valor legal* (e não o *Npred declaração*) como sujeito anafórico do verbo *dar*, como demonstra a dependência SUBJ_PRE_ANAPHO extraída incorretamente. O sujeito dessa oração subordinada (*já que [**essa declaração**] não foi dada*) encontra-se elidido, pois já ocorre como sujeito da oração principal (***A declaração** não tem valor legal*). Apesar de a STRING conter um módulo de resolução de anáforas (PEREIRA, 2010), neste caso o sistema não conseguiu capturar adequadamente o sujeito elíptico e, por isso, não extraiu a dependência SUPPORT. No entanto, o sistema captura adequadamente a construção nominal passiva em uma frase cujo sujeito esteja explícito:

(13.36a) *A **declaração** não foi **dada** em um depoimento formal.* [Ex.C]

SUPPORT[vsup-standard](declaração,dada)

SUBJ_PRE(dada,declaração)

Também foram detectadas falhas na extração da dependência SUPPORT em alguns casos de coordenação entre constituintes. Note-se que, por um lado, em alguns casos de coordenação entre nomes predicativos com o mesmo *Vsup*, a STRING extraiu corretamente a dependência de ambos, como sucede em:

(13.36a) *(**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).* [Ex.R]

COORD(e,leitura)

COORD(e,aprovação)

SUPPORT[vsup-converse](aprovação,tiveram)

SUPPORT[vsup-converse](leitura,tiveram)

(13.38) (***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.* [Ex.R]

COORD(nem,resposta)

COORD(nem,confirmação)

SUPPORT[vsup-converse](confirmação,recebeu)

SUPPORT[vsup-converse](resposta,recebeu)

Nesses casos, as dependências de complemento direto (CDIR) que foram extraídas entre (*ter*, *leitura*) e (*receber*, *resposta*) foram percoladas para os outros grupos nominais com que esses nomes se encontram coordenados, nomeadamente *aprovação* e *confirmação*, respectivamente. Assim, apesar de só esses dois últimos nomes predicativos serem o alvo em cada uma dessas frases, o sistema também extraiu a dependência SUPPORT para os primeiros membros da coordenação, já que eles também são *Npred*.

Por outro lado, em outros casos em que a conjunção coordenativa não está explícita, tendo sido substituída por uma vírgula, o sistema não conseguiu extrair a dependência COORD, pelo que também falhou na extração da dependência SUPPORT entre o *Vsup dar* e o *Npred amassos*:

(13.39) (***dar***, ***amasso***) *Ah, Lorena, você só* ***dá*** *uns* ***beijinhos*** *nele, uns* ***amassos*** *e pronto.* [Ex.R]

CDIR_POST(dá,beijinhos)

SUPPORT[vsup-converse](beijinhos,dá)

Na frase (13.39), 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*[67] foi analisada como um adjetivo e não está em um contexto formal que permita formar um grupo nominal 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, neste momento, na STRING, é tratada como um fenômeno estritamente local, ligando grupos nominais e/ou preposicionais, incluindo os casos de enumeração de três ou mais elementos, em que há elipse das conjunções intermédias (*e.g. laranjas, bananas e maçãs*). Se explicitarmos a conjunção coordenativa da frase (13.39) entre os grupos nominais *uns beijinhos* e *uns amassos* (15.39a), então o sistema extrai as duas dependências corretamente:

(13.39a) *Ah, Lorena, você só* ***dá*** *uns* ***beijinhos*** *e uns* ***amassos*** *nele e pronto.* [Ex.C]

COORD(e,beijinhos)

COORD(e,amassos)

SUPPORT[vsup-converse](amassos,dá)

SUPPORT[vsup-converse](beijinhos,dá)

Os 114 casos de falsos-negativos foram testados individualmente na cadeia de processamento, utilizando-se frases simples como exemplos (exemplos simplificados). Para quase 100 deles, foi possível obter a análise adequada, o que significa que o problema não estava nos dados linguísticos do L-G, mas resulta da dificuldade do complexo processo de análise sintática.

13.2.3 Análise dos verdadeiros-negativos: casos em que a STRING acertadamente não extraiu dependência

Dos 139 casos em que a STRING não extraiu dependência SUPPORT, há 25 frases que 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 atribuíram, para essas 25 frases, nenhuma etiqueta de CVS-STANDARD nem CVS-CONVERSO; são, portanto, verdadeiros-negativos. Trata-se de frases cujo par candidato (*Vsup*, *Npred*) não é uma *CVS*.

67 Em rigor, *e pronto* deveria ter sido analisado como um advérbio composto da classe PJC (GROSS, 1986; PALMA, 2009). Aparentemente esse composto ainda não está no léxico da STRING.

(13.40) *(**levar**, **crédito**) Segundo as autoridades, os três **levavam cartões** de crédito e passaportes falsificados quando foram presos, mas não foram encontrados produtos usados na fabricação de bombas.* [Ex.R]

(13.41) *(**ter**, **aula**) Quando estiver pronto será a sede da administração do campus e também **terá** salas de **aula**.* [Ex.R]

(13.42) *(**levar**, **ponto**) Roteiro de um dia **leva** aos **pontos** altos de San Francisco.* [Ex.R]

(13.43) *(**receber**, **título**) Por exemplo, no documento dizia que foi **recebido** a **título** de horas extras CR$ 200 mil.* [Ex.R]

(13.44) *(**ter**, **sorte**) Sua **sorte** foi **ter** sido socorrido com rapidez.* [Ex.R]

Esses exemplos ilustram diversos fenômenos. De (13.40) a (13.42), o sistema reconhece os nomes compostos *cartões de crédito*, *salas de aula* e *pontos altos*, razão por que não emparelha as regras de identificação de *CVS* com os nomes simples, que são elementos de formação desses compostos (*cartões*, *aula* e *pontos*). Uma situação semelhante sucede no caso da locução prepositiva *a título de* (13.43), já estudada por Gross (1986) e Fernandes (2007). Já no caso de (13.44), o sistema identifica a cadeia verbal passiva *ter sido socorrido*, analisada por Baptista, Mamede e Gomes (2010), que lematiza como uma flexão do verbo principal *socorrer*. Assim, o verbo *ter* deixa de estar disponível para emparelhar com as regras de identificação de *CVS*. Ainda assim, tratando-se de uma estrutura de inversão-espelho de uma frase com *ser* (*e.g. Ter sido socorrido foi sua sorte*), a configuração sintática nunca emparelharia com as regras de identificação de *CVS*.

As frases da amostra foram extraídas automaticamente utilizando-se grafos do Unitex, porém esse sistema não utiliza nenhuma informação sintática de *chunking* ou *parsing*, apenas sequências ou combinações de categorias gramaticais. Devido a isso, foram selecionadas para a amostra algumas frases em que não há nenhum tipo de relação sintática entre o *Vsup* e o *Npred*. Esses 25 casos constituem os verdadeiro-negativos, pois não deveriam ser identificados como *CVS* e, de fato, não o foram.

13.3 Casos particulares de verdadeiros-positivos

A seção anterior apresentou os erros de extração automática da dependência SUPPORT nas frases retiradas do *corpus*. Além daqueles erros, devem-se destacar ainda os acertos da STRING em relação às frases que haviam sido mal-extraídas pelo Unitex. Como se usou pouco ou nenhum processamento sintático com o Unitex, para a extração das

frases do *corpus*, o sistema selecionou as combinações de *Npred* e *Vsup* constantes da matriz do L-G, nessa ordem ou em ordem inversa, considerando uma janela com qualquer número de palavras entre os dois elementos-alvo. Assim, alguns pares de (*Vsup, Npred*) foram incorretamente selecionados para o *corpus* a ser anotado, mas foram etiquetados corretamente pelos anotadores como OTHER. Esses casos também foram analisados pela STRING e, em muitos deles, a cadeia não apenas desconsiderou o par indevido, como também encontrou um novo par de *Vsup* e *Npred*, além do par-alvo, extraindo corretamente sua dependência.

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 sem itálico no corpo do exemplo.

(13.45) *(**dar**, **valor**) O varejo, em contrapartida, pode* **dar descontos** *no **valor** cobrado à indústria por determinado espaço na loja.* [Ex.R]

SUPPORT[vsup-standard](descontos,dar)

(13.46) *(**receber**, **prêmio**) Aproveite a semana para conhecer os 12 filmes em cartaz que* **receberam indicações** *ao **prêmio**, como "O Resgate do Soldado Ryan", "Além da Linha Vermelha", "A Vida é Bela" e o recordista "Shakespeare Apaixonado" para a categoria de melhor filme.* [Ex.R]

SUPPORT[vsup-converse](indicações,receberam)

(13.47) *(**ter**, **reajuste**) As categorias com data-base em fevereiro, junho e outubro, pertencentes ao grupo B da política salarial adotada pelo governo,* **terão direito** *a um **reajuste** quadrimestral de 259,3297%.* [Ex.R]

SUPPORT[vsup-standard](direito,terão)

(13.48) *(**ter**, **vôo**) Dois homens suspeitos de pilotarem um dos aviões que se chocou contra uma das torres do World Trade Center* **tiveram aulas** *de **vôo** no aeroporto de Venice (Flórida) e ficaram por uma semana hospedados na casa de um antigo funcionário da escola, a Huffman Aviation.* [Ex.R]

SUPPORT[vsup-converse](aulas,tiveram)

(13.49) *(**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.* [Ex.R]

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, estabelecem a relação sintática relevante para serem identificados como *CVS* e ignora os que não a apresentam.

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, por exemplo:

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

SUPPORT[vsup-standard](condições,tem)

SUPPORT[vsup-converse](ajuda,contar com)

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].

13.4 Segunda avaliação da performance do sistema

Após a análise dos erros verificados na primeira avaliação, procedemos a alterações tanto nos dados do Léxico-Gramática quanto nas anotações do *corpus* de referência. Antes de submeter o *corpus* de referência a uma nova etapa de processamento automático na STRING, foram tomadas as seguintes medidas:

(i) Os dados linguísticos da matriz foram sistematicamente corrigidos, por exemplo, o nome amor estava descrito no L-G com o Vsup *standard dar* e com o Vsup converso *ter*, mas, durante a análise de erros, verificamos que ele também pode aparecer associado ao verbo *ter* em construções *standard* (e.g. A mãe tem amor pelo filho), pelo que acrescentamos esse Vsup à lista de verbos com que amor se constrói. Após revisar sistematicamente esse tipo de erro em toda a matriz, a

STRING passou a extrair, na segunda avaliação, a dependência correta de mais 56 frases. A dependência que estava sendo extraída, na primeira avaliação, para essas 56 sentenças era SUPPORT[vsup-converse], a qual passou a ser SUPPORT[vsup-standard] na segunda avaliação. Isso fez com que o número de verdadeiros-positivos aumentasse e, assim, aumentassem também os valores de precisão, cobertura, acurácia e medida-F.

(ii) Algumas anotações no *corpus* de referência, que estavam inconsistentes, foram revisadas sistematicamente. Por exemplo, em alguns casos, a construção *dar nome* pode significar "nomear alguém, atribuir um nome a alguém" (*e.g. A mãe deu o nome de Zé ao filho*) e, em outros, significa "dizer o nome de alguém" (*e.g. O político deu o nome do seu comparsa*). Quando a frase tem o segundo significado, os anotadores anotaram ora como SVC-*standard* ora como OTHER, por ser um verbo distribucional. Esse tipo de inconsistência foi revisada caso a caso e corrigida quando necessário.

(iii) Foi necessário inserir no dicionário (PB.dic), usado pela STRING, as informações de flexão/derivação de grau dos nomes terminados em *-ada/-ida* que eram marcados no léxico da STRING como verbos no particípio passado ou como adjetivos (*e.g. saidinha*, *fugidinha* e *arrumadinha*). As dependências desses nomes não estavam sendo extraídas porque o sistema não reconhecia os nomes no diminutivo como formas flexionadas/derivadas dos nomes predicativos *saída*, *fugida* e *arrumada*, respectivamente. Ao associar o paradigma de flexão/derivação a esses nomes, a STRING passou a extrair corretamente suas dependências.

(iv) Para algumas palavras, foi necessário completar as entradas do dicionário da STRING, no módulo LexMan, já que este foi construído originalmente com base no vocabulário ortográfico na variante europeia do Português. Assim, esse léxico *ainda* não continha as variantes ortográficas brasileiras *anistia*, *bônus* ou *chute*. Ao associá-las aos lemas (por convenção o sistema usa como lema a forma da variante europeia) *amnistia*, *bónus* e *chuto*, esses nomes passaram a ser corretamente analisados. Outros exemplos de nomes predicativos com grafias diferentes nas duas variantes são: *ajuda econômica* (*ajuda económica*), *chute a gol* (*chuto a golo*), *incentivo econômico* (*incentivo económico*), *indenização* (*indemnização*), *prêmio* (*prémio*), *tesoura* (*tesoira*), *tesourada* (*tesoirada*), dentre outros.

(v) por fim, 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 podem compor uma *CVS* ou, em outras situações,

podem ser expressões cristalizadas. 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.

