

UNIVERSIDADE FEDERAL DE SÃO CARLOS
CENTRO DE EDUCAÇÃO E CIÊNCIAS HUMANAS
PROGRAMA DE PÓS-GRADUAÇÃO EM LINGUÍSTICA

**DEVERBAIS DE AÇÃO EM *CORPUS* HISTÓRICO:
CONTRIBUIÇÃO COMPUTACIONAL PARA A MORFOLOGIA
CONSTRUCIONAL DO PORTUGUÊS**

Livia Aluisi Cucatto

Dissertação apresentada ao Programa de Pós-Graduação em Linguística da Universidade Federal de São Carlos, como parte dos requisitos para obtenção do Título de Mestre em Linguística.

Orientadora: Profa. Dra. Sandra Maria Aluísio

Co-orientadora: Profa. Dra. Margarita Correia

São Carlos, São Paulo, Brasil
Dezembro – 2009

**Ficha catalográfica elaborada pelo DePT da
Biblioteca Comunitária da UFSCar**

C963da

Cucatto, Livia Aluisi.

Deverbais de ação em corpus histórico : contribuição computacional para a morfologia construcional do português / Livia Aluisi Cucatto. -- São Carlos : UFSCar, 2011.
181 f.


Dissertação (Mestrado) -- Universidade Federal de São Carlos, 2009.

1. Linguística - processamento de dados. 2. Nome deverbal. 3. Corpus histórico. 4. Morfologia construcional. 5. Recuperação da informação semi-automática. I. Título.

CDD: 410.285 (20ª)

BANCA EXAMINADORA

Profa. Dra. Sandra Maria Aluisio



Prof. Dr. Thiago Alexandre Salgueiro Pardo



Profa. Dra. Gladis Maria de Barcellos Almeida



À minha querida família.

Agradecimentos

À Prof.^a Sandra Maria Aluísio. Agradeço-lhe pela confiança depositada, pela orientação, pela amizade e principalmente pelo entusiasmo contagiante, amor e comprometimento que sempre demonstrou em relação à profissão e à área desafiadora da linguística computacional.

À Prof.^a Margarita Correia pela orientação, paciência e, sobretudo, pela amizade construída.

À Prof.^a Gladis Maria de Barcellos Almeida por me ajudar a encontrar a linguística computacional e o NILC. Ao Prof. e amigo Thiago Pardo por ter me recebido tão bem no NILC, ter me ensinado muitas coisas e compartilhado de momentos filosóficos nos saudosos cafés nilquenses. À ambos pela leitura atenta do trabalho e pelas contribuições muito produtivas dadas no exame de qualificação.

A todos os colaboradores do NILC, em especial à Prof.^a e coordenadora Graça Volpe Nunes por ter me aceitado e me proporcionado o contato com pessoas tão brilhantes.

À professora Eliane Hércules por ter me direcionado para área de linguística computacional.

Aos meus chefes Jussara Melo Vieira e Edmilson da Silva Morais pela compreensão, ajuda e torcida para o desenvolvimento deste mestrado. Aos colegas da Vocalize – em especial, Bruno Cedraz Brandão e Rodolfo Rodrigues de Araujo pela acessória computacional; à Roberta Cruvinel Baptistella, Gilberto Luis Valente da Costa e Gláucya Carreiro Boechat pela acessória técnica e linguística.

À minha querida mãe, ao meu querido tio Bento, e aos meus irmãos lindos Felipe e Fábio pelo apoio e compreensão em relação a minha ausência de corpo, mente e algumas vezes de alma nos últimos anos. Ao meu pai, Judith, Yzadora e Marina por sempre me motivarem e torcerem pelo meu sucesso e felicidade. Sou eternamente grata por tudo que fizeram e que ainda fazem por mim.

Aos meus amigos por contribuírem para que minha vida seja cada vez mais completa.

Finalmente, agradeço a Deus, por cuidar de mim e por ter me dado muita energia para a conclusão deste trabalho.

A mente é um estômago. Há muitos tipos de mente-estômago. Alguns se parecem com os estômagos humanos e processam os mais variados tipos de informações. Leonardo da Vinci é um exemplo extraordinário desse estômago omnívoro, capaz de digerir poesia, música, arquitetura, urbanismo, pintura, engenharia, ciência, criptografia, filosofia. Outros estômagos se especializaram e só são capazes de digerir um tipo de alimento.

Rubem Alves, 2004.

Resumo

Nos últimos anos, têm-se desenvolvido trabalhos em morfologia construcional que visam à descrição dos mecanismos de construção de nomes deverbais, isto é, nomes derivados de verbos, numa perspectiva sincrônica. As razões para esse interesse podem encontrar-se na produtividade dos mecanismos de construção deste tipo de nomes, cujo impacto no uso do português, sobretudo em contextos formais de uso, não é irrelevante. Faltam, no entanto, estudos que nos permitam conhecer, por um lado, o Português do Brasil (PB) em nível destes mecanismos e, por outro, ter uma perspectiva diacrônica deles. Acreditamos que os mecanismos de construção atuais para o PB sejam iguais aos dos séculos XVI, XVII e XVIII, podendo existir alterações relacionadas aos sufixos mais produtivos e à mudança dos processos de construção mais produtivos para determinadas palavras. A realização do projeto *Dicionário Histórico do Português do Brasil (DHPB)* dos séculos XVI a XVIII, financiado pelo Programa Institutos do Milênio, associada à construção do *corpus* coligido para esse efeito constituem ao mesmo tempo um desafio e uma oportunidade para o aprofundamento do conhecimento sobre este aspecto da língua portuguesa. Por um lado, a análise do *corpus* dá-nos a oportunidade de verificar como se processou a evolução destes mecanismos de construção no português. Porém (e este é o desafio), não existem ainda ferramentas para realizar tais pesquisas de forma (semi-) automática, que permitam aos morfologistas obter os dados necessários para o seu trabalho com agilidade e eficiência. O objetivo desta pesquisa de mestrado foi, em primeiro lugar, descrever os diferentes mecanismos de construção de nomes deverbais em PB de acordo com o modelo de morfologia construcional SILEX (cf. Corbin 1987, 1991, 1997, Correia 1999, Rio-Torto (org.) 2004 e Rodrigues 2006) e, em segundo, desenvolver um sistema computacional denominado EXTRADEV que permitiu um acesso fácil aos seguintes dados: (a) nomes deverbais de ação históricos com as mais diversas estruturas morfológicas; e (b) variantes gráficas desses nomes deverbais (históricos), que facilitem o trabalho de recuperação de informação relativa aos mesmos. A metodologia utilizada para a construção deste sistema se embasa: (i) na descrição dos deverbais e na construção de regras; (ii) em um estudo piloto dos cinquenta verbos mais frequentes do *corpus* do projeto DHPB, extraídos por meio da ferramenta UNITEX e na análise das variantes gráficas destes verbos; (iii) no conhecimento da linguagem de programação Python e expressões regulares; e (iv) no uso de recursos criados no projeto DHPB, como o sistema de geração de variantes gráficas SIACONF. Foram encontrados 1.742.663 ocorrências de deverbais de ação e 15.633 formas distintas dos mesmos sem variação de grafia. Este número, mais as variantes extraídas no segundo módulo do EXTRADEV, somam um total de 22.442 ocorrências de deverbais históricos, sendo eles 6.809 variantes e 15.633 deverbais sem variação de grafia. Seguimos os seguintes critérios de análise: frequência, análise da estrutura morfológica com auxílio de dicionários históricos e etimológicos, e observação da lista final de deverbais históricos. Com este trabalho pretendemos contribuir para um melhor conhecimento da variação diacrônica na construção de nomes deverbais através dos dados que foram encontrados, mas, sobretudo, motivar a aliança entre linguística e a ciência da computação, particularmente o processamento de língua natural, de modo a potencializar estudos futuros sobre a língua portuguesa.

Palavras-chave: deverbais, *corpus* histórico, morfologia construcional, recuperação semi-automática

Abstract

In the past few years, several studies based on constructional morphology to describe the mechanisms of construction of deverbal nouns, that is, nouns derived from verbs, using a synchronic perspective were carried out. The interest in this topic comes from the productivity of the construction mechanism of this type of noun. The impact in the Portuguese language, especially on its formal use, is not irrelevant. There aren't studies showing the variations of these mechanisms in Brazilian Portuguese (PB), much less, a diachronic perspective of him. We believe the current mechanisms be very similar with from XVI, XVII e VXIII. The suffix productivity and formation process can be suffering some kind of changing to specific words. The project Historical Dictionary of Brazilian Portuguese (DHPB) of centuries from XVI through XVIII, sponsored by the Programa Institutos do Milênio, and the construction of a corpus collected for this project, are, at the same time, a challenge and an opportunity to deepen the knowledge about this aspect of the Portuguese language. The analysis of this *corpus* gives us an opportunity to observe how these mechanisms evolved on the Portuguese language. However, there aren't yet tools that automate this type of research, enabling the morphologists to acquire this kind of data efficiently. The objective of this research was, first of all, to describe the different mechanisms of deverbal nouns construction in PB according to the SILEX morphological construction model (cf. Corbin 1987, 1991, 1997, Correia 1999, Rio-Torto (org.) 2004 and Rodrigues 2006). A second objective was the development of a computational system, named EXTRADEV, which allows easy access to the following data: (a) deverbal nouns (current and historic) with various morphological structures; and (b) graphical variants of nominal action (historic), to allow easier information retrieval. The methodology used on the construction of this system is grounded on: (i) description of deverbal nouns and computational rule's building about these; (ii) a pilot study of the fifty most frequent verbs of the DHPB project, extracted using UNITEX, and on the analysis of the graphical variations of these verbs; (iii) the knowledge of the Python programming language and regular expressions; and (iv) the use of the resources constructed for the DHPB project, such as a system of generation of graphical variants, SIACONF. We found 1,742,663 action deverbal instances and 15,633 distinct forms of the same without change of spelling. This number more variants extracted in the second module EXTRADEV, totalling 22,442 occurrences of deverbal history (6,809 variants and 15,633 deverbal no change in spelling). We followed some criteria analysis: frequency data, analysis about deverbal form based on in etymology and history dictionaries and the last criteria was based on the observation of the final list of historical deverbal. With this study, we aim to increase the knowledge of the diachronic variations of deverbal nouns and to motivate the joint linguistics and computer science contribution, particularly the one made by the natural language processing area, to empower future studies about the Portuguese language.

Key-words: deverbal nouns, historical *corpus*, constructional morphology, automatic recovery

Lista de Figuras

Figura 1: Arquitetura dos componentes atuantes no processo de construção de palavras	57
Figura 2: Arquitetura resumida dos componentes atuantes no processo de construção de palavras	59
Figura 3: Estrutura das classes gramaticais que funcionam de base para processos construcionais	64
Figura 4: Visão geral do sistema EXTRADEV	84
Figura 5: Arquitetura representativa da etapa 1 do módulo Deverbais de ação sem variação de grafia do sistema EXTRADEV	85
Figura 6: Arquitetura representativa da Etapa 2 do sistema EXTRADEV	86
Figura 7: Arquitetura da etapa 3 do módulo de extração de variação de grafia do sistema EXTRADEV	87
Figura 8: Arquitetura do módulo do sistema EXTRADEV de extração de deverbais históricos.	88
Figura 9: Exemplo de um código em linguagem Python.	93
Figura 10: Interface do prompt (também chamado linha de comando) com o comando grep.	94
Figura 11: Token List e a frequência das formas do verbo ser.....	102
Figura 12: Interface de busca e a janela dos resultados obtidos a partir da expressão <ser.V>	102
Figura 13: A forma ser como verbo	103
Figura 14: A forma ser como substantivo	103
Figura 15: Parte da Token List ordenada alfabeticamente.	106
Figura 16: Variações de grafia possíveis de bases para o verbo escrever	107
Figura 17: Interface do buscador de palavras do Word.....	121
Figura 18: Gráfico dos sufixos deverbais de ação mais frequentes no português.	128
Figura 19: Gráfico das estruturas de deverbais de ação mais frequentes construídas com o sufixo -ção	128
Figura 20: Gráfico das estruturas de deverbais de ação mais frequentes construídas com o sufixo -mente	129
Figura 21: Gráfico das estruturas de deverbais de ação mais frequentes construídas com o sufixo -mente	131
Figura 22: Gráfico das estruturas de deverbais de ação mais frequentes construídas com o sufixo -nç(o/a).....	133
Figura 23: Gráfico das estruturas de deverbais de ação mais frequentes construídas com o sufixo -gem.....	134
Figura 24: Gráfico das estruturas de deverbais de ação mais frequentes construídas com o sufixo -ncia	137

Lista de Tabelas

Tabela 1: Definições para deverbais encontradas em dicionários de LG e LE.....	24
Tabela 2: Tabela de sufixos formadores de deverbais encontrados na literatura.....	31
Tabela 3: Comparativo entre os trabalhos estudados.....	44
Tabela 4: Sufixos escolhidos para este trabalho.....	72
Tabela 5: Sufixos formadores de deverbais de origem latina.....	73
Tabela 6: Estrutura dos deverbais de ação possíveis a partir do verbo puxar.....	74
Tabela 7: Estrutura dos deverbais possíveis a partir do verbo pesquisar.....	74
Tabela 8: Estrutura dos deverbais de ação possíveis a partir do verbo parar.....	75
Tabela 9: Distribuição do corpus por século.....	89
Tabela 10: Metacaracteres para expressões regulares.....	93
Tabela 11: Variantes detectadas para as palavras “apelido”, “mais”, “não” e “vila”.....	97
Tabela 12: Agrupamento de “não” e “naõ” em torno da forma “nam”.....	98
Tabela 13: Regras do SIACONF para formas que caíram em desuso.....	99
Tabela 14: Regras SIACONF para consoantes duplas.....	99
Tabela 15: Regras do SIACONF geradas de acordo com a norma ortográfica.....	99
Tabela 16: Regras do SIACONF baseadas na frequência.....	100
Tabela 17: Regras do SIACONF baseadas em aprendizado automático.....	100
Tabela 18: Os 50 verbos mais frequentes no corpus total.....	103
Tabela 19: Formas nominais encontradas no corpus DHPB para o verbo escrever.....	106
Tabela 20: Formas nominais do verbo mandar.....	107
Tabela 21: Variações de grafia presentes no verbo mandar.....	107
Tabela 22: Variações de grafia presentes no verbo chegar.....	108
Tabela 23: Formas nominais do verbo partir.....	108
Tabela 24: Variações de grafia presentes no verbo partir.....	108
Tabela 25: Variações de grafia presentes na vogal temática da 3ª conjugação.....	108
Tabela 26: Formas nominais do verbos fazer.....	108
Tabela 27: Variações de grafia presentes no verbo fazer.....	109
Tabela 28: Variações de grafia presentes na vogal temática da 2ª conjugação.....	109
Tabela 29: Variações de grafia presentes nos sufixos formadores de deverbais de ação.....	109
Tabela 30: Regras de recuperação automática de deverbais de ação com variação de grafia.	114
Tabela 31: quadro comparativo do desempenho do buscador Word e do sistema EXTRADEV	122
Tabela 32: Parte da lista de deverbais históricos com e sem variação de grafia.....	125
Tabela 33: Possíveis deverbais de ação terminados com o sufixo (t/d)ura encontrados no corpus histórico.....	130
Tabela 34: deverbais construídos com o sufixo -nça/o encontrados no corpus histórico.....	132
Tabela 35: Deverbais de ação construídos com o sufixo -gem.....	134
Tabela 36: Deverbais de ação construídos com o sufixo -ata.....	135
Tabela 37: Deverbais de ação construídos com o sufixo -ncia.....	136
Tabela 38: Deverbais de ação construídos com o sufixo -são.....	137
Tabela 39: Deverbais de ação construídos com o sufixo -(u/a)me.....	137
Tabela 40: Deverbais de ação construídos com o sufixo -deira.....	138
Tabela 41: Deverbais de ação construídos com o sufixo -nsia.....	138
Tabela 42: Deverbais de ação construídos com o sufixo -dela.....	138
Tabela 43: Deverbais de ação construído com o sufixo -ria.....	140

Tabela 44: Deverbais de ação resultantes do processo de conversão.....	142
Tabela 45: Deverbais resultantes do processo de desflexionação de infinitivo	143
Tabela 46: Deverbais resultantes do processo de desflexionação de participio	144
Tabela 47: Deverbais de ação que sofreram alteração sufixal <i>-ção/ -mento</i>	146
Tabela 48: Deverbais de ação históricos que atualmente são construídos por processo de conversão.....	147
Tabela 49: Deverbais em com baixa ou nenhuma frequência no português contemporâneo.	148
Tabela 50: Deverbais de ação construídos por sufixação no português histórico.....	151

Sumário

Introdução	12
1. Deverbais: revisão da literatura	23
1.1. O que dizem os dicionários, gramáticas e trabalhos acadêmicos sobre os deverbais?	23
1.1.1. Dicionários	23
1.1.2. Gramáticas	26
1.1.2.1. Celso Cunha & Lindley Cintra (1989)	26
1.1.2.2. Bechara (1992 e 2005)	28
1.1.2.3. Rocha Lima (1957)	28
1.1.2.4. Gramáticas Históricas	29
1.1.2.4.1. Said Ali (1964)	29
1.1.2.5. Considerações	30
1.1.3. Trabalhos acadêmicos	31
1.1.3.1. Trabalho com corpus contemporâneo	32
1.1.3.1.1. Representação sintático-semântica do deverbais em Português (TOLEDO, 1976)	32
1.1.3.1.2. Formação de Palavras no Português Brasileiro (Sandmann, 1989)	34
1.1.3.1.3. Polissemia sistemática em substantivos deverbais (BASÍLIO, 2004)	35
1.1.3.1.4. Um modelo teórico de formação de palavras e sua aplicação aos deverbais do português (MARONEZE, 2005)	38
1.1.3.1.5. Os sufixos -ção e -mento na construção de nomes de ação e de processo: contribuições às práticas lexicográficas (SANTOS, 2006)	39
1.1.3.2. Trabalhos com <i>corpus</i> histórico	40
1.1.3.2.1. Deverbais em um texto português do século XIV: considerações sobre o étimo (REIS, 2005)	40
1.1.3.2.2. O estudo de deverbais no vocabulário da carta de Pêro Vaz de Caminha (OLIVEIRA, 2005)	41
1.1.3.2.3. Considerações sobre participípios e nomes deverbais no Foro Real (GOMES, 2005)	42
1.1.4. Deverbais no Portal da Língua Portuguesa	46
1.1.5. Considerações	47
2. O modelo SILEX	48
2.1. Histórico	48
2.2. Princípios de um modelo associativo e estratificado	49
2.2.1. Um modelo associativo	49
2.2.1.1. Distorções aparentes no léxico de palavras construídas: significado e forma	51
2.2.1.1.1. Distorções aparentes ao sistema derivacional	53
2.2.1.1.2. A concorrência aparente de muitos processos morfológicos	53
2.2.1.2. Uma definição não natural da noção palavra construída	54
2.2.2. Um modelo estratificado	56
2.2.2.1. O Componente de Base	57
2.2.2.2. O componente derivacional	58
2.3. Considerações	59
3. Deverbais de ação: denominação, significado, sintaxe e estrutura	62
3. 1. Denominação	62

3.2. O significado dos deverbais de ação	65
3.3. Comportamento sintático dos deverbais de ação	67
3.4. Estrutura	69
3.4.1. Sufixos formadores de deverbais de ação.....	71
3.5. Mecanismos de construção de deverbais de ação	75
3.6. A RCP de deverbais de ação.....	81
3.6.1. Considerações sobre RCP de deverbais de ação e de agente	82
4. EXTRADEV: um extrator de deverbais de ação históricos e suas variantes gráficas	83
4.1. O Projeto EXTRADEV	83
4.2. Seleção do Corpus	89
4.3. Sistemas e Linguagens.....	91
4.3.1. Linguagem PYTHON.....	91
4.3.1.1. Módulo RE	92
4.3.2. Comandos para processamento de texto.....	94
4.3.3. UNITEX	95
4.3.4. SIACONF	96
4.4. Estudo Piloto.....	100
4.4.1. Etapa I: Seleção dos 50 verbos	101
4.4.2. Etapa II: Extração das variantes gráficas de deverbais de ação	105
4.5. As regras	111
4.5.1. Os constituintes linguísticos x ambiguidade	111
4.6. Avaliações do EXTRADEV	115
4.6.1. Regras hipergerativas	116
5. Análise dos Resultados do Sistema EXTRADEV	120
5.1. Um comparativo entre o sistema EXTRADEV e buscadores de processadores de texto	120
5.2. Dados coletados	123
5.3. Desafios da contagem de palavra em corpus histórico.....	124
5.4. Processos construcionais mais recorrentes no português histórico	126
5.2.1 Sufixação.....	127
5.2.2. Conversão	142
5.2.3. Desflexionação	143
5.3. Deverbais de ação dos séculos XVI, XVII e XVIII à atualidade	146
6. Conclusão	153
6.1. Considerações iniciais	153
6.2. Sobre o modelo escolhido: SILEX	153
6.3. Sobre as regras linguísticas.....	154
6.4. Sobre o sistema de recuperação semi-automática – EXTRADEV.....	154
6.5. Sobre os dados gerados.....	155
6.6. Considerações finais.....	156
6.7. Trabalhos futuros.....	157
Referências	158
Apêndice A: regras no formato de expressões regulares para deverbais de ação com variação de grafia.....	163
Apêndice B: lista de deverbais de ação sufixados em <i>-mento</i> e <i>-ção</i>	178

Introdução

A criação de recursos computacionais para auxiliar a pesquisa linguística vem crescendo exponencialmente nos últimos anos. Embora os recursos para a língua inglesa encabeçam este crescimento, os recursos para a língua portuguesa também acompanham a evolução. Uma das razões que contribuem para esse desenvolvimento é a construção e disponibilização de grandes *corpora* em língua portuguesa. Dentre os *corpora* em língua portuguesa construídos no Brasil, podemos citar o Lacio-Web¹ (ALUÍSIO et al., 2003a, 2003b, 2004), o Corpus NILC² (PINHEIRO; ALUÍSIO, 2003), o NURC³ (1970), o PLN-BR⁴, o Corpus Summ-it⁵, o Banco do Português (BP)⁶, entre outros. Os *corpora* são construídos para diversas finalidades. Além de permitirem a descrição e compreensão dos fenômenos linguísticos, contribuem notavelmente para criação de recursos computacionais que embasam as várias áreas relacionadas ao Processamento de Língua Natural (PLN) como, por exemplo, a recuperação de informação, sumarização, tradução, desenvolvimento de corretores gramaticais, *parsers*, etiquetadores morfossintáticos, sistemas de síntese de fala (TTS), sistemas de reconhecimento automático de fala (RAF).