Após adotar todas essas medidas, submetemos novamente as 580 frases do *corpus* de referência para serem processadas novamente pela STRING. O resultado da segunda avaliação pode ser comparado com o resultado da primeira na Tabela 13:

	TP	FP	FN	TN	Precisão	Cobertura	Acurácia	Medida-F
1ª avaliação	350	91	114	25	79%	75%	65%	77%
2ª avaliação	325	56	84	115	85%	87%	80%	86%

Tabela 13 – Primeira e segunda avaliações da performance da STRING
Fonte: Elaboração própria

Como se nota, todos os valores de Precisão, Cobertura, Acurácia e Medida-F aumentaram. Mesmo depois da correção dos dados do L-G e da revisão das anotações do *corpus* de referência, alguns problemas ainda se mantêm. Por exemplo, os casos de construções com verbo-operador de ligação não foram tratados; os casos aparentemente ambíguos entre SUPPORT[vsup-standard] e SUPPORT[vsup-converse], que dependem da correta identificação dos papéis semânticos, também não foram resolvidos; os casos em que a dependência não foi extraída em decorrência de erro em alguma etapa anterior do processamento na cadeia também permanecem como erros, dentre outros.

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, corrigimos muitos erros de anotação no *corpus* de referência, por isso aumentou o número de frases em que a STRING acertadamente não extraiu dependências de SUPPORT.

Este capítulo apresentou os resultados finais da pesquisa e uma discussão profunda sobre esses resultados. No início, explicamos a estratégia para a análise sintática automática das *CVS* na STRING e o processo de representação e de integração dos dados do L-G no *parser* XIP. Em seguida, apresentamos os resultados de uma primeira avaliação da performance do sistema, com base na comparação com a anotação do *corpus* descrito no capítulo anterior. Os erros encontrados foram analisados e discutidos nas seções referentes aos falsos-positivos, falsos-negativos e verdadeiros-

negativos. Foram identificados erros: (i) na descrição dos dados do Léxico-Gramática; (ii) na anotação que, em um primeiro momento, não havia sido revisada e normalizada; (iii) nos dicionários usados pelo XIP, que não continham algumas palavras que só existem na variante brasileira do Português; e (iv) na extração de algumas dependências que estavam sendo duplicadas.

Após corrigir todos os erros que eram passíveis de correção, processamos novamente o *corpus* de referência na STRING e procedemos a uma segunda avaliação da performance do sistema, o qual melhorou significativamente em todas as medidas de avaliação (precisão, cobertura, acurácia e medida-F).

14 Validação e conversão automática dos dados em regras

Para validar e converter automaticamente os dados linguísticos em regras utilizáveis pelo *parser* XIP, foram construídos em Pearl (.pl) dois programas[68]: (i) o *validador automático*, que verifica a consistência formal e linguística dos dados da matriz do L-G; e (ii) um *conversor automático*, que transforma automaticamente as informações constantes na matriz em regras de extração de dependências, que o XIP usa para determinar as relações sintáticas entre o *Vsup* e o *Npred*.

Tanto pela sua extensão, quanto pela sua complexidade, a elaboração das matrizes é um processo moroso, que está sujeito a variados tipos de erros, sendo necessário desenvolver métodos automáticos de validação formal e linguística dos dados.

14.1 Validação formal

O processo de validação formal dos dados constantes na matriz do L-G foi realizado com base no seguinte procedimento: primeiro, o validador converte cada linha da matriz em um vetor de *strings*; em seguida, valida os valores de cada uma das células com base em critérios previamente definidos.

Inicialmente, é preciso indicar, no programa: (i) quantas linhas iniciais serão ignoradas da validação; (ii) quantas linhas da tabela deverão ser validadas; (iii) em qual coluna está o nome predicativo.

```
my $linhasIniciaisIgnorar = 3; my $linhasReaisProcessar = 1500;

my $colunaLema = 1;
```

Como se pode observar pela notação acima, foram ignoradas as três primeiras linhas, que correspondem ao cabeçalho e nomes das propriedades. Foram processadas as 1.500 linhas seguintes, que abrangem as 1.489 entradas da matriz do L-G. O *Npred*, que é o alvo principal, consta na coluna 1.

Em seguida, o programa estabelece quais são os atributos de cada classe, ou seja, quais elementos podem preencher as células de uma determinada propriedade. Seguem os atributos definidos para as classes, respectivamente, de: (i) papel semântico do argumento na posição sujeito; (ii) papel semântico do argumento na posição de primeiro

68 Agradeço ao Prof. Dr. Nuno Mamede, pesquisador do Instituto de Engenharia de Sistemas e Computadores, Investigação e Desenvolvimento, em Lisboa (INESC-ID), que construiu esses dois programas em Pearl (.pl) e me auxiliou em todas as etapas do processamento automático dos dados na STRING.

complemento; (iii) papel semântico do segundo complemento; (iv) classes sintáticas em PB; (v) preposição que introduz o primeiro complemento; (vi) variantes de verbo-suporte *standard*; (vii) variantes de verbo-suporte converso; e (viii) preposição que introduz o segundo argumento, em construções conversas.

(i) my @SRN0 = ["agent-cause", "agent-gen", "agent-giver", "agent-speaker", "object-gen", "object-cl", "cause", "experiencer-gen", "experiencer-vol", "locative-place", "locative-x"];

(ii) my @SRN1 = ["-", "addressee", "co-agent", "co-object", "locative-dest", "locative-x","locative-place", "locative-source", "locative-path", "object-cl", "object-f", "object-gen", "patient"];

(iii) my @SRN2 = ["-", "topic", "message", "object-f", "object-gen", "patient", "tag", "instrument"];

(iv) my @ClassPB = ["DH1", "DN1", "DR1", "DPC1", "DH2", "DN2", "DR2", "DPC2", "DHN2","DLC2", "DQF2", "DHR2", "D2LC", "DRN2", "D3"];

(v) my @prep1 = ["-", "a", "até", "de", "em", "com", "contra", "para", "por", "sobre", "acerca de", "[preploc]"];

(vi) my @VsupCDIR = ["dar", "fazer", "pegar", "ter", "ser de", "pregar", "apresentar", "atribuir", "aplicar", "ministrar", "bater", "pedir", "tirar", "conceder", "prestar", "impor"];

(vii) my @VsupConv = ["-", "levar", "receber", "pegar", "ter", "tomar"];

(viii) my @Prep1Conv = ["-", "de", "da parte de", "por parte de", "por"];

É importante notar que o rol de papéis semânticos para preencher a posição de sujeito é diferente do rol da posição de complemento. Isso ocorre porque um CO-AGENT, por exemplo, só pode existir como complemento quando o sujeito da construção é um AGENT; outros papéis semânticos também têm posições sintáticas fixas.

Também vale notar que quase todas as propriedades podem ser preenchidas por "-", à exceção do papel semântico na posição sujeito, da classe em PB e do verbo-suporte *standard*. Nos casos em que a construção só possui um argumento, por exemplo, os papéis semânticos do segundo e terceiro argumento são preenchidos com "-". Nos casos em que a construção não aceita a conversão, as posições de variantes de verbo-suporte converso e a da preposição que introduz o segundo argumento, em construções conversas, também serão preenchidas com "-".

Após essa validação, o programa verifica o preenchimento das células que só admitem valores binários. Algumas colunas da matriz só podem ser preenchidas com valores de "+" ou "-". Assim, o programa verifica se essas colunas estão preenchidas apenas por um desses valores. Segue um exemplo dessa validação formal para o preenchimento da posição de *N*1.

["N1=Hum", "S", ["+", "-"]],

["N1=nHum", "S", ["+", "-"]],

["N1=Npc", "S", ["+", "-"]],

["N1=Nloc", "S", ["+", "-"]],

["N1=Npred_de_N", "S", ["+", "-"]],

["N1=N1Vinfw", "S", ["+", "-"]],

["N1=QueFconj", "S", ["+", "-"]],

["N1=QueFind", "S", ["+", "-"]],

["N1=Dativo", "S", ["+", "-"]],

Se houver erro no processo de digitação ou mesmo na formalização dos dados e alguma dessas células estiver preenchida com um valor diferente de "+" ou "-", o programa acusa o erro, sendo necessário corrigir os dados do Léxico-Gramática.

14.2 Validação linguística

O processo de validação linguística considera as relações entre diferentes propriedades de uma mesma entrada (numa mesma linha) e valida os valores de algumas células, de acordo com o preenchimento prévio de outras células. Por exemplo, a coluna referente ao número de argumentos pode ser preenchida com os valores "1", "2" e "3". Se for preenchida com "1", as demais células referentes às preposições, tipos de argumento na segunda e terceira posições e os papéis semânticos desses argumentos devem ser preenchidas com "-", já que essas construções possuem apenas um argumento, que ocupa a posição sujeito.

Para validar os argumentos locativos, o programa verifica a correspondência entre o tipo de argumento (*Nloc*), o papel semântico desse argumento (LOCATIVE-X) e o tipo de preposição ([preploc]). Indicamos a seguir um exemplo de regra de validação que verifica a correspondência entre os argumentos locativos e seus papéis semânticos.

```
if (  ($colunas[26] eq “Loc”) && ($colunas[37] ne “locative-place”) &&
      ($colunas[37] ne “locative-dest”) && ($colunas[37] ne “locative-dest”) &&
      ($colunas[37] ne “locative-path”) )
            { print “**Error col\[26\]=’Loc’ & col\[37\]=/=’locative-...’.
            (line $numeroLinha - $colunas[$colunaLema]).\n”; $error++; }

if (  ($colunas[38] eq “Loc”) && ($colunas[49] ne “locative-place”) &&
      ($colunas[49] ne “locative-dest”) && ($colunas[49] ne “locative-dest”) &&
      ($colunas[49] ne “locative-path”) )
            { print “**Error col\[38\]=’Loc’ & col\[49\]=/=’locative-...’.
            (line $numeroLinha - $colunas[$colunaLema]).\n”; $error++; }
```

Nesses casos, a coluna que indica o tipo de argumento deve ser preenchida com “+” em *Nloc*, com “-” em todos os outros tipos; o papel semântico desse argumento deve ser LOCATIVE-X e a coluna da preposição precisa estar preenchida com o valor “[preploc]”.

Para validar os papéis semânticos dos argumentos, o validador verifica o tipo de argumento. Assim, se o papel semântico está marcado como AGENT, esse mesmo argumento deve ser um *Nhum*. Se o papel semântico indicado for OBJECT-F, então alguma das colunas de completivas deve ser preenchida com “+”.

O programa também faz a validação das classes sintáticas com base em suas características. Para exemplificar, citemos a classe “DH1”, que agrupa construções com apenas um argumento e esse argumento é do tipo humano na posição sujeito. Se a coluna referente à classificação foi preenchida por “DH1”, então o programa verifica: (i) se a coluna referente ao número de argumentos está preenchida por “1”; (ii) se o argumento na posição sujeito está preenchido com “+” na coluna “N0=Nhum” e com “-” em todas as outras colunas referentes ao tipo de sujeito; (iii) se a coluna da preposição está preenchida com “-”, já que não há outros argumentos; e (iv) se todas as demais colunas referentes aos argumentos 2 e 3 estão preenchidas com “-”. O validador possui regras específicas para cada classe sintática.

Para validar os predicados simétricos, ou seja, aqueles que aceitam a transformação da simetria, o validador verifica inicialmente se a coluna referente à simetria está

preenchida com "+". Em seguida, verifica se o papel semântico do primeiro argumento é AGENT-GEN e o papel semântico do segundo argumento é CO-AGENT. A regra que verifica essa correspondência é a seguinte:

```
if (  ($colunas[8] eq "2") && ($colunas[51] eq "+") &&
      ($colunas[18] eq "agent-generic") && ($colunas[37] ne "co-agent"))
            { print "**Error col\[8\]='2' &col\[51\]='+' & col\[18\]='agent-generic' &
            col\[37\]=/='co-agent' (line $numeroLinha - $colunas[$colunaLema]).\n";
            $error++; }
```

Há ainda várias outras regras de correspondência entre colunas da matriz para fazer a validação linguística dos dados do L-G. As que foram aqui citadas são apenas exemplificativas do tipo de validação linguística que o programa executa.

Ao todo, foram construídas 42 regras para a validação formal e linguística dos dados. A construção dessas regras é o resultado da elicitação do conhecimento linguístico representado pelas matrizes e sua formalização, razão pela qual se impõe uma estreita relação entre linguistas e programadores.

14.3 Conversão dos dados em regras de extração de dependência

Dadas as diferenças sintáticas e semânticas que existem entre uma construção com verbo-operador causativo e as construções com verbo-suporte (*standard* e conversas), as regras de dependência também consideram essa distinção. Assim, podem ser extraídas duas dependências diferentes: (i) a dependência SUPPORT e (ii) a dependência VOP-CAUSE. A dependência SUPPORT pode receber ainda dois atributos diferentes, indicados entre parênteses retos, dependendo da orientação de sentido: (i) as construções de orientação ativa dão origem à dependência SUPPORT[vsup-standard]; e (ii) as construções de orientação passiva, à dependência SUPPORT[vsup-converse].