A grande maioria dos *corpora* em língua portuguesa é do português contemporâneo, concentrando, desse modo, as pesquisas sobre a língua no período de tempo corrente. Nesse contexto, encontramos pouca informação sobre a língua portuguesa de períodos passados baseada em grandes quantidades de dados.

No Brasil, uma das primeiras iniciativas de criação de *corpus* histórico realizou-se em 1998 na Faculdade de Letras da UFRJ com o projeto *Para uma História do Português Brasileiro* (PHPB)⁷ com o intuito de promover o conhecimento e a descrição da realidade linguística brasileira nos 500 anos de Brasil. Contemporâneo ao PHPB, é o *corpus* histórico Tycho Brahe⁸ (PAIXÃO and TRIPPEL, 2006), que foi desenvolvido junto ao projeto temático Padrões Rítmicos, Fixação de Parâmetros e Mudança Lingüística (Fase I, 1998-2003) e continuou sendo expandido na Fase II do projeto temático que cobre os anos 2005 a 2009. Esse projeto contempla o português europeu (PE) dos séculos XVI, XVII, XVIII e XIX -

¹<http://www.nilc.icmc.usp.br/lacioweb/index.htm>

² <http://www.nilc.icmc.usp.br/nilc/index.html>.

³ <http://www.letras.ufrj.br/nurc-rj/>

⁴ <http://www.nilc.icmc.usp.br/plnbr/>

⁵ <http://www.nilc.icmc.usp.br:8180/portal/index.jsp?option=related.jsp>

⁶ <http://www2.lael.pucsp.br/corpora/bp/index.htm>

⁷ <http://www.letras.ufrj.br/phpb-rj/>

⁸ <http://www.tycho.iel.unicamp.br/~tycho/corpus/index.html>

período esse que abarca, também, a formação do português brasileiro (PB) -, e suas análises realizam-se nos campos da prosódia, fonética e fonologia.

Há também o *corpus* BIT-PROHPOR (Banco Informatizado de Textos do Programa para a História da Língua Portuguesa) que é parte do projeto do Programa para a História da Língua Portuguesa⁹, desenvolvido por pesquisadores da Universidade Federal da Bahia.

Sobre ferramentas de PLN geradas a partir desses *corpora*, visando à dinamização e à ampliação das investigações, temos conhecimento do analisador morfossintático desenvolvido para o *corpus* Tycho Brahe (FINGER, 1998; FINGER, 2000) e, também, do *parser* estatístico inspirado no sistema do Penn-Helsinki Parsed Corpus of Middle English (PPCME) (FINGER, BRITTO and GALVES, 2000). A inexistência de ferramentas pode ser justificada pela pequena quantidade de *corpora* – como já foi explicitado – e pela complexidade do tratamento computacional: estado dos manuscritos, disponibilidade, qualidade dos documentos fontes, processo trabalhoso e minucioso de digitalização, quantidade de fenômenos peculiares que requerem codificações especiais, unidades lexicais de mesma significação e que apresentam alguma alteração gráfica, entre outros. Particularmente, quanto aos dois últimos problemas citados, RYDBERG-COX (2003) reforça que a ausência de hifenização, junções de palavras (exemplo: “éamor”), símbolos tipográficos incomuns para a grafia de palavras e a alta frequência de abreviaturas são comuns em *corpus* históricos de várias línguas.

Nesse cenário de ausência de grande quantidade de dados históricos digitalizados e, por conseguinte, escasso número de ferramentas computacionais para esse período do português do Brasil, surge o projeto intitulado Dicionário Histórico do Português Brasileiro (DHPB), preconizado por Maria Tereza Biderman (UNESP-Araraquara) e financiado pelo CNPq no âmbito dos Institutos do Milênio. O projeto visa à construção de um dicionário – inédito – do português do Brasil dos séculos XVI, XVII, XVIII e XIX. Para a concretização de tal objetivo, foi compilado um *corpus* de mais de sete milhões de palavras por meio do trabalho de uma equipe constituída por diversas universidades brasileiras e duas portuguesas. O *corpus* vem sendo utilizado para diversas pesquisas; uma delas – inserida na grande área de PLN – resultou no trabalho de mestrado de Arnaldo Candido Jr. (CANDIDO JR, 2008), entre outros trabalhos, recursos e ferramentas disponíveis no site HPC¹⁰ (CANDIDO JR, ALUÍSIO, 2008; VALE, et al, 2008; GIUSTI, et al, 2007).

⁹ <http://www.prohpor.ufba.br/projetos.html>

¹⁰ <http://www.nilc.icmc.usp.br/nilc/projects/hpc/>

A constituição desse *corpus* foi a nossa principal motivação, pois, com base nele, pudemos investigar o fenômeno linguístico eleito para esta pesquisa – os deverbais. Nos últimos anos, têm-se desenvolvido trabalhos em morfologia construcional¹¹ que visam à descrição dos mecanismos de construção de nomes deverbais, isto é, nomes derivados de verbos, numa perspectiva sincrônica, sobretudo em Portugal. As razões para esse interesse podem encontrar-se na produtividade dos mecanismos de construção deste tipo de nomes, cujo impacto no uso do português, sobretudo em contextos formais de uso, não é irrelevante. No Brasil, existem diversos trabalhos sobre deverbais, contudo, são raros os que se inserem nas áreas de Morfologia ou Lexicologia como veremos no decorrer deste trabalho (vide Seção 1.1.3). Faltam, portanto, estudos que nos permitam conhecer, por um lado, o Português do Brasil (PB) em nível destes mecanismos de construção e, por outro, ter uma perspectiva diacrônica deles. Nossa hipótese é que ao longo dos séculos os mecanismos sejam os mesmos. A mudança, se ela existir, pode estar relacionada à produtividade dos processos de construção de deverbais e de alguns sufixos em detrimento de outros; a existência ou participação de possíveis sufixos na construção de deverbais que, atualmente, caíram em desuso, ou seja, não pertencem mais ao paradigma de construção de deverbais de ação, ou participam da construção de outras categorias da linguagem.

Portanto, nos propusemos, neste trabalho, realizar uma descrição dos mecanismos de construção dos deverbais de ação, construir um sistema de extração semi-automática de deverbais de ação denominado EXTRADEV e analisar os dados encontrados no *corpus*. A construção do sistema só foi possível devido à formalização das regras linguísticas (no formato de expressões regulares¹²). Essas regras são capazes de reconhecer, automaticamente, estruturas de deverbais de ação. Além disso, este trabalho tem maior alcance por possibilitar que outras categorias – que seguem padrões construcionais semelhantes – sejam reconhecidas, através de uma proposta metodológica que aponta caminhos possíveis para recuperação de dados de maneira semi-automática. As regras podem ser reutilizadas para outras investigações, sendo necessário que o pesquisador altere os dados de entrada das regras de acordo com seu objetivo. O linguísta, conhecendo os constituintes, - sendo estes constituintes portadores de forma e significado previstos - e as operações

¹¹ *Morfologia construcional* é o termo adotado pelo modelo SILEX semelhante ao termo *morfologia derivacional* da GT, contudo o modelo SILEX aproxima-se cada vez mais de uma teoria do léxico.

¹² Uma expressão regular denota uma linguagem regular na Hierarquia de Chomsky (Hopcroft, Motwani & Ullman, 2001), isto é, é uma maneira formal de se especificar um padrão de texto, ou seja, uma composição de símbolos – caracteres com funções especiais – que agrupados entre si e com caracteres literais ou não formam uma seqüência, uma expressão.

linguísticas que formam as unidades lexicais, é capaz de reportar tais operações por meio de regras e realizar a equivalência usando a sintaxe das expressões regulares, por exemplo.

Os deverbais de ação, também chamados “deverbais eventivos”, “pós-verbais”, “*nomina actionis*” são substantivos formados a partir de verbos que podem expressar ação (ex.: do verbo (V) *combinar* temos o nome (N) *combinação*). Observemos o que diz Correia (2004, p.196):

Os substantivos deverbais, normalmente designados ‘nomes de ação’ ou ‘nomina action’ exprimem basicamente a ação de Vb.¹³, e são definidos lexicograficamente através de perífrases do tipo: <<ação de Vb>>, <<acto ou efeito de Vb>>. (...) É sabido também que muitos dos substantivos deverbais podem adquirir outros significados, entre os quais se contam os parafraseáveis por <<efeito de Vb>>, produto da acção de Vb, “agente da (acção) de Vb”, objecto da (acção) de Vb’, “instrumento da acção de Vb’, ‘local de acção de Vb’.

O excerto acima revela a polissemia dessa categoria e a existência de uma certa hierarquia entre esses sentidos. Ou seja, há sempre um significado central que nesse caso é a ação e significados periféricos que estão relacionados à ação. Por exemplo, o substantivo *casamento* é um deverbais do verbo *casar* que designa a ação de *casar* e pode ser parafraseado por ação de *casar*, contudo, parafraseá-lo por efeito ou resultado da ação de *casar* é mais preciso. Ex: *O casamento da Luciana foi lindo*.

Os constituintes dos deverbais de ação são o tema do verbo, ou seja, o radical (parte da palavra que contém o seu significado de natureza lexical) mais a vogal temática (indicativa da conjugação verbal a qual o verbo pertence). Em alguns casos, é possível obtermos bases verbais a partir do tema do participio passado, como, por exemplo, *escritamento*^o cuja base verbal é *escrita*- retirada da forma do verbo *escrever* no participio. Há deverbais também que são construídos juntamente com sufixos¹⁴.

Segundo o modelo adotado para este trabalho – SILEX (Syntaxe, Intreprétation, et Lexique) – os deverbais são palavras construídas definidas pelo modelo de análise como palavras portadoras de um significado previsível, linguístico, que é dado pela própria estrutura. Isso quer dizer que é possível prever o significado de uma dada palavra por ela mesma, independente do contexto no qual ela é empregada. Diferentemente das palavras de estrutura simples, que quando apresentam grafias iguais são dependentes do contexto no qual

¹³ ‘Vb’ quer dizer ‘Verbo base’.