Uma vez validada a matriz do L-G, outro programa gera automaticamente as regras de dependência que serão, em seguida, utilizadas pelo *parser* XIP.

Para analisar algumas regras de dependência, tome-se como exemplo o nome predicativo *beijo*. Conforme mencionado anteriormente, a matriz do L-G contém as informações de quais verbos funcionam como *Vsup standard* e quais funcionam como *Vsup* conversos para cada *Npred* e informações sobre a função sintática que o *Npred*

exerce em relação ao *Vsup*. No caso da *CVS* com *dar*, o nome predicativo é sempre o objeto direto do verbo (*Rui deu um beijo na Ana*), portanto é o sujeito numa construção passiva (*Um beijo foi dado pelo Rui à Ana*). Além da apassivação, também é possível construir uma relativa (*O beijo que o Rui deu na Ana*). Nesse caso, o verbo-suporte deixa de operar diretamente sobre o nome predicativo, mas ainda está inserido na oração relativa, agindo sobre o pronome relativo (*que*), que tem como antecedente o *Npred*. A relativização também pode operar em uma frase passiva (*O beijo que tinha sido dado por Rui na Ana*). Todos esses casos precisam ser levados em consideração no processo de geração das regras.

Para a *CVS dar um beijo*, são geradas quatro regras. Analisemos a primeira delas:

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

As regras de dependência são constituídas de duas partes: primeiramente, uma série de condições (if) são enunciadas. Essas condições precisam ser atendidas para que a dependência seja extraída; em segundo lugar, a dependência a ser extraída é declarada na última linha da regra. Nesse exemplo, a primeira linha determina uma variável (#2) cujo lema é o verbo-suporte *dar*[69]; em seguida, lista-se um conjunto de condições em alternativa. A disjunção entre essas condições é representada por "||", e a conjunção, por "&".

- a segunda condição (linha 2) corresponde à situação em que a *CVS* está em uma oração relativa, o que é dado pela dependência MOD[post,relat] ligando o verbo-suporte no particípio passado com o nome predicativo, que é o antecedente do pronome relativo (*que*). Isso acontece tanto na voz ativa (*Pedro viu o beijo que Rui deu na Ana*) quanto na passiva (*Pedro viu o beijo que foi dado na Ana pelo Rui*);

69 Esse verbo é o verbo principal de uma cadeia verbal constituída por um ou mais verbos auxiliares e uma forma nominal do verbo-suporte (infinitivo, gerúndio ou particípio passado). Sobre o tratamento dos verbos auxiliares na STRING, veja-se Baptista, Mamede e Gomes (2010).

- a terceira condição (linha 3) estabelece que a regra vai disparar se o *Npred* for complemento direto (chamada de dependência CDIR) do verbo *dar*, quando ela não possui o atributo [transf-passiva], o qual é introduzido previamente quando a estrutura passiva é identificada (*Rui deu um beijo na Ana*);
- a quarta condição (linha 4) é complementar à anterior. Ela verifica se já foi extraída a dependência de sujeito (SUBJ) entre o *Npred* e o *Vsup*, quando a construção está na voz passiva (*O beijo foi dado pelo Rui à Ana*);
- a quinta condição é uma conjunção (&) de duas condições (nas linhas 5 e 6). Na linha 5, a condição verifica se o sujeito da passiva tem um pronome relativo como seu antecedente (#4[pronrel]), o qual será o sujeito do *Vsup* na passiva, enquanto
- a linha 6 verifica a dependência de sujeito (SUBJ) entre o pronome relativo e o verbo-suporte na construção passiva. Apesar de parecer que essa condição é redundante com a declarada na linha 4, ela captura uma outra dependência que existe nas orações relativas, a ANTECEDENT na relação anafórica entre o nome e o pronome relativo. Nesse sentido, o *parser* assegura que, mesmo se a relativa não tiver sido identificada pelas dependências anteriores, ainda assim o *parser* consegue extrair a relação de SUPPORT, com base na dependência ANTECEDENT;
- e, finalmente (na linha 7), a regra verifica se a dependência SUPPORT[vsup-standard] ainda não foi extraída entre o *Npred beijo* e o *Vsup dar*.

Se essas condições se verificarem, a regra é aplicada e é extraída a dependência SUPPORT[vsup-standard] entre o nome predicativo e o verbo.

Uma segunda regra para o nome predicativo *beijo* também é produzida:

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

Essa regra difere da anterior por causa da informação de contexto, que é sinalizada entre barras verticais (|) no início e fim da primeira linha. O contexto verifica se o verbo *dar* está na forma de particípio passado (*pastpart*) e se foi etiquetado com o atributo pass-ser, ou seja, se a construção passiva com o verbo auxiliar *ser* já foi identificada previamente. Nesse contexto, a primeira condição (linha 2) verifica se existe uma dependência MOD[ifier] entre o particípio e o nome predicativo para extrair a dependência SUPPORT.

A necessidade dessa regra advém do fato de que o particípio passado é tratado pela STRING como uma categoria gramatical diferente, portanto não dispara as dependências SUBJ or CDIR, agindo como se fosse modificador adjetival.

Um procedimento parecido é realizado para as *CVS* conversas (*e.g. Ana recebeu um beijo do Rui*), produzindo-se as regras correspondentes. As próximas regras cobrem os casos em que a *CVS* admite conversão, desde que essa informação esteja codificada na matriz do L-G.

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

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

A diferença entre essas duas regras e as duas explicadas anteriormente é basicamente o lema do verbo, que passa a ser *receber*, em vez de *dar* (linha 1 em ambas as regras), e, portanto, o tipo de dependência a ser extraída passa a ser SUPPORT[vsup-converse], em vez de SUPPORT[vsup-standard] (linha 8 da primeira regra e linha 4 da segunda regra).

Em capítulos anteriores, apresentamos a metodologia de descrição e formalização dos dados linguísticos em uma matriz do Léxico-Gramática. Essa descrição, por ser feita manualmente, está sujeita a muitos erros de digitação, de formatação e até mesmo de consistência na descrição linguística. No intuito de sanar esses erros, foi criado um programa que valida os dados constantes na matriz do L-G, o qual foi explicado no início deste capítulo. A validação foi feita em dois níveis: (i) nível formal, que corrige problemas

de digitação, formatação *etc.*; e (ii) nível linguístico, que verifica a consistência das propriedades linguísticas descritas na matriz.

Por fim, este capítulo também apresentou o programa que converte os dados do L-G (as *CVS* com suas propriedades sintático-semânticas) em regras de dependência que são usadas pelo XIP para extrair a dependência entre os constituintes de uma *CVS*.

Essas regras foram apresentadas e explicadas. No próximo capítulo, será apresentada a estratégia, a metodologia e os resultados do processo de integração das *CVS* na STRING, por meio do *parser* XIP.

Conclusão

Neste *e-book*, foram analisadas e descritas as construções do Português Brasileiro formadas pelo verbo *dar*. Além de constituírem um número muito significativo de predicados, formando uma parte não negligenciável do léxico do Português, elas ocorrem também com bastante frequência na linguagem cotidiana, tanto em PB quanto em PE. Essas construções também apresentam grande complexidade em suas estruturas e formações sintáticas. Com base no recenseamento dos dados em *corpus*, foram descritas quatro categorias para as construções com o verbo *dar*: como verbo pleno, operador causativo, verbo-suporte e constituinte de frase fixa.

Ao longo do livro, todas essas categorias foram analisadas, utilizando-se exemplos reais ou exemplos construídos para descrever as propriedades sintático-semânticas dessas construções. A mesma classificação proposta nesta pesquisa poderá ser adaptada e expandida para a análise e classificação de outros verbos do Português.

Além dessa proposta de tipologia das construções com o verbo *dar*, esta pesquisa também propôs uma descrição das construções nominais formadas pelo verbo-suporte *dar* e um nome predicativo. Os dados foram descritos a partir de suas propriedades formais, distribucionais e transformacionais, e, em seguida, essas propriedades foram formalizadas em um formato tabular, cujas informações podem ser lidas e utilizadas por computadores, para PLN, bem como por linguistas, para a certificação ou refutação dessas informações.

As 1.489 construções predicativas nominais com o verbo-suporte *dar* foram, posteriormente, integradas no *parser* XIP (MOKHTAR; CHANOD; ROUX, 2002), que é utilizado pela cadeia de processamento do Português chamada STRING (MAMEDE *et al.*, 2012). O *parser* analisa automaticamente as *CVS* com *dar* e identifica-as extraindo a dependência SUPPORT entre o *Vsup* e o *Npred*. O *parser* distingue entre as construções *standard*, as construções conversas e ainda as construções causativas com o verbo *dar*. A inserção dos dados no *parser* foi feita por meio de regras que são geradas automaticamente a partir dos dados codificados na matriz do Léxico-Gramática.

A identificação das *CVS* permite (ou permitirá no futuro) um conjunto de outras operações, distinguindo entre essas construções e as frases em que o verbo *dar* é um verbo pleno, ordinário. Nomeadamente, a atribuição de papéis semânticos aos argumentos do nome predicativo é determinada a partir da informação constante na matriz das *CVS* – e não como sucederia caso estas não tivessem sido identificadas, a partir dos verbos quando empregados como verbo pleno. Por outro lado, no processo de identificação dos

eventos (predicados semânticos) expressos na frase, os *Vsup* não são considerados, mas sim os nomes predicativos que com eles se constroem. Em consequência disso, os argumentos (e outros constituintes) dependentes dos predicados semânticos expressos pelos *Npred* podem assim ser associados aos eventos que os *Npred* representam. Essa estruturação da informação veiculada pelo texto é, pois, semanticamente mais adequada do que a que se teria se os verbos-suporte continuassem a ser analisados como verbos ordinários.

No início do trabalho, ressaltamos quatro objetivos gerais que esta pesquisa visava a atingir: i) uma discussão teórica e prática sobre cada categoria que envolve o verbo *dar*; ii) a descrição formal das propriedades das construções constituídas pelo verbo *dar*, com vistas a aplicações computacionais; iii) a elaboração de uma tipologia para o verbo *dar* em PB, com vistas a subsidiar futuras pesquisas e análises linguísticas; e (iv) a identificação de uma metodologia que permita o processamento automático (ou uma análise sintática automática) das *CVS*. Consideramos, ao fim do estudo, que todos eles foram atendidos.

A discussão teórica das categorias verbais, bem como a elaboração de uma tipologia para as construções com o verbo *dar*, foi feita na Parte I. A proposta tipológica reconheceu cinco categorias sintático-semânticas de que o verbo *dar* pode fazer parte: (i) verbo pleno, em pelo menos oito construções sintáticas diferentes; (ii) verbo-operador causativo, podendo veicular a causa de, pelo menos, quatro tipos semânticos de predicados de base; (iii) constituinte fixo, tanto em expressões cristalizadas verbais quanto em provérbios; e (iv) verbo-suporte, quando selecionado por um nome predicativo.

Especificamente a última categoria (verbo-suporte) foi longamente analisada na Parte II. O objetivo inicial não consistia apenas em identificar o verbo *dar* como verbo-suporte, mas descrever sistemática e formalmente as propriedades do maior número possível de *CVS* que selecionam o *Vsup dar*. Para tanto, foram recenseados 1.489 predicados nominais formados pelo *Vsup dar* e um nome predicativo. Suas propriedades formais, distribucionais e transformacionais foram analisadas, descritas e formalizadas em uma matriz do Léxico-Gramática. Assim, acreditamos que esta obra contribui com um avanço no estado da arte da descrição de predicados nominais para o Português. Contando com os resultados do presente trabalho, estima-se atualmente em cerca de 10.000 os predicados nominais do Português, já descritos e formalizados nos moldes do L-G. O acúmulo de dados e de evidências contribui com o avanço da área e pode subsidiar a construção de recursos léxico-computacionais para o processamento automático da língua portuguesa.

O quarto objetivo visava à identificação de uma metodologia que permitisse o processamento automático das *CVS*. Neste trabalho, não apenas expusemos essa metodologia (RASSI *et al.*, 2014), como também a implementamos e a testamos com uma amostra de frases de um *corpus* de língua real, produzido especificamente para tal, no âmbito deste trabalho (RASSI; BAPTISTA; VALE, 2015). Os resultados alcançados nessa tarefa (Precisão: 79%, Cobertura: 87%, Acurácia: 80% e Medida-F: 86%) nos permitem afirmar que a abordagem baseada em regras para o processamento automático das *CVS* é uma abordagem válida, que poderá ser reproduzida para uma quantidade maior de dados da língua.