¹⁴ Na derivação regressiva a questão do sufixo causa divergências entre estudiosos. Alguns pontos de vistas serão reportados no Capítulo 2.

estão inseridas para ocorrer desambiguação. Observemos a palavra *manga*. Se um estrangeiro nos perguntasse o significado dessa palavra responderíamos que em determinado contexto é x e em outro contexto é y. O que não ocorreria com a palavra *desenvolvimento*. Responderíamos “*ato ou efeito de desenvolver, ação de desenvolver*”. A base verbal carrega a ação do verbo que adjunta ao sufixo formador de nomes de ação costroi o sentido inerente à palavra.

A estrutura do deverbal de ação e de todas as unidades lexicais consideradas construídas é resultante da combinação de constituintes linguísticos já existentes na língua. Essa combinação dá-se de maneira associativa, ou seja, no momento da construção da unidade lexical, forma e significado se relacionam, além de poderem sofrer a ação de operações semânticas e fonéticas como veremos mais adiante (vide Seção 1.2).

É justamente essa autonomia da palavra construída proporcionada pela característica associativa do léxico que proporciona a construção de uma ferramenta automática com elevada taxa de acerto, sem a necessidade de recorrer a heurísticas sofisticadas que requeiram o plano sintático, que no caso das palavras simples seria necessário.

Descrevemos no Capítulo 3 os mecanismos de construção dos deverbais do português brasileiro histórico, através da regra de construção de palavra (RCP) responsável pela formação dos deverbais de ação - sobre quais bases essa categoria é construída, quais são os afixos participantes, quais são os processos de construção de deverbais de ação.

As chamadas RCP designam os paradigmas de construção de palavras existentes em determinada língua. É uma forma de mostrar que a construção de palavra também é passível de ser sistematizada, pois as palavras não se justapõem de maneira aleatória, pelo contrário, elementos semânticos, estruturais, fonológicos e pragmáticos interferem no processo.

Em suma, tradicionalmente definido como o conjunto de palavras de uma língua, o léxico é visto em diferentes teorias lingüísticas, explícita ou implicitamente como um conjunto arbitrário de palavras ou itens lexicais. No entanto, ao lado de signos arbitrários, as raízes ou bases primitivas, **o léxico apresenta também um alto teor de regularidade em seus processos de expansão, que constituem uma parte fundamental da organização linguística**¹⁵. (BASÍLIO, 2004 p.53)

Foram três as razões principais que nos motivaram a escolher os deverbais como nosso objeto de pesquisa:

¹⁵ Grifo meu.

a) por serem escassos os estudos acerca desta categoria no Brasil na área de Morfologia Construcional, haja vista a dificuldade encontrada na investigação da literatura (ver Seção 1.2);

b) por apresentarem um padrão de construção predizível e, por isso, altamente passível de automatização;

c) por serem unidades lexicais de alta produtividade¹⁶ no Português. “A formação de substantivos de ação deverbais é um dos processos mais produtivos de formação de palavras no português, por causa de sua motivação categorial de cunho gramatical” (BASÍLIO, 1980 KASTOVSKY, 1986 *apud* BASÍLIO, 2004 p.54).

Como já mencionado, os deverbais de ação serão descritos segundo o modelo SILEX desenvolvido na França na Universidade de LILE III, cuja criadora foi Danielle Corbin (1987). SILEX é um modelo de morfologia construcional que tem aspectos gerativistas por estar de acordo que todo falante tem uma competência linguística que é inata ao ser humano, o que autoriza qualquer falante a interpretar e produzir um número infinito de palavras mesmo que estas sejam desconhecidas. Contudo, distancia-se dos modelos gerativos por não priorizar os aspectos formais em relação aos aspectos semânticos. Acreditamos que este modelo é o mais adequado para nossa proposta por três razões principais:

a) por ser um modelo associativo que relaciona aspectos formais e semânticos (estrutura e significado);

Notre choix théorique fondamental est l’associativité : il repose sur l’idée que le sens d’un mot construit est construit en même temps que sa structure morphologique, et compositionnellement par rapport à celle-ci, et que la représentation grammaticale doit refléter cette construction simultanée de la structure et du sens (Corbin, 1991, p.294)

A associatividade repousa no fato de que cada parte gramatical da palavra deve refletir esta construção simultânea de estrutura e significado. Apesar de, como já dissemos, o modelo apresentar semelhanças com o modelo gerativo, é no carácter associativo (uma das características mais importantes do modelo ao lado do aspecto estratificado) que se apresenta uma considerável divergência do modelo gerativo, por ser ele dissociativo, ou seja, conceber

¹⁶ Defini-se produtividade como a capacidade de construir palavras a partir de processos regulares.

a estrutura como principal aspecto da palavra derivada e secundariamente seu significado que é dado por modelos interpretativos.

b) por ser um modelo estratificado:

Em vez de propor análises baseadas na evidência do léxico observável propõe uma análise baseada na estratificação e na reconstrução do léxico descritível; em vez de dar prioridade a análise morfológica sobre a análise semântica, propõe uma análise que associa forma e significado (cf. D. Corbin 1989a: 49 *apud* CORREIA, 2004, p.31)

A forma estratificada tenta representar, por um lado, a hierarquização entre as regularidades de diversos tipos, e situar as regularidades em seu adequado lugar, as sub-regularidades e as irregularidades; por outro lado, situar estes fenômenos em relação a outros componentes da gramática, e principalmente aos componentes semântico e fonológico. As diversas operações necessárias para construir a forma e significado de uma palavra são ajustadas às diferentes variáveis que intervêm, e sua ordem é tomada para reproduzir a hierarquia que vai do mais regular ao menos regular, o que encobre normalmente o intervalo que separa o previsível do observável.

c) o terceiro aspecto, e um dos mais importantes para a realização deste trabalho, é o fato de compreender a construção lexical como possibilidade. Isso quer dizer, o ser humano é capaz de gerar e compreender palavras mesmo que elas não sejam atestadas em dicionários. Na mesma direção está Basílio (2006, p. 10):

Do ponto de vista interno, ou mental, o léxico corresponde não apenas às palavras que um falante conhece, mas também ao conhecimento de padrões gerais de estruturação, que permitem a interpretação ou produção de novas formas. Assim o léxico interno é constituído por uma lista de formas já feitas e por um conjunto de padrões, os processos de formação de palavras, que determinam estruturas e funções tanto de formas já existentes quanto de formas ainda a serem construídas.

Por que os seres humanos falantes de língua portuguesa sem necessitar de uma gramática saberiam como expressar, por exemplo, o nome de um local onde se vende verdura: *verdura - verduraria*; onde se vende lanche: *lanche - lancheria*; *roupa - rouparia*; *tabaco-*

tabacaria; *bolsa* - *bolsaria*; será que todas essas palavras podem ser encontradas em dicionários?

Se, ao invés de designar um local, quiséssemos designar um golpe desferido com uma bolsa em alguém, diríamos: *Maria deu uma "bolsaria"¹⁷ em Pedro*? Não. Sem consultar um dicionário, diríamos: *Maria deu uma bolsada em Pedro*. Não diríamos **verdurada*, **lançamento*, **roupada*, **bolsadura*, por exemplo. Da mesma forma, diríamos: *João deu uma garrafada em Pedro*; e não: *João deu uma *garrafaria em Pedro*. “A proposta teórica prevê uma “competência derivacional”, revelada pela elaboração inconsciente de regras e não pela memorização de unidades lexicais isoladas (CORBIN, 87, p.53 *apud* CORREIA, 2004). Há um conjunto disponível de regras que diminui a dependência da memória e possibilita comunicação imediata.

Podemos observar uma diferença na construção das duas palavras (*bolsaria* e *bolsada*), e com mais atenção, notamos que a distinção está na terminação da unidade lexical. Para facilitar a compreensão, segmentamos as palavras, ou seja, dividimos os constituintes de formação das respectivas unidades. Observemos o exemplo abaixo:

<i>verdura</i>	→	<i>verdura-</i>	+ <i>-ria</i>
<i>banana</i>	→	<i>banana-</i>	+ <i>-ria</i>
<i>roupa</i>	→	<i>roupa-</i>	+ <i>-ria</i>
<i>bolsa</i>	→	<i>bolsa-</i>	+ <i>-ria</i>
<i>bolsa</i>	→	<i>bolsa-</i>	+ <i>-da</i>

A partir da segmentação podemos inferir que há regularidade formal na construção desses nomes. Tal regularidade formal não se dissocia do significado: há uma regra de construção de palavras que autoriza a construção de nomes de lugar a partir de bases substantivais mais o sufixo *-aria*. Os derivados denominais por meio do sufixo *-aria* assumem, entre outros, um significado locativo; já os substantivos que designam, entre outros, um golpe desferido com Nb não são palavras construídas e sim nomes desflexionados¹⁸ (nome da base – exs.: *joelhada*, *cotovelada*).

¹⁷ * e ° são símbolos utilizados para designar palavras agramaticais e palavras possíveis não atestadas, respectivamente.

¹⁸ A descrição dos processos de construção é apresentada na Seção 3.5.

O objetivo do modelo SILEX é construir uma teoria sincrônica do léxico capaz de atribuir uma estrutura e uma interpretação adequadas às palavras construídas, atestadas ou não nos dicionários, de caracterizar a natureza da <<gramaticalidade lexical>> e de determinar as restrições das regras de formação de palavras (CORBIN, 1987:1 apud CORREIA, 2004)

Dois aspectos do modelo SILEX são fundamentais para este trabalho:

- a) Associativo e Estratificado
- b) Hipergerativo: a possibilidade de lidar com um *corpus* histórico;

O caráter associativo permite maior precisão na busca pela categoria devido a sincronia entre forma e significado. Desse modo, o algoritmo conhece os constituintes que podem ou não se unir e evita construções aleatórias. O caráter estratificado hierarquiza e separa cada componente, facilitando também a organização do algoritmo de busca: o componente lexical 1 determina a entrada – as partes que se combinam; o componente 2 determina o ‘como combinar os constituintes’ por meio das RCP.

A partir do conhecimento da RCP de deverbais de ação pode-se gerar todas as estruturas morfológicas possíveis, o que não significa que elas sejam atestadas por dicionários e tampouco usadas na língua. Essa hipergeratividade é fundamental para recuperarmos formas deverbais possíveis dada a RCP, porém não necessariamente em uso no português atual. Se partíssemos do léxico observável de hoje, ou mesmo de gramáticas históricas, poderíamos perder diversas ocorrências no que tange a diversidade das construções morfológicas deverbais.

Além do desconhecimento das formas usadas na época, há também o fenômeno da variação de grafia o que permite aumento da variabilidade entre as formas deverbais. Em se tratando de um *corpus* histórico, é fundamental que tais formas sejam recuperadas.