Vários outros autores que propuseram diferentes metodologias para o processamento automático de *CVS* não trataram as complexas relações entre os constituintes de uma *CVS*, considerando-as antes como um bloco único de elementos, tal como outras expressões multipalavras[70]. A postura adotada nesta pesquisa, de analisar e descrever as relações entre os constituintes das *CVS*, é linguisticamente mais motivada do que considerar todos os subtipos de expressões multipalavra (palavras compostas, locuções verbais, *CVS*, expressões adverbiais, expressões cristalizadas, *phrasal verbs*, dentre outros) como um fenômeno único. Todos esses subtipos são objetos linguísticos diferentes, que apresentam diferentes níveis de complexidade e funcionamento e, portanto, devem ser estudados e descritos um a um.

Todos os dados e recursos produzidos no âmbito desta pesquisa estão disponíveis para a comunidade acadêmica, a fim de que possam ser utilizados, validados ou contestados em outros trabalhos. As equipes que promovem a construção de dicionários, gramáticas e materiais didáticos poderão se beneficiar com os resultados produzidos. A área do PLN também poderá se beneficiar desses dados no sentido de enriquecer os léxicos computacionais já existentes e de avançar na tarefa de processamento automático de construções nominais. Em geral, reconhece-se a importância da análise sintática dos verbos, mas pouco se conhece sobre a análise sintática dos nomes, que podem funcionar como predicadores da mesma forma que os verbos.

Como contribuições desta pesquisa, ressaltamos:

(i) a proposta tipológica das construções verbais, que poderá ser utilizada ou adaptada para a análise de outros verbos da língua (Capítulo 4);

70 Inversamente, seria possível analisar as *CVS* com *dar* como colocações desse verbo, cada uma com uma acepção diferente. Tal faz perder de vista a unicidade que a noção de *Vsup* e *Npred* aqui adotada permite utilizar, bem como ignora a unicidade do predicado expresso por um *Npred* quando este se constrói com múltiplos *Vsup*, incluindo, em especial, a oposição produzida pelo fenômeno da conversão.

(ii) o recenseamento de cerca de 600 nomes predicativos que ocorrem com o verbo-operador causativo *dar* (Capítulo 5 e Apêndice B). Essas construções não foram descritas na matriz do L-G porque não são construções de base, mas essa lista de *Npred* foi utilizada para descrever as construções de base com outros verbos-suporte, por exemplo, o *Vsup ter* (SANTOS, 2015);

(iii) a proposta formal para a identificação de expressões fixas verbais, com base em critérios formais organizados dicotomicamente (Capítulo 6, Figura 4);

(iv) a classificação das expressões fixas verbais (Capítulo 6, Tabela 2), dos provérbios (Capítulo 6, Tabela 3) e das *CVS* com o verbo *dar* (Parte II), todas elas baseadas em critérios formais e reprodutíveis;

(v) a produção do L-G das *CVS* com o *Vsup dar* em Português do Brasil, que inclui informações linguísticas de diferentes níveis: (a) em nível *morfológico*, em termos de nome autônomo ou nominalização; nome deverbal, deadjetival ou denominal e diferentes sufixos que estão envolvidos na formação dos *Npred*; (b) em nível *sintático*, quanto ao número e tipo de argumentos, determinantes e preposições, além da alternância sintática entre construções *standard* e conversas; e (c) em nível *semântico*, com relação aos traços semânticos dos argumentos do *Npred*, seus papéis semânticos, variantes aspectuais e/ou estilísticas do *Vsup* (ver Apêndice A);

(vi) o tratamento e a descrição dos nomes predicativos que funcionam como predicados *dicendi*, como nomes predicativos de fala (subseção 10.2.5);

(vii) o processamento automático das *CVS* enquanto estruturas sintáticas complexas, que parece ser uma contribuição original na área do Processamento Automático de Línguas Naturais (Capítulo 13).

Além das contribuições apontadas, consideramos ainda que este trabalho poderá ajudar a difundir o modelo do Léxico-Gramática no Brasil. O modelo foi proposto entre as décadas de 70 e 80 na França e ganhou visibilidade em outros países da Europa para o tratamento de diversas línguas, tais como, Catalão, Checo, Coreano, Espanhol, Francês, Italiano, Romeno e várias outras. Apesar de haver alguns trabalhos publicados no Brasil utilizando a metodologia do Léxico-Gramática, esse modelo de análise e descrição linguística ainda não é muito conhecido. Por reconhecermos as vantagens desse tipo de descrição (formalizada e sistemática, visando a exaustividade), acreditamos que a divulgação do modelo do L-G poderá contribuir para o desenvolvimento de novas pesquisas linguísticas no Brasil.

Referências

ABREU, D. T. B. de. *A semântica de construções com verbos-suporte e o paradigma da Framenet*. 2011. Dissertação (Mestrado em Linguística Aplicada) – Unidade Acadêmica de Pesquisa e Pós-Graduação, Universidade do Vale do Rio dos Sinos, São Leopoldo, 2011.

ASSIS, K. L. P. de. *Dar/Fazer/Ter queixa*: queixar-se? A alternância entre construções perifrásticas e verbos plenos correspondentes. 2009. Tese (Doutorado em Letras) – Programa de Pós-Graduação em Letras Vernáculas, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2009.

AUWERA, J. van der; PLUNGIAN, V. Modality's semantic map. *Linguistic Typology*, v. 2, n. 1, p. 79–124, 1998.

BAPTISTA, J. *Estabelecimento e formalização de classes de nomes compostos*. 1994. Dissertação (Doutorado em Linguística Portuguesa) – Faculdade de Letras, Universidade de Lisboa, Lisboa, 1994.

BAPTISTA, J. Sermão, tareia e facada: uma classificação das expressões conversas *dar-levar*. *Seminários de Linguística*, v. 1, p. 5–38, 1997.

BAPTISTA, J. Instrument nouns and fusion. Predicative nouns designating violent actions. *In*: LECLÈRE, C. *et al.* (ed.). Lexique, Syntaxe et Lexique-Grammaire, Hommage à Maurice Gross. *Linguisticae Investigationes Supplementa*, Amsterdam/Philadelphia: John Benjamins Publishing Comp., p. 31–40, 2004.

BAPTISTA, J. Construções simétricas: argumentos e complementos. *Estudos de Homenagem a Mário Vilela*, Porto, p. 353–367, 2005.

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

BAPTISTA, J. Verba dicendi: a structure looking for verbs. *In*: NAKAMURA, T. *et al.* (ed.). Les Tables. La grammaire du français par le menu. Mélanges en hommage à Christian Leclère. *Cahiers du CENTAL*, Louvain-la-Neuve: CENTAL/Presses Universitaires de Louvain, v. 6, p. 11–20, 2010.

BAPTISTA, J. ViPEr: A Lexicon-Grammar of European Portuguese Verbs. *In*: RADIMSKY, J. (ed.). *Actes du 31e Colloque International sur le Lexique et la Grammaire*. République Tchèque: Université de Bohême du Sud, 2012. p. 10–17.

BAPTISTA, J.; CORREIA, A.; FERNANDES, G. Frozen Sentences of Portuguese: Formal Descriptions for NLP. *In*: *Proceedings of ACL Workshop on Multiword Expressions: Integrating Processing (MWE'04)*. Barcelona, Spain: Association for Computational Linguistics (ACL), 2004. p. 72–79.

BAPTISTA, J.; MAMEDE, N.; GOMES, F. Auxiliary verbs and verbal chains in European Portuguese. *In*: PARDO, T. A. S. (ed.). *Proceedings of 9th International Conference on Computational Processing of Portuguese (PROPOR'10)*. Porto Alegre-RS: LNAI 6001, 2010. p. 110–119.

BAPTISTA, J.; MAMEDE, N.; MARKOV, I. Integrating verbal idioms into an nlp system. *In*: BAPTISTA, J. *et al.* (ed.). *Proceedings of Computational Processing of Portuguese Language (PROPOR'14)*. São Carlos: Springer, 2014.

BAPTISTA, J.; RANCHHOD, E. M. Propriétés des phrases élémentaires associées à l'expression de grandeurs mesurables. Exemples du Portugais. *In*: KLEIN, J. R.; LAMIROY, B.; PIERRET, J.-M. (ed.). *Actes du 16a Colloque Européen sur la Grammaire et le Lexique comparés*. Louvain, Belgique: Peeters, v. 1, n. 24, p. 49–61, 1998.

BARREIRO, A. *et al*. Linguistic Evaluation of Support Verb Constructions by OpenLogos and Google Translate. *In*: CALZOLARI, N. *et al.* (ed.). *Proceedings of LREC'14*. Reykjavik, Iceland: European Language Resources Association (ELRA), 2014.

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

BASILIO, M. M. P. Construções morfológicas e construções lexicais: expressões VSN com *dar* e *fazer*. *In*: *Anais do Congresso de Letras da UERJ (CLUERJ)*. São Gonçalo-RJ: Botelho Editora, 2007. vol. único, n. 4, n.p.

BATISTA, Z. N. *Estrutura linguística e informação*: uma introdução à abordagem de Zellig S. Harris sobre os fenômenos da língua. 2008. Dissertação (Mestrado em Letras e Linguística) – Faculdade de Letras, Universidade Federal de Goiás, Goiânia, 2008.

BICK, E. *The parsing system PALAVRAS: automatic grammatical analysis of Portuguese in a constraint grammar framework*. 2000. Tese (Doutorado em Linguística) – Departamento de Linguística, University of Aarhus (AU), Aarhus, DK, 2000.

BICK, E. Automatic Semantic Role Annotation for Portuguese. *In*: *Proceedings of the 5th Workshop in Information and Human Language Technology (TIL'07)*. Rio de Janeiro: Sociedade Brasileira de Computação, 2007. p. 713–1716.

BOONS, J. P.; GUILLET, A.; LECLÈRE, C. *La structure des phrases simples en Français: constructions intransitives*. Genève: Droz, 1976.

BORBA, F. S. *Teoria Sintática*. São Paulo: EdUSP, 1979.

BORBA, F. S. *Dicionário Gramatical de Verbos do Português Contemporâneo do Brasil*. Araraquara: Editora UNESP, 1991.

BORILLO, A. Remarques sur les verbes symétriques. *Langue Française*, v. 11, n. 1, p. 17–31, 1971.

BORILLO, A. Un congé de trois jours, trois jours de congé. *Cahiers de Grammaire*, v. 9, p. 3–20, 1985.

BRATÁNKOVÁ, L. Le costruzioni italiane a verbo supporto: un'analisi condotta sul corpus parallelo ceco-italiano. *Acta Universitatis Carolinae Philologica*, v. 2, p. 55–70, 2013.

BRUCKSCHEN, M. *et al. Anotação Linguística em XML do Corpus PLN-BR*. Série de relatórios do NILC (NILC-TR-09-08). São Carlos, 2008.

BUSSE, W. *Dicionário sintáctico de verbos portugueses*. Coimbra, Portugal: Almedina, 1994.

BUTT, M. Machine translation and complex predicates. *In*: TROST, H. (ed.). *Proceedings of Konvens' 94*. Wien: Springer, 1994. p. 62–71.

BUTT, M.; GEUDER, W. On the (semi)lexical status of light verbs. *In*: CORVER, N.; RIEMSDIJK, H. van (ed.). *Semi-lexical categories: the function of content words and the content of function words*. Amsterdam: Mouton de Gruyter, 2001. p. 323–370.

CALZOLARI, N. *et al.* Towards best practice for Multiword Expressions in Computational Lexicons. *In*: *Proceedings of the Third International Conference on Language Resources and Evaluation (LREC'02)*. Las Palmas, Canary Islands, Spain: European Language Resources Association (ELRA), 2002. p. 1934–1940.

CANÇADO, M. Argumentos: complementos e adjuntos. *Revista Alfa*, v. 53, n. 1, p. 35–59, 2009.

CANÇADO, M. Verbal alternations in Brazilian Portuguese: a lexical semantic approach. *Studies in Hispanic and Lusophone Linguistics*, v. 3, n. 1, p. 77–111, 2010.

CANÇADO, M.; GODOY, L.; AMARAL, L. The construction of a catalog of Brazilian Portuguese verbs. *In*: JANCSARY, J. (ed.). *Proceedings of Workshop on Recent Developments and Applications of Lexical-Semantic Resources (LexSem'12)*. Wien: Ögai, 2012. p. 438–445.

CANTARINI, S. *Costrutti con verbo supporto*: italiano e tedesco a confronto. Itália: Pàtron, 2004. (Le scienze del linguaggio).

CASTELEIRO, J. M. *Sintaxe transformacional do adjectivo*: regência das construções completivas. 1981. Tese (Doutorado em Linguística Portuguesa) – Instituto Nacional de Investigação Científica, Lisboa, Portugal, 1981.

CHACOTO, L. *O verbo FAZER em construções nominais predicativas*. 2005. Tese (Doutorado em Linguística) – Faculdade de Ciências Humanas e Sociais, Universidade do Algarve, Faro, Portugal, 2005.

COELHO, L. C.; SILVA, J. A. A. da; SOUSA, V. V. Análise do fenômeno de gramaticalização do verbo *dar*. *Colóquio do Museu Pedagógico*, v. 10, n. 1, p. 3205–3214, 2013.

COHEN, J. A coefficient of agreement for nominal scales. *Educational and Psychological Measuring*, v. 20, n. 1, p. 37–46, 1960.

CONEJO, C. R. *O verbo-suporte 'fazer' na Língua Portuguesa*: um exercício de base funcionalista. 2008. Dissertação (Mestrado em Letras) – Centro de Ciências Humanas, Letras e Artes, Universidade Estadual de Maringá, Maringá, 2008.

CRUYS, T. V. de; MOIRÓN, B. V. Semantics-based Multiword Expression extraction. *In*: *Proceedings of the Workshop on a Broader Perspective on Multiword Expressions (MWE'07)*. Morristown, NJ, USA: Association for Computational Linguistics (ACL), 2007. p. 25–32.

DAVEL, A. P. C. *Um estudo sobre o verbo-suporte na construção* dar *+ SN*. 2009. Dissertação (Mestrado em Estudos Linguísticos) – Centro de Ciências Humanas e Naturais, Universidade Federal do Espírito Santo, Vitória, 2009.

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)*. Stroudsburg, PA, USA: Association for Computational Linguistics (ACL), 2009. p. 17–22.

DURAN, M.; ALUÍSIO, S. *Verbos auxiliares no Português do Brasil*. Relatório Técnico NILC-05, São Carlos, 2010.

DURAN, M. S.; ALUÍSIO, S. M. Propbank-Br: a Brazilian Portuguese corpus annotated with semantic role labels. *In*: *Proceedings of 8th Brazilian Symposium in Information and Human Language Technology (STIL 2011)*. Cuiabá-MT: Association for Computational Linguistics (ACL), 2011. p. 164–168.

DURAN, M. S. *et al.* Identifying and analyzing Brazilian Portuguese complex predicates. *In*: *Proceedings of Workshop from Parsing and Generation to the Real World (MWE'11)*. Stroudsburg, PA, USA: Association for Computational Linguistics (ACL), 2011. p. 74–82.

FAUCI, N. L.; MIRTO, I. M. *Fare. Elementi di sintassi*. Itália: ETS editore, 2003.

FAZLY, A.; STEVENSON, S. Distinguishing Subtypes of Multiword Expressions Using Linguistically-Motivated Statistical Measures. *In*: *Proceedings of the Workshop on a Broader Perspective on Multiword Expressions (MWE'07)*. Prague, Czech Republic: Association for Computational Linguistics (ACL), 2007. p. 9–16.

FERNANDES, M. da G. M. A. *Léxico-Gramática das frases fixas do Português Europeu - construções intransitivas*. 2007. Dissertação (Mestrado em Linguística) – Faculdade de Ciências Humanas e Sociais, Universidade do Algarve, Faro, 2007.

FILLMORE, C.; BACH, E.; HARMS, R. *The Case for Case*. Berkeley, California, USA: Eric Document Service, 1968.

FLEISS, J. L. Measuring nominal scale agreement among many raters. *Psychological Bulletin*, v. 76, n. 5, p. 378–382, 1971.

FONSECA, J. Verbos simétricos. *Boletim de Filologia – Homenagem a Rodrigo Lapa*, v. 29, n. 2, p. 383–404, 1984.

FRANCHI, C.; NEGRÃO, E. V.; VIOTTI, E. Sobre a gramática das orações impessoais com *ter/haver*. *DELTA*: *Documentação de Estudos em Linguística Teórica e Aplicada*, v. 14, n. especial, p. 105–131, 1998.

GILDEA, D.; JURAFSKY, D. Automatic labeling of semantic roles. *Computational Linguistics*, p. 245–288, 2002.

GIRJU, R. Automatic detection of causal relations for question answering. *In*: *Proceedings of the ACL Workshop on Multilingual Summarization and Question Answering*. Sapporo, Japan: Association for Computational Linguistics (ACL), v. 12, p. 76–83, 2003.

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.

GIRY-SCHNEIDER, J. Les compléments nominaux des verbes de parole. *Langages. Sélection et sémantique. Classes d'objets, compléments apropriés, compléments analysables*, v. 28, n. 115, p. 104–125, 1994.

GREFENSTETTE, G.; TEUFEL, S. Corpus-based method for automatic identification of support verbs for nominalizations. *In*: *Proceedings of the Seventh Conference of the European Chapter of the Association for Computational Linguistics (EACL'95)*. Belfield, Dublin, Ireland: Association for Computational Linguistics (ACL), p. 98–103, 1995.

GROSS, G. Un cas des constructions inverses: *donner* et *recevoir*. *Linguisticae Investigationes*, n. 2, p. 1–44, 1982.

GROSS, G. *Les construction converses du français*. Genève: Droz, 1989.

GROSS, G. Classes d'objets et description des verbes. *Langages*, Paris: Larousse, v. 115, n. 1, p. 15–30, 1994.

GROSS, G. *Les expressions figées en français. Noms composés et autres locutions*. Paris: Editions OPHRYS, 1996. (Collection L'essentiel français).

GROSS, M. *Grammaire transformationelle du français*: *1 - Syntaxe du verbe*. Paris: Cantilène, 1968.

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

GROSS, M. Les bases empiriques de la notion de prédicat sémantique. *Langages*, n. 63, p. 7–52, 1981.

GROSS, M. Une classification des phrases "figées" du français. *Revue Québécoise de Linguistique*, v. 11, n. 2, p. 151–185, 1982.

GROSS, M. *Grammaire transformationnelle du français*: *3 - Syntaxe de l'adverbe*. Paris: ASSTRIL, 1986.

GROSS, M. Methods and Tactics in the Construction of a Lexicon-Grammar. *In*: *Linguistics in the Morning Calm, Selected Papers from SICOL*. Seoul: Hanshin Pub. Co., 1988. p. 177–197.

GROSS, M. La fonction sémantique des verbes supports. *Travaux de Linguistique*, v. 37, n. 1, p. 25–46, 1998.

GUILLET, A.; LECLÈRE, C. Restructuration du groupe nominal. *Langages*, v. 15, n. 63, p. 99–125, 1981.

GUILLET, A.; LECLÈRE, C. *La structure des phrases simples en français*: constructions transitives locatives. Genève: Droz, 1992.

GURRUTXAGA, A.; ALEGRIA, I. Automatic extraction of NV Expressions in Basque: Basic Issues on Cooccurrence Techniques. *In*: *Proceedings of Workshop from Parsing and Generation to the Real World (MWE'11)*. Portland, Oregon, USA: Association for Computational Linguistics (ACL), 2011. p. 2–7.

HALLIDAY, M. A. K. Notes on transitivity and theme in english. part ii. *Journal of Linguistics*, v. 3, p. 199–244, 1967.

HARRIS, Z. S. From morpheme to utterance. *Language*, v. 22, n. 3, p. 161–183, Zellig S. Harris Language July - September 1946.

HARRIS, Z. S. *Methods in Structural Linguistics*. Chicago, Illinois, USA: University of Chicago Press, 1951.

HARRIS, Z. S. *Discourse Analysis*. Dordrecht, Holland: D. Reidel Publishing Company, 1952.

HARRIS, Z. S. *Co-occurrence and transformation in linguistic structure*. Dordrecht, Holland: D. Reidel Publishing Company, 1955. p. 143–210.

HARRIS, Z. S. Strings and transformations in language description. *Papers on formal linguistics*, v. 1, 1961.

HARRIS, Z. S. *Papers on Syntax*. Dordrecht, Holland: D. Reidel Publishing Company, 1964.

HARRIS, Z. S. *Mathematical Structures of Language*. Pennsylvania, USA: Interscience publishers, 1968.

HARRIS, Z. S. *Notes du Cours de Syntaxe*. Paris: Editions du Seuil, 1976.

HARRIS, Z. S. Operator-grammar of English. *Linguisticae Investigationes*, v. 2, p. 55–92, 1978.

HARRIS, Z. S. *A Grammar of English on Mathematical Principles*. New York, USA: John Wiley & Sons, 1982.

HARRIS, Z. S. *Language and Information*. New York: Columbia University Press, 1988.

HARRIS, Z. S. *A Theory of Language and Information: a Mathematical Approach*. New York: Clarendon Press, 1991.

ISTVÁN, N.; VINCZE, V.; FARKAS, R. Full-coverage Identification of English Light Verb Constructions. *In*: *Proceedings of 6th International Joint Conference on Natural Language Processing, organized by the Asian Federation of Natural Language Processing (IJCNLP)*. Nagoya, Japan: Association for Computational Linguistics (ACL), 2013. p. 207–215.

KRIPPENDORFF, K. Estimating the reliability, systematic error, and random error of interval data. *Educational and Psychological Measurement*, v. 30, n. 1, p. 61–70, 1970.

LABELLE, J. Le prédicat nominal avec support *avoir*: contribution à l'étude de la phrase simple. *In*: GUILLET, A. (ed.). *Lexique-grammaire des langues romanes*. Amsterdam, Philadelphia: John Benjamins, 1984. p. 165–198.

LABOREIRO, G.; OLIVEIRA, E. Avaliação de métodos de desofuscação de palavrões. *Linguamática*, v. 6, n. 2, p. 25–43, 2014.

LAMIROY, B. Le lexique-grammaire: essai de synthèse. *Travaux de Linguistique*, v. 37, p. 7–23, 1998.

LAMIROY, B. Les expression figées: à la recherche d'une définition. *Les séquences figées de la langue au discours*, p. 85–88, 2008.

LANDIS, J. R.; KOCH, G. G. The measurement of observer agreement for categorical data. *Biometrics*, v. 33, p. 159–174, 1977.

LANGER, S. A linguistic test battery for support verb constructions. *Linguisticae Investigationes (Special Issue) Verbes supports: nouvel état des lieux*, v. 27, n. 2, p. 171–184, 2004.

LAPORTE, E. The science of Linguistics. *Inference: International Review of Science*, v. 1, n. 2, 2015. Disponível em: http://inference-review.com/article/the-science-of-linguistics. Acesso em: 3 nov. 2022.

LAURA, F. I. *Abordagem multissistêmica da marcação de Tema no Português paulista*. 2013. Tese (Doutorado em Estudos Linguísticos) – Instituto de Estudos da Linguagem, Universidade Estadual de Campinas, Campinas, 2013.

LECLÈRE, C. Remarques sur les substantifs óperateurs. *Langue française. Syntaxe transformationelle du français*, v. 11, n. 1, p. 61–76, 1971.

LECLÈRE, C. Organisation du lexique-grammaire des verbes français. *Langue Française. Les dictionaires électroniques*, v. 87, p. 112–122, 1990.

LECLÈRE, C. Sur une restructuration datif. *Language research Institute Seoul National University*, v. 31, n. 1, p. 179–198, 1995.

LEVIN, B. *English verb classes and alternations*: A preliminary investigation. Chicago, Illinois, USA: University of Chicago Press, 1993.

LOPES, A. C. M. *Texto Proverbial Português*: elementos para uma análise semântica e pragmática. 1992. Tese (Doutorado em Estudos Clássicos) – Faculdade de Letras, Universidade de Coimbra, Coimbra, Portugal, 1992.

MACHONIS, P. A. Support verbs: an analysis of *be prep* X idioms. *The SECOL Review*, v. 12, n. 32, p. 95–125, 1988.

MACHONIS, P. A. The support verb *make*. *In*: *Actes du colloque Computational Lexicography*. Balatonfured, Hungary: Hungarian Academy of Sciences, 1991. p. 141–153.

MAMEDE, N. *et al.* STRING: An hybrid statistical and rule-based Natural Language Processing chain for Portuguese. In: *Proceedings of International Conference on Computational Processing of Portuguese (PROPOR'12)*. Coimbra, Portugal: Springer, 2012. Disponível em: http://www.inesc-id.pt/pt/indicadores/Ficheiros/8578.pdf. Acesso em: 03 nov. 2022.

MAZIERO, E. G.; PARDO, T. A. S.; NUNES, M. G. V. Identificação automática de segmentos discursivos: o uso do parser PALAVRAS, São Carlos, 2007. *NILC-TR-07-08*, Agosto, 2007. Disponível em: https://sites.icmc.usp.br/taspardo/NILCTR0708-MazieroEtAl.pdf. Acesso em: 07 mar. 2023.

MEJRI, S. *Las construcciones verbo-nominales libres y fijas*. Aproximación contrastiva y traductológica. Université Paris XIII: LDI, 2008. cap. Constructions a verbes supports, collocations et locutiones verbales, p. 191–202.

MEL'ČUK, I. Verbes supports sans peine. *Linguisticae Investigationes*, v. 27, n. 2, p. 203–217, 2004.

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

MUNIZ, M. *et al.* Taming the tiger topic: an XCES compliant corpus Portal to generate subcorpus based on automatic text topic identification. In: *Proceedings of the Corpus Linguistics Conference (CL2007)*. Birmingham, UK: University of Birmingham, 2007. Disponível em: http://ucrel.lancs.ac.uk/publications/CL2007/paper/257_Paper.pdf. Acesso em: 03 nov. 2022.