O objetivo deste trabalho foi descrever os mecanismos de construção de deverbais de ação no português do Brasil e propor regras linguísticas de construção de deverbais de ação adequadas e adaptadas a um sistema de recuperação semi-automático. Para sua realização, seguimos a seguinte metodologia de trabalho:

- a) estudo das ferramentas e linguagens de programação adequadas à criação de regras que fossem robustas e fáceis de se usar por um pesquisador com formação na área de linguística;
- b) realização de um estudo piloto que abarcou os seguintes passos:

b.1) extração dos cinquenta verbos mais frequentes do *corpus* de trabalho por meio da ferramenta UNITEX (ver Seção 4.3.1 do Capítulo 4);

b.2) busca de uma amostra de variações de grafia de verbos na lista de desconhecidas – *Unknown List* – do mesmo software para detectarmos possíveis padrões na variações de grafia (ver Seção 4.3.2 do Capítulo 4);

b.3) construção de regras no formato de expressões regulares (ver Seção 4.3.3.2 do Capítulo 4) a partir dos estudos realizados previamente no Capítulo 1, as quais foram submetidas ao *corpus*. O algoritmo¹⁹ de recuperação de deverbais foi implementado na linguagem Python²⁰ (ver Seção 4.2.1 do Capítulo 4), utilizando a biblioteca de expressões regulares. Também foram utilizados outros comandos de processamento de texto (*Text Processing Commands*²¹) reconhecidos pelo sistema operacional Linux, que também suporta expressões regulares. É válido ressaltar que as expressões regulares são muito úteis para se lidar com uma língua natural, pois, além de muito flexíveis, elas possuem uma metalinguagem quase universal, com pequenas variações de linguagem para linguagem.

c) com o estudo piloto realizado, foi proposta uma arquitetura para o sistema e os procedimentos requeridos para a construção do sistema de extração de deverbais de ação denominado EXTRADEV;

d) as regras foram avaliadas manualmente segundo os seguintes aspectos:

d.1) todas as unidades lexicais da lista são possíveis deverbais de ação?

d.2) foram encontradas unidades lexicais agramaticais?

d.3) há palavras não construídas (ou não derivadas) na lista?

e) por fim, procedemos à análise dos dados para responder as seguintes questões:

e.1) quais são os mecanismos de construção mais recorrentes no português histórico?

e.2) quais são os sufixos mais frequentes? Além disso, fizemos uma análise comparativa em relação aos dados coletados do trabalho de Margarida Basílio (1996).

Esta dissertação é composta de seis Capítulos. O primeiro Capítulo traz a revisão da literatura - investigação sobre deverbais de ação em dicionários, gramáticas e trabalhos acadêmicos. O segundo Capítulo consiste na descrição do modelo de Morfologia

¹⁹ Vejamos algumas acepções para o termo algoritmo: um algoritmo é um procedimento computacional definido que recebe um ou mais valores (**entrada**) e produz um ou mais valores (**saída**) (<http://algoritmos.tiagomadeira.net/algoritmos-o-que-e-um-algoritmo>).

Um algoritmo não representa, necessariamente, um [programa de computador](http://www.python.org.br/), e sim os passos necessários para realizar uma tarefa (<http://pt.wikipedia.org/wiki/Algoritmo>)

²⁰ <http://www.python.org.br/>

²¹ <http://tldp.org/LDP/abs/html/textproc.html>

Construcional (SILEX) adotado para esta pesquisa de mestrado. No terceiro Capítulo, primeiramente, apresentamos uma proposta de denominação para a categoria de verbal, em seguida, realizamos a descrição dos verbais de ação segundo aspectos semânticos, estruturais e sintáticos e, sobretudo, descrevemos seus mecanismos de construção. No quarto Capítulo, após a revisão da literatura e a descrição do quadro teórico passamos à metodologia que abarca a descrição de um sistema para a extração semi-automática de verbais – chamado de EXTRADEV; o *corpus*; as linguagens e ferramentas utilizadas para os tratamentos computacionais; as regras; e um estudo piloto para os 50 verbos mais frequentes no português brasileiro. A última Seção desse Capítulo compreende a avaliação das regras do sistema EXTRADEV. O Capítulo cinco destina-se à análise dos dados e, finalmente, no Capítulo seis apresentamos a conclusão.

1. Deverbais: revisão da literatura

1.1. O que dizem os dicionários, gramáticas e trabalhos acadêmicos sobre os deverbais?

O principal objetivo deste Capítulo é a caracterização dos nomes deverbais de ação sob os aspectos morfológicos, semânticos e sintáticos. Antes de iniciarmos a descrição dos constituintes linguísticos que participam da construção dos nomes de ação, relataremos dados coletados de dicionários, gramáticas e trabalhos acadêmicos sobre o tema. Essa investigação objetivou conhecer como a literatura em geral compreende o verbal – como são apresentados, conceituados e qual lugar ocupam nos estudos de nossa língua. A explanação segue a seguinte ordem: Dicionários (Seção 1.1.1), Gramáticas (Seção 1.1.2), Trabalhos acadêmicos (Seção 1.1.3), e Deverbais no Portal da língua portuguesa (Seção 1.1.4).

1.1.1. Dicionários

Devido ao pouco material (verbetes) encontrado nos dicionários mais renomados – tanto nos de língua especializada quanto geral – ampliou-se a gama de dicionários a fim de obter maior quantidade de informação sobre deverbais. A Tabela 1 reporta os dicionários consultados de língua geral (LG) e de linguagens e especialidades (LE), com suas respectivas definições.

Tabela 1: Definições para deverbais encontradas em dicionários de LG e LE.

Dicionário	Definição
Dicionário Houaiss (LG)	<p>1 Diz-se de ou substantivo formado a partir de verbo (p.ex.: <i>constrangimento</i>, de <i>constranger</i>)</p> <p>2. Derivação: frequentemente. diz-se de ou palavra formada por <i>derivação regressiva</i>, a partir de um verbo; regressivo verbal, pós-verbal, regressivo (ex.: <i>embarque</i> < <i>embarcar</i>; <i>pesca</i> < <i>pescar</i>; <i>amparo</i> < <i>amparar</i>).</p>
Dicionário do Português (web) ²²	<p>Diz-se da palavra, ou a palavra derivada que resulta da "redução" da derivante, por isso que a esta se subtrai um segmento terminal, como, p. ex., choro(ô) (de chorar), toque (de tocar), compra (de comprar), sarampo (de sarampão). (A esse tipo de derivação também se chama regressiva, e, mais particularmente, deverbativo, em razão de ter grande produtividade na criação de substantivos tirados de verbos.</p>
Dicionário Aulete (LG)	<p>1 Diz-se de substantivo cuja origem é um verbo: 'Casamento' é um substantivo deverbal de 'casar'.</p> <p>2 Diz-se de substantivo formado de um verbo por derivação regressiva (p.ex.: corte, de cortar)</p> <p>3 Qualquer desses substantivos</p>
Dicionário de Uso do Português do Brasil (LG) ²³	-
Dicionário de Filologia e Gramática (LE)	<p>Segundo CÂMARA (1974 p.139), deverbais são nomes de ação, isto é, substantivos verbais abstratos (v.) que correspondem a verbos cognatos sem deles se derivarem por meio de um sufixo, ou em outros termos, derivados por um sufixo (0). Do ponto de vista diacrônico há três espécies de deverbais: derivados regressivos do verbo respectivo (ex: paga de pagar – ataque de atacar) onde se tem apenas o tema (v.); 2) derivados latinos em que na evolução fonética se obliterou o sufixo (ex: perda). 3) deverbais já existentes em latim (ex: fábrica fabrica de <i>fabricare</i>), havendo neste ultimo caso a possibilidade de em português de uma alternância prosódica em face de formas verbais (v.).</p>
Dicionário de Lingüística e Fonética (Crystal, 2000)(LE) ²⁴	-

²² <http://www.dicio.com.br/deverbal/>

²³ Não foi encontrada nenhuma entrada para deverbal.

²⁴ Não foi encontrada nenhuma entrada para deverbal.

Dicionário de lingüística (Dubois, et. al. 97/98, p.73)(LE)	Deverbais são os substantivos formados a partir de radicais verbais. Assim os substantivos <i>o gasto</i> (de <i>gastar</i>), a <i>marcha</i> (de <i>marchar</i>), etc., são deverbais. Às vezes, o termo deverbal é reservado aos substantivos derivados formados com o sufixo zero, como o fr. <i>bond</i> , “salto”, de <i>bondir</i> , “saltar”. (Sind. Pós-verbal.)
---	--

De maneira geral podemos observar que o termo deverbal está atrelado aos substantivos construídos a partir do processo de derivação de palavra intitulado “derivação regressiva”: todos os dicionários investigados, tanto os LE quanto os LG citam direta, ou, indiretamente, esse processo de formação de palavra. Todos enfatizam a estrutura de formação dos deverbais, salvo o dicionário de filologia e gramática que traz definições do ponto de vista semântico, apesar de restringir os deverbais a somente designadores de ação, deixando de lado outras possibilidades, como por exemplo, os nomes de qualidade. Devido à essa restrição semântica, acaba-se restringindo também, a classe gramatical do deverbal ao substantivo abstrato – que de fato são a maioria deles -, porém, por este ser polissêmico, assume também a forma de substantivo concreto: *O vendedor não sabe como resolver a questão* – agente da ação de Vb (*vender*) -, *Juliana dormiu no banco e não viu a parada de ônibus* – local da ação de Vb (*parar*). Temos, então, deverbais que são substantivos concretos que denotam, no primeiro, agente e, no segundo, local onde se para, cujos processos de formação são característicos do processo de construção de deverbais de ação – sufixação e desflexionação²⁵.

Dos sete dicionários consultados, apenas um não contém a entrada deverbal – Dicionário de Uso do Português do Brasil – o que pode ser explicado pelo fato de *deverbal* não ser um termo de conhecimento popular.

No Dicionário de Lingüística e no Houaiss, a definição inicia-se pelo aspecto de construção da palavra, revelando ser o radical verbal a origem dos deverbais “Deverbais são os substantivos formados a partir de radicais verbais”. Tal afirmação é mais ampla, levando-nos a pensar, em um primeiro momento, em qualquer processo de formação de palavra que tenha na sua base um radical verbal. Entretanto, em seguida, o Dicionário de Lingüística traz exemplos somente de derivados regressivos que são formados a partir da junção do sufixo 0 (zero) e faz uma ressalva sobre a relação direta do termo deverbal e a derivação regressiva. Por outro lado, o Houaiss, assim como o Aulete, são os únicos que, indiretamente, ou seja, a partir da exemplificação, nos dão a idéia de que deverbais também podem ser formados por

²⁵ Tanto a conversão quanto a desflexionação são concebidas segundo o modelo de morfologia construcional que veremos em seguida.

sufixos diferentes de zero: “Diz-se de substantivo cuja origem é um verbo: 'Casamento' é um substantivo deverbal de 'casar’”. O Aulete é um dicionário de origem portuguesa, talvez seja essa a razão de encontrarmos exemplos de deverbais por sufixação, pois os estudos acerca dos deverbais são mais explorados, ou recebem, aparentemente, mais atenção em Portugal.

De modo geral, as definições são focadas nos aspectos formais – caráter dissociativo – e mesmo assim de maneira breve e superficial. Deixa de lado outras categorias que também têm no verbo a sua origem como os adjetivos e os próprios substantivos que nomeiam agente ou qualidade e são formados por processos de sufixação e não de derivação regressiva. Verbos podem formar substantivos e adjetivos que exprimem diversos sentidos os quais estão, intrinsecamente, relacionados aos processos de construção que envolvem os deverbais e, por conseguinte, aos morfemas agregados a eles.