NEVES, M. H. M. Estudo das construções com verbo-suporte em Português. *In*: KOCH, I. (ed.). *Gramática do Português Falado*. Campinas: Editora da UNICAMP, 1996.

NEVES, M. H. M. *Gramática de Usos do Português*. São Paulo: Editora UNESP, 2000.

NEVES, M. H. M. *A gramática*: história, teoria, análise e ensino. São Paulo: UNESP, 2002.

NOIMANN, A. *Um olhar sobre os fraseologismos (locuções) em um dicionário bilíngue escolar Espanhol-Português/Português-Espanhol*. 2007. Dissertação (Mestrado em Letras) – Instituto de Letras, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2007.

OLIVEIRA, M. E. M. *Syntaxe des Verbes Psychologiques Du Portugais*. 1979. Tese (Doutorado em Sciences du Langage) – Université Paris Diderot, Paris 7, 1979.

OLIVEIRA, M. E. M. Les structures de complétives du Portugais. *Linguisticae Investigationes*, v. 7, n. 1, p. 117–134, Amsterdam: John Benjamins B.V. 1981.

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, France: Association for Computational Linguistics (ACL), 2014. p. 419-424.

PALMA, C. *Expressões fixas adverbiais*: descrição léxico-sintática e subsídios para um estudo contrastivo Português-Espanhol. 2009. Dissertação (Mestrado) – Faculdade de Ciências Humanas e Sociais, Universidade do Algarve, Faro, Portugal, 2009.

PALMER, M.; GILDEA, D.; KINGSBURY, P. The Proposition Bank: An Annotated Corpus of Semantic Roles. *Computational Linguistics*, v. 31, n. 1, p. 71–106, 2005.

PEREIRA, S. ZAC.PB: An Annotated Corpus for Zero Anaphora Resolution in Portuguese. *In*: *Proceedings of the Student Research Workshop*. Borovets, Bulgaria: Association for Computational Linguistics (ACL), 2010. p. 53–59.

PRINCE, E. *Radical pragmatics*. New York, USA: Academic Press, 1981. cap. Toward a taxonomy of given-new information, p. 223–256.

RADFORD, A. *Transformational Grammar*: a first course. Cambridge, UK: Cambridge University Press, 1988.

RADIMSKY, J. Noms prédicatifs, noms de résultat et noms concrets dans les constructions à verbe support. *Linguisticae Investigationes*, v. 34, n. 2, p. 204–227, 2011.

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

RANCHHOD, E. M. Uma imagem da gramática. *Revista da Faculdade de Letras*, v. 1, n. 15, p. 109–119, 1993.

RAPOSO, E. P. Uma restrição derivacional global sobre o Infinitivo em Português. *Boletim de Filologia*, v. 24, p. 275–293, 1975.

RASSI, A.; BAPTISTA, J.; VALE, O. Proverb variation: Experiments on Automatic Detection in Brazilian Portuguese Texts. *In*: BAPTISTA, J. *et al.* (ed.). *Porceedings of the 11th International Conference on Computational Processing of the Portuguese Language (PROPOR 2014)*. São Carlos: Springer-Verlag, 2014. v. 8775, p. 137–148.

RASSI, A.; VALE, O. Predicative Nouns Suffixation associated to the verb *dar* (*give*) in Brazilian Portuguese. *In*: BAPTISTA, J.; MONTELEONE, M. (ed.). *Pre-Proceedings of the 32nd International Conference on Lexis and Grammar*. Faro, Portugal: UALG/FCHS, 2013. v. 32, p. 1–7.

RASSI, A. P. *Estatuto sintático-semântico do verbo 'fazer' no Português escrito do Brasil*. 2008. Dissertação (Mestrado em Letras e Linguística) – Faculdade de Letras, Universidade Federal de Goiás, Goiânia, 2008.

RASSI, A. P.; BAPTISTA, J.; VALE, O. Automatic detection of proverbs and their variants. *In*: PEREIRA, M. J. V.; LEAL, J. P.; SIMÕES, A. (ed.). *Proceedings of 3rd Symposium On Languages, Applications and Technologies (SLATE'14)*. Bragança, Portugal: Schloss Dagstuhl, 2014. p. 1–15.

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

RASSI, A. P. *et al.* Estudo contrastivo sobre as construções conversas em PB e PE. *In*: NADIN, O. L. *et al.* (org.). *Léxico e suas interfaces*: descrição, reflexão e ensino. São Paulo: Cultura Acadêmica, 2016. p. 199-218. (Série Trilhas Linguísticas, 29)

RASSI, A. P. *et al.* The fuzzy boundaries of operator verb and support verb constructions with dar "give" and ter "have" in Brazilian Portuguese. *In*: *Proceedings of Workshop on Lexical and Grammatical Resources for Language Processing (LG-LP/COLING)*. Dublin, Ireland: Association for Computational Linguistics (ACL), 2014. p. 92–101.

RUPPENHOFER, J. *et al*. *FrameNet II*: Extended Theory and Practice. Berkeley, California: International Computer Science Institute, 2006.

RUWET, N. Les verbes météorologiques et l'hypothèse inaccusative. *Grammaire et histoire de la grammaire. Mélanges à la mémoire de J. Stefanini. Publications de l'Université de Provence*, n. Aix-en-Provence. Publications de l'Univ. de Provence, p. 383–402, 1986.

SAG, I. A. *et al.* Multiword Expressions: A Pain in the Neck for NLP. *In*: GELBUKH, A. (ed.). *Proceedings of 3rd Computational Linguistics and Intelligent Text Processing (CICLing)*. Mexico City, Mexico: Association for Computational Linguistics (ACL), 2002. p. 1–15.

SALOMÃO, M. M. M. *Polysemy, aspect and modality in Brazilian Portuguese*: the case of a cognitive explanation of grammar. 1990. Tese (Doutorado em Phisolophy in Linguistics) – University of California, Berkeley, Califórnia, 1990.

SALOMÃO, M. M. M. Construções modais com *dar* no Português do Brasil: metáfora, uso e gramática. *Revista Estudos Linguísticos*, v. 16, p. 83–115, 2008.

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

SAUSSURE, F. *Cours de Linguistique Générale*. Lausanne: Payot, 1916.

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

SCHER, A. P. As construções com o verbo leve (CVLs) *dar* e as nominalizações em *-ada* no Português do Brasil. *Sínteses*, v. 10, n. 1, p. 509–522, 2005.

SCHER, A. P. Nominalizações em *-ada*, com o verbo leve *dar* em Português Brasileiro. *Letras de Hoje*, v. 143, n. 1, p. 29–48, 2006.

SMARSARO, A.; ROCHA, L. H. P. da. O uso do verbo *dar* no jogo da linguagem. *Revista (Con)textos linguísticos*, v. 5, n. 5, p. 45–56, 2011.

STEVENSON, S.; FAZLY, A.; NORTH, R. Statistical Measures of the Semi-Productivity of Light Verb Constructions. *In*: *Proceedings of ACL Workshop on Multiword Expressions*: *Integrating Processing (MWE'04)*. Barcelona, Spain: Association for Computational Linguistics (ACL), 2004. p. 1–8.

SUISSAS, G. A. R. *Verb Sense Disambiguation*. 2014. Dissertação (Mestrado em Information Systems and Computer Engineering) – Instituto Superior Técnico/INESC-ID Lisboa - Spoken Language Laboratory, Universidade de Lisboa, Lisboa, Portugal, 2014.

TALHADAS, R. P. *Automatic Semantic Role Labeling for European Portuguese*. 2014. Dissertação (Mestrado em Ciências da Linguagem) – Faculdade de Ciências Humanas e Sociais, Universidade do Algarve (UAlg), Faro, Portugal, 2014.

TAN, Y. F.; KAN, M.-Y.; CUI, H. Extending corpus-based identification of light verb constructions using a supervised learning framework. *In*: *Proceedings of the Workshop on Multiword Expressions*: *Identifying and Exploiting Underlying Properties (MWE'06)*. Trento, Italy: Association for Computational Linguistics (ACL), 2006. p. 49–56.

TOLONE, E.; SAGOT, B.; CLERGERIE, É. V. de la. Evaluating and improving syntactic lexica by plugging them whithin a parser. In: *Proceedings of 8th International Conference on Language Resources and Evaluation (LREC'12)*. Istanbul, Turkey: European Language Resources Association (ELRA), 2012. p. 2742–2749.

TRAVAGLIA, L. C. *Linguagem, conhecimento e aplicação*: estudos de língua e linguística. Rio de Janeiro: Editora Europa, 2003. cap. A gramaticalização de verbos, p. 306–321.

TU, Y.; ROTH, D. Learning English Light Verb Constructions: Contextual or Statistics. *In*: *Proceedings of the Workshop on Multiword Expressions*: from Parsing and Generation to the Real World (MWE'11). Portland, Oregon, USA: Association for Computational Linguistics (ACL), 2011. p. 31–39.

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

VAZA, A. *Estruturas com nomes predicativos e o verbo-suporte 'dar'*. 1988. Dissertação (Mestrado em Estudos Linguísticos) – Universidade Clássica de Lisboa, Portugal, 1988.

VINCZE, V. Light verb constructions in the SzegedParalellFX English–Hungarian Parallel Corpus. *In*: *Proceedings of 8th International Conference on Language Resources and Evaluation (LREC'12)*. Istanbul, Turkey: European Language Resources Association (ELRA), 2012. p. 2381–2388.

WANG, Y.; IKEDA, T. Translation of the Light Verb Constructions in Japanese-Chinese Machine Translation. *Advances in Natural Language Processing and Applications*, v. 33, p. 139–150, 2008.