Após a explanação sobre os deverbais do ponto de vista dicionarístico, partiremos para análise das principais gramáticas.

1.1.2. Gramáticas

Em uma breve análise dos sumários das gramáticas constatamos que não existe Capítulo ou Seção destinada aos deverbais. Então, a primeira pergunta que surge é: onde devemos procurar pelos deverbais? Sendo eles nomes e, sobretudo, derivados, essas unidades lexicais devem estar alocadas na parte da gramática que trata o fenômeno da formação de palavras.

1.1.2.1. Celso Cunha & Lindley Cintra (1989)

No Capítulo 6 (p.83), cujo título é “Derivação e Composição”, encontramos a Seção sobre derivação sufixal e, em seguida, uma subseção sobre sufixos nominais. Nesta Seção, o autor lista sete tipos de derivação sufixal guiados pela classe gramatical da forma derivada e da derivante. Dentre elas a quinta intitula-se ‘formam substantivos de verbos’ e traz uma tabela com todos os sufixos possíveis, o sentido e o exemplo. A ordem de prioridade dos aspectos linguísticos dessa gramática é a seguinte: primeiro, o papel determinante dos sufixos na nominalização (Sufixos Nominais); em seguida, o efeito da junção do afixo à palavra base do ponto de vista categorial (categorias gramaticais); e, por último, o sentido conferido à

palavra derivada. Notamos, a partir dessa estrutura, a priorização do aspecto formal em relação ao significado.

Logo em seguida, há algumas notas de rodapé sobre os sufixos, a última delas faz uma consideração sobre os sufixos: *-ção* e *-são*: “Os sufixos *-são* e *-ção* depreendem-se de substantivos deverbais, quase todos formados no próprio latim (p. 89).

Os autores ao denominarem a tabela como “formam substantivos de verbos” focam somente o processo derivacional, sem mencionar o termo *deverbal*. O termo *deverbal* significa “*deverbal*” = de verbo – aquilo que vem do verbo; poderíamos trocar o nome da tabela por substantivos deverbais, ou simplesmente *deverbais*. Entretanto, eles escrevem algumas notas (conforme reportada acima) que nos induzem a enquadrar somente os sufixos *-são* e *-ção* como formadores de *deverbais*. Qual seria a distinção entre substantivos que vêm de verbos e *deverbais*? Notamos aqui a mesma inconsistência encontrada nos dicionários. Esse uso aparentemente aleatório do termo *deverbal* pode estar relacionado à carga semântica que é atribuída à palavra *deverbal* na maioria dos dicionários e gramáticas como pudemos observar na literatura relatada até agora. O excerto abaixo, encontrado na mesma gramática, é mais uma ilustração da hipótese que levantamos: “A derivação regressiva tem importância maior na criação dos substantivos *deverbais* ou pós-verbais, formados pela junção de uma das vogais *-o*, *-a*, ou *-e* ao radical do verbos” (p.102). Nessa passagem, o termo *deverbal* aparece não mais como nota e sim no corpo do texto, fazendo referência ao processo de derivação regressiva assim como encontrado nos dicionários.

A diferença na definição de Cunha & Cintra em relação às que encontramos até agora, está em como se constrói o *deverbal* pelo processo de derivação regressiva. Os autores não mencionam sufixo zero como nas definições dos dicionários, mas apenas que existe a junção das vogais. Na obra de Sandmann (1991), há uma reflexão sobre a questão do sufixo zero na derivação regressiva que reportaremos no Capítulo 3.

A partir desses dados podemos constatar que para Cunha & Cintra há *deverbais* formados por sufixação – quando sufixos *-são* e *-ção* – e no mais, por derivação regressiva, a qual parece ser a principal forma de construção de *deverbais* segundo os autores.

1.1.2.2. Bechara (1992 e 2005)

Na gramática de Evanildo Bechara encontramos na Seção sobre derivados, subseções cuja primeira intitula-se “I - Principais sufixos formadores de substantivos”, contendo oito tipos de formação. Tal divisão é baseada nos significados das construções: 1 - Para formação dos nomes de agente; e a segunda, a qual nos interessa: “2 - Para formação de nomes de ação, ou resultado de ação; estado, qualidade”. Essa subseção, por sua vez, apresenta três subdivisões: derivados de verbos; derivados de substantivos; derivados de adjetivos. Portanto, foi nessa Seção que encontramos os deverbais de ação. Vejamos em seguida os deverbais citados por Bechara (2005, p.358):

Derivados de verbos: *-ame*: gravame; *-ção, -são*: coroação, perdição, compreensão e ascensão; OBSERVAÇÃO: Há de se atentar para a correta grafia de *-ção* e *-são*. *-mento*: casamento, descobrimento; *-ura, -dura, -tura*: feitura, moderadura e formatura; *-ança (-ancia), -ença (-ncia)*: mudança, esperança, parecença, abundância, convalescença, (ou convalescência); *-ata*: passeata; *-ada*: estada (estadia na norma de Portugal); *-ida* (verbos de segunda e terceira conjugação): acolhida e partida; *-agem*: (vadiagem); *-ário*: lapidário

Seguimos a investigação esperando encontrar algo na Seção derivação regressiva, porém, o termo deverbal não é mencionado, há apenas a frase: “em regra isto se dá com muitos substantivos retirados de verbo (Bechara, 1992, p.185). Bechara em relação à Cunha & Cintra coloca o significado como anterior a categoria gramatical e faz considerações muito relevantes para o esclarecimento dos aspectos formais e semânticos que envolvem os processos derivacionais.

1.1.2.3. Rocha Lima (1957)

O Capítulo denominado *Formação de palavra* é iniciado com a definição de derivação e composição. Em seguida, o autor discorre sobre a derivação sufixal e afirma:

Ao contrário dos prefixos, que, como vimos, guardam certo sentido, com o qual modificam, de maneira mais ou menos clara, o sentido da palavra primitiva, os sufixos, vazios de significação, tem por finalidade formar séries de palavras da mesma categoria gramatical. (p.193)

Logo a seguir, apresenta uma lista de sufixos latinos, definindo cada um apenas em relação à categoria gramatical “forma substantivo de substantivo”. Sufixos como *-ada*, *-agem*, *-ão*, *-aria*, *-eria*, *-ato*, *-ado* aparecem todos como formadores de substantivos a partir de substantivos. Enfim, aparecem os primeiros sufixos que formam substantivos de verbos *-dor*, *-tor*, *-sor*, e *-tura*, *-sura*, *-dura*, *-mento* e *-tório*, *-douro*. Não há nenhuma menção a possíveis significados, o que era previsível, baseado na forma como os conceitua.

Na Seção sobre derivação regressiva Rocha Lima afirma que a derivação regressiva é um processo de redução do derivante e é altamente produtiva na criação de substantivos tirados de verbos, também chamado de derivação deverbal (p. 199). Na sequência cita Said Ali e sua divisão em quatro grupos de acordo com o gênero: masculino em -o; masculino em -e; feminino em -a; e masculino e feminino. Mais uma vez, a palavra deverbal aparece atrelada à derivação regressiva.

Rocha Lima faz uma descrição puramente formal o que restringe certos sufixos a uma única e exclusiva categoria, quando, na verdade, existem outras possibilidades de combinação, mas que somente poderiam ser recuperadas se o autor considerasse aspectos semânticos juntamente com os formais.

1.1.2.4. Gramáticas Históricas

A visitação às gramáticas históricas tornou-se relevante devido a esta pesquisa utilizar um *corpus* histórico, portanto, dados diacrônicos. Nesse sentido, consultá-las pôde contribuir com a análise dos dados e, além disso, com uma avaliação do conteúdo das mesmas.

1.1.2.4.1. Said Ali (1964)

Na *Gramática Histórica da Língua Portuguesa*, a palavra deverbal aparece somente vinculada ao tema derivação regressiva. Segundo Said Ali: “Aos substantivos assim derivados de verbos costumam os linguistas chamar deverbais ou pós-verbais, ou simplesmente substantivos verbais”. As escolhas lexicais na formação do enunciado deixam transparecer falta de clareza no que, de fato, abarca o termo deverbal - “(...) **costumam** os lingüistas chamar (...)”.

A gramática é organizada de modo diferente em relação às outras, apresentando micro seções para sufixos. Por isso, procuramos pelos sufixos característicos na formação de deverbais e encontramos afirmações como: “-*mento*: Ajunta-se este sufixo a temas verbais para constituir substantivos que denotam ação.”

Nessa definição, Said Ali consegue reunir tanto aspectos formais quanto semânticos, portanto, uma definição mais completa, embora somente esse sufixo dentre os formadores de deverbais apresentam tal definição. Já em relação aos sufixos -*são* e -*ção*: “Servem para derivar nomes abstratos dos verbos.”. Nessa definição o aspecto formal é priorizado, representando um lado mais funcional de geração de substantivos abstratos, ao invés de especificar o tipo de significação que esses abstratos assumem. Ambos os sufixos têm papel importante na construção de nomes de ação com bases verbais, entretanto são definidos diferentemente, não de forma incorreta, mas incompleta.

1.1.2.5. Considerações

Notamos que os autores estudados ora focam na estrutura lexical da palavra, ora no sentido, quando deveriam abarcar todos os níveis linguísticos que uma unidade lexical contém. Outro aspecto relevante é o fato de não existir consenso sobre as definições dos processos de formação, tampouco entre os sufixos formadores de deverbais nos gramáticos visitados. Desse modo, observamos um distanciamento da gramática em relação à realidade linguística, por não dar a devida atenção ao significado juntamente com os processos de formação de palavras e, principalmente, por não trazer explicações sobre a pluralidade, polissemia e produtividade das palavras construídas.

Observemos na Tabela 2 os sufixos para formação de nomes a partir de verbos que denotam ação ou resultado dela, os coincidentes estão em negrito.

Tabela 2: Tabela de sufixos formadores de deverbais encontrados na literatura.

Cunha & Cyntra		-ção -são	-mento	-(d)ura -(t)ura -(s)ura	-ança -ância -ença -ência	-tório	-douro		
Bechara	-ame	-ção -são	-mento	-ura -dura -tura	-ança -ancia	-ata	-ada -ida	-agem	-ário
Rocha Lima			-mento	-tura -sura -dura		-tório	-douro		

Tal definição sem critérios fixos de descrição acarretam falta de clareza e, por conseguinte, dificuldade de compreensão por parte do consulente.

Abaixo, segue a Seção que reporta a visão sobre deverbais do ponto de vista da academia.

1.1.3. Trabalhos acadêmicos

Foram encontrados poucos trabalhos sobre deverbais na língua portuguesa do Brasil em geral. Há alguns sobre o português histórico, porém português europeu (ver Seções 1.1.3.2.1, 1.1.3.2.2 e 1.1.3.2.3). Em relação ao português contemporâneo encontramos trabalhos sobre deverbais de ação nas áreas de lexicografia, lexicologia, sintaxe e em morfologia. São dois os trabalhos inseridos em morfologia: o primeiro (Sandmann, 1989) aborda questões de derivação passando pela categoria dos deverbais; o segundo (Maronezi, 2005) propõe um modelo para descrever o processo de formação dos deverbais numa perspectiva morfológico fonológica. Todos serão descritos em seguida.