Apêndice A – As categorias

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53
	Nome predicativo		Classificação			Propriedades N0										Det Npred							Propriedades N1								Propriedades N2										Nominalização						Conversão				
	Npred	PB	PE	Classe_PE	Classe_PB	Arg	N0=Nhum	N0=Nhum	N0=Npc	N0=Nloc	N0=Vinfw	N0=QueFsubj	N0=QueFind	N0=QueFact	Pl obrig N0	N0=Semantic Role	DET=E	DET=Def	DET=Def + Mod	DET=indef	DET=indef + Mod	DET=Poss0	DET=fixo	Prep1_PB	N1=Nhum	N1=Nhum	N1=Npc	N1=Nloc	N1=Vinfw	N1=QueFsubj	N1=QueFind	N1=Semantic Role	Prep2_PB	N2=Nhum	N2=Nhum	N2=Npc	N2=Nloc	N2=Vinfw	N2=QueFsubj	N2=QueFind	N2=Semantic Role	Simetria	N-n	Nome corresp.	Adj-n	Adj corresp.	V-n	V-pleno corresp.	variantes Vsup_Standard	Conversão	variantes Vsup_Converso	Prep1_conversa	P_obr	Exemplo
abafada		+	-	32R	DN1	1	-	+	-	-	-	-	-	-	-	object-gen	-	-	-	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	abafar	dar		-	-	-	-	O tempo deu uma abafada.
abafada		+	-	32R	DHN2	2	+	-	-	-	-	-	-	-	-	agent-gen	-	-	-	+	+	-	-	em	-	+	-	-	-	-	-	object-gen	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	abafar	dar		+	ter	por parte de	-	O Rui deu uma abafada no caso.
abaixada		+	-	31H	DH1	1	+	-	-	-	-	-	-	-	-	agent-gen	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	baixo	+	abaixar	dar		-	-	-	-	A Ana deu uma abaixada.
abaixada		+	-	xxx	DN1	1	-	+	-	-	-	-	-	-	-	object-gen	-	-	-	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	baixo	+	abaixar	dar, ter		-	-	-	-	O nível de água deu uma abaixada.
abalada		+	-	32C	DR1	1	+	+	-	-	-	-	-	-	-	object-gen	-	-	-	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	abalar	dar		-	-	-	-	O prédio deu uma abalada.
abanada		+	-	32C	DHN2	2	+	-	-	-	-	-	-	-	-	agent-gen	-	-	-	+	+	-	-	em	-	+	-	-	-	-	-	object-gen	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	abanar	dar		-	-	-	-	A Ana deu uma abanada na toalha.
abanada		+	-	32CL	DPC2	2	+	-	-	-	-	-	-	-	-	agent-gen	-	-	-	+	+	-	-	em	-	-	+	-	-	-	-	object-cl	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	abanar	dar		-	-	-	-	O cão deu uma abanada na cauda.
abastecida		+	-	36R	DHN2	2	+	-	-	-	-	-	-	-	-	agent-gen	-	-	-	+	+	-	-	em	-	+	-	-	-	-	-	object-gen	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	abastecer	dar		-	-	-	-	O Rui deu uma abastecida na geladeira.
abertura		+	-	xxx	D3	3	+	-	-	-	-	-	-	-	-	experiencer-vol	+	+	+	+	+	-	-	-	a, para	+	-	-	-	-	-	-	patient	para	-	-	-	+	+	-	object-f	-	-	-	-	-	-	+	abrir	conceder, dar		+	receber, ter	de, por parte de	-	O Rui deu abertura para a Ana falar.
abertura		+	-	xxx	DHN2	2	+	-	-	-	-	-	-	-	-	agent-gen	+	-	-	-	+	-	-	a	-	+	-	-	-	-	-	object-gen	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	abrir	dar, fazer		+	ter	-	-	O Rui deu abertura à solenidade.
abocanhada		+	-	32CL	DHN2	2	+	-	-	-	-	-	-	-	-	agent-gen	-	-	-	+	+	-	-	em	-	+	-	-	-	-	-	object-gen	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	abocanhar	dar		-	-	-	-	O Rui deu uma abocanhada no pão.
abonada		+	-	xxx	DHN2	2	+	-	-	-	-	-	-	-	-	agent-gen	-	-	-	+	+	-	-	em	-	+	-	-	-	-	-	object-gen	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	abonar	dar		+	ter	por parte de	-	A Ana deu uma abonada nas faltas.
abordada		+	-	32H	DHR2	2	+	-	-	-	-	-	-	-	-	agent-gen	-	-	-	+	+	-	-	em	+	+	-	-	-	-	-	object-gen	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	abordar	dar		+	levar, receber, ter	de, por parte de	-	O Rui deu uma abordada na Ana.
abordagem		+	-	32C	DHN2	2	+	-	-	-	-	-	-	-	-	agent-gen	+	+	+	-	+	-	-	a, para	-	+	-	-	-	-	-	object-gen	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	abordar	dar, fazer		+	receber, ter	de, por parte de	-	O Rui deu uma abordagem natural ao problema.
abotoada		+	-	32C	DHN2	2	+	-	-	-	-	-	-	-	-	agent-gen	-	-	-	+	+	-	-	em	-	+	-	-	-	-	-	object-gen	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	abotoar	dar		-	-	-	-	O Rui deu uma abotoada na camisa.
abraçada		+	-	32H	DHR2	2	+	-	-	-	-	-	-	-	-	agent-gen	-	-	-	+	+	-	-	em	+	+	-	-	-	-	-	object-gen	-	-	-	-	-	-	-	-	-	-	+	abraço	-	-	+	abraçar	dar		+	levar, receber	de, por parte de	-	O Rui deu uma abraçada na Ana.
abraço		+	+	32H	DH2	2	+	-	-	-	-	-	-	-	-	agent-gen	+	+	+	+	+	+	-	em	+	-	-	-	-	-	-	patient	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	abraçar	conceder, dar		+	receber, ter	de	-	O Rui deu um abraço na Ana.
abrandada		+	-	xxx	DN1	1	-	+	-	-	-	-	-	-	-	object-gen	-	-	-	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	brando	+	abrandar	dar		-	-	-	-	O fogo deu uma abrandada.
abrasileirada		+	-	xxx	DR1	1	+	+	-	-	-	-	-	-	-	object-gen	-	-	-	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	brasileiro	+	abrasileirar	dar, ter		-	-	-	-	O cinema deu uma abrasileirada.
abreviada		+	-	32TA	DHN2	2	+	-	-	-	-	-	-	-	-	agent-gen	-	-	-	+	+	-	-	em	-	+	-	-	-	-	-	object-gen	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	abreviar	dar		+	ter	por parte de	-	O Rui deu uma abreviada no discurso.
abrida		+	-	32C	DHN2	2	+	-	-	-	-	-	-	-	-	agent-gen	-	-	-	+	+	-	-	em	-	+	-	-	-	-	-	object-gen	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	abrir	dar		-	-	-	-	O Rui deu uma abrida na janela.
abrigo		+	+	38L2	D2LC	2	-	-	-	+	-	-	-	-	-	locative-place	+	+	+	+	+	-	-	a, para	+	+	-	-	-	-	-	patient	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	abrigar	conceder, dar		+	receber, ter	de	-	A caverna deu abrigo à Ana.
abrilhantada		+	-	32C	DN1	1	-	+	-	-	-	-	-	-	-	object-gen	-	-	-	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	brilhante	+	abrilhantar	dar		-	-	-	-	O espetáculo deu uma abrilhantada.
absolvição		+	+	32H	DH2	2	+	-	-	-	-	-	-	-	-	agent-gen	+	+	+	-	+	-	-	a, para	+	-	-	-	-	-	-	patient	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	absolver	conceder, dar		+	receber, ter	de, por parte de	-	O Rui deu absolvição à Ana.
absorvida		+	-	32C	DN2	2	-	+	-	-	-	-	-	-	-	object-gen	-	-	-	+	+	-	-	em	-	+	-	-	-	-	-	object-gen	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	absorver	dar		-	-	-	-	A esponja deu uma absorvida no óleo.
abstraída		+	-	35R	DH1	1	+	-	-	-	-	-	-	-	-	experiencer-gen	-	-	-	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	abstrair	dar		-	-	-	-	A Ana deu uma abstraída.
abusada		+	-	34	DHR2	2	+	-	-	-	-	-	-	-	-	agent-gen	-	-	-	+	+	-	-	em	+	+	-	-	-	-	-	object-gen	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	abusar	dar		-	-	-	-	A Ana deu uma abusada no decote.
acabada		+	-	31R	DR1	1	+	+	-	-	-	-	-	-	-	object-gen	-	-	-	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	acabar	dar		-	-	-	-	O Rui + O pneu deu uma acabada <depois do tratamento>.
acabamento		+	+	xxx	DHN2	2	+	-	-	-	-	-	-	-	-	agent-cause	+	-	+	+	+	-	-	em	-	+	-	-	-	-	-	object-gen	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	aplicar, atribuir, dar, fazer		+	receber, ter	por parte de	-	O Rui deu um acabamento na tela.
acalentada		+	-	32H	DPC1	1	-	-	+	-	-	-	-	-	-	object-cl	-	-	-	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	acalentar	dar		-	-	-	-	O coração da Ana finalmente deu uma acalentada.
acalmada		+	-	4	DH1	1	+	-	-	-	-	-	-	-	-	experiencer-gen	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	calmo	+	acalmar	dar		-	-	-	-	A Ana deu uma acalmada.
acalmada		+	-	4	DN1	1	-	+	-	-	-	-	-	-	-	object-gen	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	calmo	+	acalmar	dar		-	-	-	-	O vento deu uma acalmada.
acampada		+	-	35LS	DH1	1	+	-	-	-	-	-	-	-	-	agent-gen	-	-	-	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	acampar	dar		-	-	-	-	O Rui deu uma acampada <no parque>.

Clique aqui para visualizar de maneira completa

Apêndice B – O processo de anotação

Npred	Classe	Exemplo
aborrecimento	<SEN>	A desconfiança de Ana dá aborrecimento ao Rui.
abrangência	<ATR>	O dicionário deu abrangência ao conceito.
acanhamento	<SEN>	A presença de Zé deu acanhamento em Eva.
acanho	<SEN>	A presença de Zé deu acanho em Eva.
aceleração	<DOE>	Anfetamina pode dar aceleração no coração.
aceleração	<ATR>	O motor deu aceleração ao carro.
acesso de loucura	<DOE>	Isso me deu uma acesso de loucura.
acesso de raiva	<DOE>	Isso me deu um acesso de raiva.
acesso de riso	<DOE>	Isso me deu um acesso de riso repentino.
acesso de tontura	<DOE>	Isso me deu uma acesso de tontura
acesso de tosse	<DOE>	Isso me deu um acesso de tosse.
aconchego	<ATR>	A madeira deu um aconchego à casa.
acréscimo	<ATR>	O governo deu uma diminuição na taxa de juros.
adrenalina	<SEN>	O esporte radical deu adrenalina aos competidores.
afeição	<ATR>	O papel de parede deu afeição ao ambiente.
afetividade	<ATR>	A presença das crianças deu afetividade à creche.
aflição	<SEN>	A ausência de Rui deu uma aflição em Ana.
agilidade	<ATR>	O juiz deu agilidade no processo.
agitação	<ATR>	A boate deu agitação ao bairro.
agonia	<SEN>	Dia de prova sempre dá agonia.
agressividade	<ATR>	O jogador deu agressividade ao ataque.
água na boca	<SEN>	Sorvete dá água na boca.
aids	<DOE>	Transar sem proteção com um soropositivo dá aids.
alcance	<ATR>	A família deu um alcance universal para o drama.
alegria	<SEN>	Dinheiro dá alegria.
alento	<SEN>	A vitória deu alento ao povo argentino.
alergia	<DOE>	Comer camarão dá alergia.
aliviada	<SEN>	O medicamento deu uma aliviada na dor.
alívio	<SEN>	A vitória deu um alívio nos brasileiros.
alma	<ATR>	A atriz deu alma à personagem.
alucinação	<DOE>	A droga dá alucinação.
amargor	<ATR>	O álcool dá amargor na bebida.
amargura	<SEN>	O fim do casamento deu amargura em Eva.
ambiente	<ATR>	A luminária deu um ambiente agradável à casa.
ambiguidade	<ATR>	A dupla personalidade deu ambiguidade à cena.
âmbito	<ATR>	Leo deu outro âmbito para o problema.
amnésia	<DOE>	Vodka dá amnésia.
amolecimento	<ATR>	O leite deu amolecimento à massa.
amplidão	<ATR>	O espelho deu amplidão ao ambiente.
amplificação	<ATR>	A caixa deu amplificação ao som.
amplitude	<ATR>	O advogado deu amplitude ao caso.
anemia	<DOE>	A falta de vitamina dá anemia.
angina	<DOE>	Má circulação dá angina.
angústia	<SEN>	A morte de Zé deu angústia em Leo.
animação	<ATR>	A presença dos palhaços deu animação à festa.
ânimo	<SEN>	O trabalho dá ânimo.
ânsia	<DOE>	A presença de Rui deu ânsia em Ana.
ansiedade	<SEN>	A iminência da prova deu ansiedade em Ana.
aparência	<ATR>	A roupa deu uma aparência agradável a Zé.
apatia	<DOE>	O excesso de carboidrato dá apatia.
aperto	<SEN>	A saudade deu um aperto no coração da Ana.
apetite	<SEN>	Malhar me dá muito apetite.
aplicabilidade	<ATR>	O professor deu aplicabilidade à teoria.
apresentação	<ATR>	O ícone deu uma apresentação amigável ao sistema.
aproveitamento	<ATR>	A vitória deu um aproveitamento de 100% ao time.
aptidão	<ATR>	O curso deu aptidão a Ana para fazer maquiagem.
aquecimento	<ATR>	A inflação deu um aquecimento na economia.
ar	<ATR>	O verde deu um ar elegante à sala.
ardor	<SEN>	O remédio deu ardor no machucado.
arejamento	<ATR>	A janela deu arejamento à varanda.
aroma	<ATR>	O perfume deu um aroma agradável à casa.
arrelia	<SEN>	A história deu arrelia em Ana.
arrependimento	<SEN>	Ter casado deu um arrependimento em Rui.
arrepio	<SEN>	Esse casarão deu arrepio em Ana.
aspecto	<ATR>	A interface deu um aspecto amigável ao programa.
astenia	<DOE>	Falta de vitamina C dá astenia.
ataque	<DOE>	Choro de bebê pode dar ataque em idosos.
ataque de nervos	<DOE>	Excesso de nervosismo dá ataque de nervos.