Com vistas a uma melhor organização do trabalho faremos a seguinte divisão: trabalhos que se utilizam de *corpus* contemporâneo e trabalhos com *corpus* histórico.

1.1.3.1. Trabalho com corpus contemporâneo

1.1.3.1.1. Representação sintático-semântica do deverbal em Português (TOLEDO, 1976)

A autora inicia sua pesquisa pelos sintagmas nominais e delimita para aqueles somente transformados por nominalização afixal cuja base é um verbo. E assume em seu trabalho o princípio de que a nominalização é uma forma particular de derivação que produz substantivos a partir de substantivos, adjetivos e verbos, e é nesta última categoria que é realizada uma segunda delimitação de seu tema de investigação:

Substantivos criados a partir de verbo foram, portanto, nosso ponto de partida, uma vez que se evidenciou uma dificuldade de sistematização em torno do assunto, existindo apenas discussão entre os autores a respeito de seu estatuto, de sua origem, com divergências entre posições transformacionistas e lexicalistas, parecendo serem relevantes para esta falta de sistematização a escassez de material das investigações estruturalistas bem como o enfoque inadequado da gramática diacrônica e normativa. (p.7)

Ela segue suas delimitações no tema de pesquisa baseadas em critérios sintático-semânticos, justificadas pela aparição do agente como predicativo do deverbal, portanto deverbais de agente (*-dor -ante*), por apresentarem no próprio substantivo a agentividade, são excluídos. Nesse sentido exclui também deverbais de lugar e de resultado de ação. “o lavatório das mãos...” (p.8). “não gostei /desta decoração do apartamento/.”. E, em seguida, chega, enfim, ao seu recorte de pesquisa, o que ela designa como deverbais de atividade e justifica que somente esses deverbais apresentam uma relação predicativa entre os actantes de uma predicação sintático-semântica. Escolhido o objeto – deverbais de atividade – uma descrição da categoria é iniciada, nesse sentido Toledo (1976, p.12) faz uma observação pertinente sobre a classificação dos deverbais em relação aos seus processos de formação:

Recebem o nome de deverbais só as lexias substantivais resultantes da derivação por supressão. Modernamente, chamam-se deverbais os nomes formados a partir de radicais verbais, por sufixação: --> a recusa sufixo o; a adoção -->sufixo -ção; diverge pois da divisão tradicional que reserva o nome “deverbal” só aos derivados formados com sufixos zero.

Esse excerto pode trazer uma explicação sobre o fato de deverbais estarem sempre atrelados àqueles construídos por derivação regressiva (como pudemos verificar na maioria dos dicionários e gramáticas consultados). Segundo a autora, é por volta de 1976 que passa a se reconhecer nomes provindos de verbos formados por sufixação na categoria deverbal: “Modernamente, chamam-se deverbais os nomes formados a partir de radicais verbais, por sufixação:”. Embora no meio acadêmico tenha se iniciado a aceitação de nomes derivados de verbos por processo de sufixação, sabemos que os dicionários e as gramáticas tradicionais são bastante conservadores e demoram a incorporar certas mudanças linguísticas.

Dentre as possibilidades de alocação de um deverbal no enunciado, Toledo escolhe somente os deverbais que são núcleos de sujeitos ou de complementos verbais. A seleção dos deverbais para análise foi baseada na tese de John Clifton Duncan (DUNCAN, 1971), que apresentou as palavras mais frequentes do português, através de um tratamento estatístico. Segundo Duncan, os substantivos deverbais ocorrem mais frequentemente em gêneros textuais que requerem uma linguagem mais especializada, técnica e formal. Portanto, são comumente encontrados em obras técnicas e também em jornais, que exigem um tipo de linguagem voltada para a transmissão de informação, e apresentam baixa ocorrência em romances. O autor reuniu textos dos mais diversos gêneros textuais: obras científicas, ensaios, crônicas, romance, teatro e jornalismo, num total de 34 obras, abrangendo trinta autores no período de 20 anos (1955 – 1975).

A autora considera os sufixos zero, *-ção*, *-mento*, *-agem*, *-ura*, *-ência*, *-ança* como formadores de deverbais e as análises são realizadas segundo a teoria gerativa transformacional, no nível – sintático-semântico. Ela conclui:

As construções nominais com deverbais, inseridos em enunciados, funcionam como sujeito ou complemento. E, ocupando essas posições, exercem papéis sêmicos. Assim, seja como sujeito ou complemento funcionam semanticamente como: instrumental, Causativo, Objetivo e Resultativo. No entanto, exercem funções de Meta, Origem, Direcional, Propósito e Material somente na posição de complemento e de Estativo somente como sujeito”. (p.359)

Com a realização da sua pesquisa, além de alcançar o objetivo esperado, a autora também proporcionou o levantamento de temas relacionados aos deverbais ainda não muito explorados e que se mostram bastante problemáticos.

A ambigüidade, fenômeno muito comum nas nominalizações com deverbais, devido às supressões do discurso, é um assunto bastante rico (...) e ainda entre outros, podemos citar o problema da lexicalização envolvendo deverbais, isto é, o Nv acompanhando determinados verbos e formando com eles expressões lexicalizadas. (p.360)

Não há nenhuma análise estatística sobre produtividade de processos e sufixos para que pudéssemos realizar um comparativo em relação aos dados encontrados no nosso *corpus* histórico.

1.1.3.1.2. Formação de Palavras no Português Brasileiro (Sandmann, 1989)

Essa obra consiste na pesquisa e na descrição dos modelos de formação de palavra produtivos ou mais produtivos do português brasileiro, e os deverbais estão entre os mais produtivos.

Para sua pesquisa, Sandmann utilizou um *corpus* de 42 edições dos seguintes jornais do ano de 1984 - *O Globo*, *Jornal do Brasil*, *O Estado de São Paulo*. Como critério de seleção de palavras foi estabelecido que uma palavra só integraria o *corpus* se não estivesse registrada no novo dicionário Aurélio.

Os deverbais de ação encontrados são terminados em: *-ada*, *-ção*, *-dura*, *-mento*, *-dor* e *-deira*. Ele faz uma separação baseada no sentido que cada um ajuda a construir (1989, p. 50-51):

No que se refere ao conteúdo das derivações formadas com esses sufixos merece destaque o fato de que as primeiras são substantivos designativos de ação, enquanto as em *-dor* e *-deira* são substantivos designativos de agente, para o que com certeza contribui a base constituída de verbo.

O trabalho de Sandmann contribui para o conhecimento da formação de novas unidades lexicais em português brasileiro e, por isso, abarca todos os processos de formação possíveis para a língua portuguesa, enquanto esta monografia faz um recorte e se aprofunda somente na descrição dos deverbais de ação e seus processos de construção com vistas a criar um recurso computacional.

1.1.3.1.3. Polissemia sistemática em substantivos deverbais (BASÍLIO, 2004)

O trabalho de Basílio aborda a questão da polissemia sincrônica na construção lexical, focando a construção e expansão do léxico. Observemos:

Processos de formação de palavras apresentam sobretudo duas funções: a função gramatical e a função denotativa. A função gramatical corresponde à necessidade de utilizar palavras de uma classe gramatical em contextos lingüísticos que exigem uma palavra de outra classe. (p.52)

Ou seja, essa necessidade gramatical está intrinsecamente relacionada às escolhas lexicais do enunciado, que muitas vezes requerem mudanças de classes, entretanto, mudanças gramaticais que preservem o sentido. Podemos dizer:

O presidente está satisfeito porque a economia cresce.

Ou, em outro contexto:

O presidente está satisfeito com o crescimento da economia.

Na segunda frase não caberia uma forma verbal e sim um nome que designasse a ação de crescer. As mudanças de classes gramaticais são constantes no nosso léxico diário ao realizarmos os enunciados. Por exemplo, podemos usar o adjetivo *belo* para qualificar algo ou alguém: *A menina é bela*, porém, se a necessidade for nomear o *belo*: *A beleza do quadro*; *Era uma beleza eterna* - passamos de um adjetivo a um substantivo. Essa alteração se dá por meio de uma mudança estrutural, mais especificamente sufixal requerida pelo contexto lingüístico.

Já sobre a função denotativa, a autora afirma que “corresponde à necessidade de aproveitamento do material simbólico já existente numa palavra para formar outra palavra, que designe algo diferente, mas relacionado à palavra primitiva” (p.53). Por exemplo, das palavras que designam objetos como *sapato*, *malha*, *camisa*, *relógio*, *bicicleta*, podemos formar *sapataria*, *malharia*, *camisaria*, *relojoaria*, *bicicletaria*. Todas são palavras que designam lugar onde se compra ou se conserta os objetos supracitados.

Observemos o que ela diz sobre a parte gramatical do deverbais de ação:

Nomes de ação deverbais apresentam sobretudo função gramatical, isto é, são formados com o objetivo de utilização da noção verbal em contextos sintáticos que exigem um substantivo. Nesta função, virtualmente todos os verbos apresentam um substantivo deverbais correspondente. Mas substantivos deverbais também podem ter motivação denotativa, ou seja, também podem ser formados com o objetivo de denotar seres, processos, eventos, situações, etc. a partir da noção verbal. (p.54)

Na oração *Os servidores estão em greve*, o deverbais é denominado de agente, por formar um nome que designa um ser que pratica alguma ação. Não se perde o sentido de ação, entretanto, altera-se toda a construção sintática, pois agora o deverbais não apenas substitui uma estrutura verbal que pede um nome, mas sim nomeia um agente, que pode e ocupa posições distintas na sentença.

Em seguida a autora divide a obra em seções, a primeira intitula-se: Nomes de ação com função gramatical. Nesta Seção atenta para o uso de deverbais em sentenças que requerem uma expressão verbal, mas que a estrutura sintática da mesma pede um substantivo.

A próxima Seção intitula-se “Polissemia sistemática em nomes de ação” e assim ela a define:

Chamamos de polissemia sistemática a multiplicidade de interpretações possíveis de caráter pré-determinado forma lingüística. (...) No caso das nominalizações deverbais, a polissemia sistemática corresponde a diferentes instâncias de uso freqüente e necessário, todas relacionadas a um processo geral de desverbalização do verbo. (p.56)

Ela assume que cada deverbais tem diversas interpretações previamente sabidas de acordo com a representação conceitual e estrutura sintática requerida. Por isso segue uma polissemia sistemática.

Em seguida a Seção “Nomes de ação com função denotativa” considera que os nomes de ação também têm função denotativa no sentido de nomear entidades, eventos e fenômenos a partir do significado do verbo. Ela acrescenta que esses deverbiais são importantes na constituição de terminologias acadêmicas, técnicas e profissionais, embora também ocorram no vocabulário cotidiano.