Clique aqui para visualizar de maneira completa

Apêndice C – Matriz do Léxico-Gramática

Classe	N0=:Nhum	N0=:N-hum	N0=:QueF	N0=:Nnr	Neg Obrig	V	Prep_C1	Det	Prep+Det	Modif_C1	C1	Modif_C1	Prep_C2 (Gen)_Prep1		C2	Prep	N0=:Nhum	N0=:N-hum	N0=:QueF	N0=:Nnr	Prep_C3	C3	Exemplo
C1PN	+	-	-	-	-	dar	-	o	-	-	anel	-	-	-	-	a, para	+	-	-	-	-	-	O João deu o anel para o Rui
C1PN	+	-	-	-	-	dar	-	o	-	-	anel	de couro	-	-	-	a, para	+	-	-	-	-	-	O João deu o anel de couro para o Rui
C1PN	+	-	-	-	-	dar	-	o	-	-	ânus	-	-	-	-	a, para	+	-	-	-	-	-	O João deu o ânus para o Rui
C1	+	-	-	-	-	dar	-	-	-	-	ares	de importância	-	-	-	-	-	-	-	-	-	-	A Rita dá ares de importância
C1PN	+	-	-	-	+	dar	-	-	-	-	assunto	-	-	-	-	a, para	+	+	-	-	-	-	O João não deu assunto para a Ana
C1PN	+	+	+	+	-	dar	-	-	-	-	assunto	-	-	-	-	a, para	+	+	-	-	-	-	A tragédia deu assunto para a reportagem
C1PN	+	+	-	-	-	dar	-	o	-	-	baile	-	-	-	-	em	+	+	-	-	-	-	O Goiás deu o baile no Vila Nova
C1	+	-	-	-	-	dar	-	-	-	-	baixa	-	-	-	-	-	-	-	-	-	-	-	A Ana deu baixa <na festa>
C1PN	+	-	-	-	-	dar	-	-	-	-	banana	-	-	-	-	a, para	+	-	-	-	-	-	O jogador deu banana para a torcida
C1	+	-	-	-	-	dar	-	-	-	-	bandeira	-	-	-	-	-	-	-	-	-	-	-	A Ana deu bandeira <na festa>
C1PN	+	-	-	-	-	dar	-	o	-	-	banho	-	-	-	-	em	-	+	-	-	-	-	O ladrão deu o banho na carteira do transeunte
C1PN	+	-	+	-	-	dar	-	um	-	-	banho	de água fria	-	-	-	em	+	+	-	-	-	-	O João deu um banho de água fria na Ana
C1PN	+	-	-	-	-	dar	-	um	-	-	banho	de loja	-	-	-	em	+	-	-	-	-	-	O João deu um banho de loja na Ana
C1PN	+	-	-	-	-	dar	-	um	-	-	banho	de entusiasmo	-	-	-	em	+	-	-	-	-	-	O João deu um banho de entusiasmo na Ana
C1	+	-	-	-	-	dar-se	-	-	-	-	bem	-	-	-	-	-	-	-	-	-	-	-	A Maria se deu bem.
C1PN	+	-	-	-	-	dar-se	-	-	-	-	bem	-	-	-	-	com	+	-	-	-	-	-	O João deu-se bem com a Ana
C1PN	+	-	-	-	-	dar	-	o	-	-	biscoito	-	-	-	-	a, para	+	-	-	-	-	-	O João deu o biscoito para o Rui
C1	-	-	+	-	-	dar	-	-	-	-	bode	-	-	-	-	-	-	-	-	-	-	-	A CPI do Mensalão deu bode.
C1PN	+	-	-	-	-	dar	-	o	-	-	boga	-	-	-	-	a, para	+	-	-	-	-	-	O João deu o boga para o Rui
C1PN	+	-	-	-	-	dar	-	-	-	-	bola	-	-	-	-	a, para	+	+	-	-	-	-	O João deu bola para a Ana
C1	+	-	-	-	-	dar	-	-	-	-	bola	fora	-	-	-	-	-	-	-	-	-	-	O João deu bola fora <na festa>
C1PN	+	-	-	-	-	dar	-	-	-	-	bola	-	-	-	-	a, para	+	-	-	-	-	-	A Ana deu bola para o Rui
C1PN	+	-	-	-	-	dar	-	o	-	-	botão	-	-	-	-	a, para	+	-	-	-	-	-	O João deu o botão para o Rui
C1P2	+	-	-	-	-	dar	-	o	-	-	braço	-	-	a	torcer	-	-	-	-	-	-	-	O João deu o braço a torcer
N0=0	-	-	-	-	-	dar	-	um	-	-	branco	-	-	-	-	em	+	-	-	-	-	-	Deu um branco na Ana
C1PN	+	-	-	-	-	dar	-	o	-	-	brioco	-	-	-	-	a, para	+	-	-	-	-	-	O João deu o brioco para o Rui
C1PN	+	-	-	-	-	dar	-	a	-	-	bunda	-	-	-	-	a, para	+	-	-	-	-	-	O João deu a bunda para o Rui
CPP	+	-	-	-	-	dar	com	os	-	-	burros	-	-	na	água	-	-	-	-	-	-	-	O João deu com os burros n'água
C1PN	+	-	-	-	-	dar	-	-	-	-	cabo	-	-	-	-	de	+	-	-	-	-	-	A Ana deu cabo do Rui
C1P2	-	+	+	-	-	dar	-	-	-	-	calo	-	-	na	vista	-	-	-	-	-	-	-	A beleza da mulher dá calo na vista
C1P2	+	-	-	-	-	dar	-	a	-	-	cara	-	-	a	tapa	-	-	-	-	-	-	-	O João deu a cara a tapa
C1P2	+	-	-	-	-	dar	-	a	-	-	cara	-	-	pra	bater	-	-	-	-	-	-	-	O João deu a cara para bater
CPP	+	-	-	-	-	dar	com	a	-	-	cara	-	-	na	porta	-	-	-	-	-	-	-	O João deu com a cara na porta
CPP	+	-	-	-	-	dar	com	a	-	-	cara	-	-	na	parede	-	-	-	-	-	-	-	O João deu com a cara na parede
CPP	+	-	-	-	-	dar	de	-	-	-	cara	-	-	com	Nhum	-	+	-	-	-	-	-	O João deu de cara com a Ana
C1PN	+	-	-	-	-	dar	-	-	-	-	carta	branca	-	-	-	a, para	+	-	+	-	-	-	O governo deu carta branca para o advogado soltar os presos
C1PN	+	-	-	-	-	dar	-	a	-	-	cartada	final	-	-	-	a, para	+	+	+	+	-	-	O João deu a cartada final para a Ana
C1P2	+	-	-	-	-	dar	-	as	-	-	cartas	-	-	do	jogo	-	-	-	-	-	-	-	O João deu as cartas do jogo
C1PN	+	-	-	-	-	dar	-	-	-	-	casa, comida e roupa lavac	-	-	-	-	a, para	+	-	-	-	-	-	O João deu casa, comida e roupa lavada à Ana
CPP	+	+	-	-	-	dar	de	-	-	-	cem	-	-	a	zero	em	+	+	-	-	-	-	O João dá de cem a zero no Zé

Clique aqui para visualizar de maneira completa

Apêndice D – *Guidelines* para a anotação

Esta *guideline* fornece as instruções para a anotação de construções com verbos-suporte e construções com verbo operador causativo em Português, a fim de constituir um *golden standard* para a avaliação de tarefas de identificação automática de construções com verbo-suporte e com verbo operador causativo.

1. As categorias

Esta anotação será feita usando a ferramenta Corpus Annotator.Class. A anotação consiste em assinalar, para cada frase, um código (convencional) que corresponde ao tipo de construção sintática indicada pelo par (*Vsup*,*Npred*) entre parênteses no início de cada frase. As etiquetas a serem usadas são:

• **CVS-STANDARD** - para as construções com verbo-suporte *standard*;

Ex.: Ana deu um tapa em Rui.

• **CVS-CONVERSE** - para as construções com verbo-suporte converso;

Ex.: Rui levou um tapa da Ana.

• **VOPC** - para as construções com verbo operador causativo;

Ex.: O vento sombrio deu medo em Ana.

Ex.: O vento sombrio fez com que Ana tivesse medo.

• **Nenhuma das anteriores/Não sei** - para qualquer outro tipo de construção (com verbo pleno, expressão fixa ou outros);

Ex.: Rui deu zebra ao ganhar a competição.

Ex.: Dilma fez a egípcia ao ser questionada sobre corrupção.

Ex.: Rui deu uma bicicleta para Ana.

Ex.: Ana tem uma bicicleta.

Ex.: Rui tem Max sob seu controle.

Na amostra, certamente há apenas um par candidato a *CVS*. Esse par (*Vsup*, *Npred*) está indicado entre parênteses no início da frase a ser anotada. No caso de haver dois ou mais possíveis candidatos a *CVS* na mesma frase, considere APENAS o par que está indicado nos parênteses antes da frase. Por exemplo:

(ter, acesso) Além de ter pouco espaço para as pernas dos ocupantes do banco de trás, o carro não tem bom acesso.

Da frase acima, poderíamos retirar dois candidatos a CVS: (ter,espaço) e (ter,acesso), porém somente o par que está entre parênteses no início da frase é que deve ser anotado, portanto, anote-se APENAS o par (*ter,acesso*) e desconsidere o(s) outro(s).

Em algumas sentenças da amostra, há uma notação como "[CVS: Vsup Det Npred] (Vsup,Npred)". Esses casos não foram corretamente normalizados, portanto, deve-se desconsiderar essa notação, mas anotar a frase de acordo com o par candidato (*Vsup,Npred*) que está entre parênteses no início da frase. Por exemplo:

(fazer,pesquisa) Quando [CVS: receber o sinal] (receber,sinal) verde, a pesquisa será feita pela GV Consult, consultoria subordinada à Fundação Getúlio Vargas em São Paulo, já contratada.

Nesse caso, deve-se desconsiderar o par (*receber,sinal*) e anotar APENAS o par entre parênteses no início da frase (*fazer,pesquisa*)

2. O processo de anotação

O procedimento para a anotação é o seguinte:

- Crie uma pasta no seu computador (*p. ex.* em "MeusDocumentos") com as iniciais do seu nome (JB, OV, AR, CB, CT);

- Copie para essa pasta todos os arquivos em anexo a este *e-mail*:

• SVC-amostra-1.0.txt – que contém as frases a serem anotadas para cada par de candidatos a *CVS* (*Vsup,Npred*);

• SVC-amostra-1.0-Parametrizacao.txt – que contém todas as formas dos nomes predicativos; e

• o arquivo 'CorpusAnnotator.class', que é o programa a ser usado para a anotação.

O programa de anotação 'CorpusAnnotator.class' corre em Java, pelo que precisa instalar previamente o JRE environment (se já tiver instalado o Unitex, não precisa fazer nada, pois ele também corre em Java). O 'CorpusAnnotator.class' foi desenvolvido especificamente (SUÍSSAS, 2014) para que o processo de anotação seja mais rápido e fácil.

O programa precisa de dois arquivos:

• arquivo SVC-amostra-1.0.txt; e

• arquivo SVC-amostra-1.0-Parametrizacao.txt

Para abrir o programa de anotação, abra uma janela (Terminal) com linha de comando (cmd) e digite o caminho <path> na árvore de diretórios para a pasta para onde copiou os seus arquivos:

cd <path>/AXXXXX-Verbos

Instale o programa de anotação digitando o seguinte comando:

java CorpusAnnotator

ATENÇÃO: não use a extensão .class!

Abre-se então a janela de diálogo inicial. Nessa janela clique nos botões "Escolher ficheiro" para selecionar:

• o arquivo SVC-amostra-1.0.txt; e

• o arquivo SVC-amostra-1.0-Parametrizacao.txt

Em seguida, inicie o processo clicando no botão "Continuar". O programa abrirá outra janela, mostrando a primeira frase a ser anotada. Repare que o programa indica os nomes predicativos em negrito, mas o anotador deve desconsiderar essa formatação e considerar, para a anotação, APENAS o par (Vsup,Npred) indicado entre parênteses no início da frase.

Para anotar, clique em uma das etiquetas disponíveis (SVC-STANDARD, SVCCONVERSE, VOPC ou Nenhuma das anteriores/Não sei) e, em seguida, clique em "seguinte" para mostrar a próxima frase e continuar a anotação. No final de toda a anotação, clique no botão "Guardar e sair" para salvar sua anotação.

Quando clicar em "Guardar e sair" ou em "Guardar processo", pela primeira vez, você será direcionado ao diretório onde deseja salvar a sua anotação. Peço que salve o arquivo com as iniciais do seu nome, seguido de "-out", por exemplo: JB-out, OT-out, ARout, CB-out CT-out, na mesma pasta onde estão todos os outros arquivos da anotação.

De tempos em tempos, sugiro clicar em "Guardar processo", só para garantir que você não perderá o trabalho feito, em caso de queda de energia ou outros.

Depois de anotar as 2.646 frases, envie-me por *e-mail* apenas o arquivo XX-out (amandarassi85@gmail.com) para que eu possa calcular o inter-annotator agreement.

Se você precisar interromper o trabalho no meio, tome nota do número da frase em que se encontra e salve o seu arquivo (ver adiante). Quando quiser voltar à anotação, você poderá abrir o programa e recomeçar a anotação da frase em que parou, indicando o respectivo número na janela "Começar a partir da linha:". Se você não pedir para começar da linha "x", o programa te conduzirá a anotar todas as frases de novo, desde o início.

No canto superior esquerdo, um contador indica o número da frase corrente. O total de frases a serem anotadas por cada anotador é 2.646.

Além das opções de etiquetas (SVC-STANDARD, SVC-CONVERSE, VOPC ou Nenhuma das anteriores, Não sei), o programa também oferece outras duas opções ("Tenho dúvidas" e "Erro na frase").

- A caixa "Tenho dúvidas" pode ser selecionada mesmo que tenha selecionado alguma das etiquetas anteriores, para poder retornar, posteriormente, somente às frases em que você ficou com dúvida.

- A caixa "Erro na frase" deve ser selecionada caso a frase esteja truncada, incompleta ou não apresente contexto suficiente para determinar qualquer etiqueta.

Os restantes botões servem para o funcionamento geral do programa:

- "Anterior": recua para a frase anterior

- "Seguinte": avança para a frase seguinte

ATENÇÃO: É aconselhável marcar "Tenho dúvidas" sempre que não estiver absolutamente seguro(a) da classificação a dar à CVS. Assim, permite que outra pessoa possa vir a decidir, mais tarde, qual é a melhor solução nesse caso.

CAPES

Publique com a gente e
compartilhe o conhecimento

Letraria®

www.letraria.net

Letraria®