Vejamos os exemplos que a autora utiliza para explicar tal ponto de vista:

(8) a. coincidência – direção – coordenação – luta – ocorrência – desvio

b. derivação – nominalização – referência – transferência – depressão

Em (8a), formas nominalizadas representam conceitualmente fatos, situações, condições, modos, processos, cargos, etc., informados pelo significado do verbo. Assim, por exemplo, coincidência designa a situação de uma coisa coincidir com outra, luta é um evento esportivo em que dois adversários lutam, coordenação é o conjunto articulado de atividades ou providências requeridas quando se coordena algo, e assim por diante. Do mesmo modo, em (8b) temos um elemento básico do significado da forma no significado do verbo derivante; mas, tratando-se de linguagem técnica, há sempre especificidades que se situam fora da esfera lingüística e que dependem das circunstâncias do ato de nomeação (...) transferência pode ter um significado psicanalítico ou de alocação profissional; depressões podem ser psiquiátricas, geográficas ou econômicas; e assim por diante.

A autora diz que interpretações específicas não podem ser preditas, entretanto, o aspecto verbal daquilo que está sendo nominalizado permanece. E, logo após, conclui que o processo também apresenta polissemia sistemática, embora o elemento semanticamente pré-determinado seja apenas parcial.

Em suma, em ambos os casos temos direções claras em relação à interpretação polissêmica, embora as particularidades de cada instância sejam oriundas de informação extra-lingüística e, portanto, de caráter lingüisticamente imprevisível. (p.59)

Assim como Basílio, este trabalho defende que o deverbal tem natureza polissêmica e que sentido e significado são predizíveis (mesmo que parcialmente) à medida que conhecemos sua RCP.

1.1.3.1.4. Um modelo teórico de formação de palavras e sua aplicação aos deverbais do português (MARONEZE, 2005)

Esse trabalho contribui para os estudos sobre o processo de formação de palavras na língua portuguesa. É uma alternativa aos modelos já estabelecidos que apresentam lacunas, em decorrência da complexidade que é o estudo lexical. O autor sugere um modelo de formação de palavra que se opõe ao modelo de Aronoff²⁶ que, segundo ele, tem recebido diversas críticas, entre elas, o tratamento da formação de palavras baseado em regras semelhantes às regras sintáticas.

O modelo seguido pelo autor é o proposto por Bybee (1988) denominado Rede de conexões lexicais e é pautado basicamente na semântica e na fonologia. Para Byber: “O léxico é organizado como uma rede de semelhanças: as unidades lexicais apresentam uma série de semelhanças nos dois planos: o semântico e o fonológico” (p.2).

O modelo é aplicado aos deverbais, relacionando as partes equivalentes do verbo e do nome e seu sufixo. Quando a correspondência não é imediata como em palavras em que a nominalização não é regular *atuar* → *atuação* e *conectar* → *conexão*, no modelo de Bybee isso é tratado como uma correlação com uma palavra irregular ou semi-regular. Outro fenômeno que também é contemplado pelo modelo é o chamado de falso sufixo. Observemos o que Maronezi discorre sobre este fenômeno:

O modelo de Byber possibilita ainda outro tipo de análise, que pode ser chamada de “falso sufixo”, exemplificada com a lexia *adição*:

a t u a ç ã o
| | | |
a d i ç ã o

A semelhança fônica entre as lexias *atuação* e *adição* é grande; porém, enquanto *atuação* é um derivado sufixal “legítimo” a partir do verbo *atuar*, o mesmo não pode ser dito em relação à *adição*. Poder-se-ia concluir, assim,

²⁶ <http://www.linguistics.stonybrook.edu/faculty/mark.aronoff>

que se trata de um caso de semelhança fônica sem semelhança semântica correspondente, a exemplo de conceber, perceber e receber, anteriormente mencionados. Porém não parece ser esse o caso. A lexia adição apresenta semelhanças inegáveis com outras lexias formadas com o sufixo -ção: é um substantivo abstrato, pode figurar em construções com verbos-suporte (como efetuar uma adição) e apresenta o traço semântico de “ação”. Acrescente-se a isso o fato de possuir um verbo semanticamente relacionado (embora morfologicamente derivado), que é o verbo adicionar. Parece, portanto, razoável analisar o -ção de adição como sendo o mesmo sufixo nominalizador de atuação, por exemplo. Casos semelhantes seriam confecção, seleção, tradição, refeição (com o sufixo -ção) e aumento, alimento, elemento, experimento (com o sufixo -mento) (p.6).

O modelo não deixa muito claro quais aspectos são de fato relevantes para o estabelecimento das conexões. Talvez a sistematização do modelo - considerando e elegendo os traços principais, quais os aspectos fonológicos e semânticos que são priorizados e os dispensáveis, e qual o peso da morfologia e da fonologia nesse modelo - ajude os estudiosos a compreenderem melhor a viabilidade e a sua amplitude.

Tentativas como essa de abarcar aspectos semânticos e fonológicos da formação lexical evidenciam cada vez mais que para uma análise lexical suficientemente completa é necessário recorrer a distintos níveis de descrição linguística, apesar desse modelo focar apenas em dois.

1.1.3.1.5. Os sufixos -ção e -mento na construção de nomes de ação e de processo: contribuições às práticas lexicográficas (SANTOS, 2006)

Nesse trabalho, a autora com sua pesquisa visa melhorar a prática lexicográfica, no português, dos verbetes afixais, por estar ciente da necessidade de embasamento linguístico na elaboração dos verbetes dessa categoria. Dentre os afixos existentes, ela foca sua análise nos sufixos -ção e -mento, que embora sejam sufixos concorrentes, assumem diferentes papéis na língua. Ambos caracterizam-se por serem formadores de deverbais, entretanto, também, podem ser simples terminações. Tal análise é realizada a partir do modelo de morfologia construcional (CORBIN, 1987), aplicado a um *corpus* de 1.225 palavras recolhidas no Novo Dicionário Aurélio (2004) e no Dicionário Eletrônico Houaiss da Língua Portuguesa (2004).

Após a investigação da literatura, a autora constata que ambos os sufixos são aplicados às bases verbais, formam substantivos abstratos, associam-se a um sentido múltiplo, pertencem a um conjunto de sufixos nominalizadores e dentre esse conjunto são os mais

produtivos; e a escolha de um ou de outro é considerada aleatória pelos gramáticos e linguístas. Porém, em alguns casos, há preferências em decorrência de restrições morfológicas/fonéticas.

Reportadas as principais características dos sufixos escolhidos, a autora segue seu trabalho descrevendo os sufixos a partir do modelo da morfologia construcional que é usado nesta dissertação também. E, a partir de tal descrição, sugere novas formas de elaborar verbetes de maneira a contemplar todos os aspectos linguísticos que envolvem os sufixos.

1.1.3.2. Trabalhos com *corpus* histórico

O PROHPOR (Programa para a História da Língua Portuguesa) destina-se ao estudo da língua portuguesa num período que vai desde suas origens até meados do século XVI (período arcaico). Nesse âmbito está inserido o Projeto DEPARC (Dicionário Etimológico do Português Arcaico), coordenado pelo professor doutor Américo Venâncio Lopes Machado Filho. O DEPARC tem por objetivo principal a elaboração de um Dicionário Etimológico do Português Arcaico, compreendido entre os séculos XIII e meados do XVI. O Projeto não se restringe, contudo, a oferecer, como resultado, apenas um dicionário etimológico alfabético impresso, mas disponibilizará, através do suporte da informática (em que se sustenta o sistema de informação, denominado PRODEPARC, especialmente desenvolvido para a sua elaboração), diversas formas de consulta e pesquisa.

Concomitantemente ao desenvolvimento do dicionário surgiram outras pesquisas motivadas por ele. Merecem destaque o projeto de Iniciação Científica denominado Deverbais em um texto do português do século XIV: considerações sobre o étimo (REIS, 2005) e Considerações sobre participios e nomes deverbais no Foro Real (GOMES, 2000). Abaixo discorreremos sobre o conteúdo de cada um.

1.1.3.2.1. Deverbais em um texto português do século XIV: considerações sobre o étimo (REIS, 2005)

Esse trabalho, a partir do levantamento de deverbais e seus respectivos étimos em um texto do século XIV, discute quais os processos de constituição morfológica desses itens no âmbito da derivação regressiva e sugere uma proposta de classificação para os deverbais opondo-se à classificação tradicional dada pela Gramática Normativa.

O trabalho inicia-se com uma investigação acerca das definições encontradas em gramáticas e dicionários. As obras citadas são: Dicionário de Lingüística e Gramática (CÂMARA, 2000); Dicionário Gramatical Globo (1962); Dicionário Houaiss da língua portuguesa (HOUAISS, 2004); Novo Dicionário da Língua Portuguesa (Buarque de Holanda, 1986), entre outros. Após sua análise dicionarística, a autora constata que: “O que nos mostram os dicionários consultados é que a dúvida quanto à melhor proposta de classificação de palavras deverbais ainda existe” (p.4).

Essa incoerência de definições reforça o objetivo da pesquisa em sugerir uma nova proposta de classificação. A nova proposta consiste na alteração do tema - os deverbais que apresentam vogal distinta em relação aos verbos de origem, ou seja, deverbais formados por derivação regressiva cuja vogal que se junta ao radical não coincide com a vogal temática do verbo em questão (*sonhar - sonho*); e os que a mantêm (*pesquisar - pesquisa*). Ao final, sugere que sejam considerados nomes deverbais apenas aqueles que permaneceram com a vogal temática do verbo de origem e apenas retiraram sua marca verbal para tornarem-se nomes.

O trabalho contempla apenas aspectos morfológicos e restringe os deverbais à somente àqueles formados por derivação regressiva.

1.1.3.2.2. O estudo de deverbais no vocabulário da carta de Pêro Vaz de Caminha (OLIVEIRA, 2005)

O trabalho inicia-se alertando para as inconsistências sobre a categoria deverbais e sugere o levantamento dos problemas, reflexões e possíveis soluções. O *corpus* de trabalho é a carta de Pêro Vaz de Caminha, no qual a autora pretende recuperar todas as formas deverbais para uma análise diacrônica. Abaixo a autora define seu objetivo e expõe os desafios:

(...) pode-se afirmar que serão abordados aspectos principalmente referentes ao uso das denominações “deverbal” e “derivação regressiva”, em relação às quais se verificam questionamentos, nem sempre esclarecidos satisfatoriamente e que se evidenciam como matéria-prima para constantes debates e discussões (p.4).

Como a proposta é um estudo diacrônico, torna-se necessária a busca em dicionários

etimológicos:

Torna-se árdua a tarefa de identificar um “autêntico” deverbal, na medida em que não existe uma regularidade no emprego da referida terminologia. Seja na prática dos dicionários, ou na teoria de lingüistas e filólogos, encontram-se não poucas imprecisões ou incertezas no uso e conceituação do termo. Na busca por esclarecimento em relação ao termo deverbal e, por extensão, à derivação regressiva, dada a vinculação semântica existente entre ambos, necessitou-se consultar, mais atentamente, não só dicionários como o Houaiss, os etimológicos de José Pedro Machado e de A. G. Cunha, bem como aqueles produzidos por lingüistas, filólogos e gramáticos (p.6).

Da mesma maneira, como foi constatado neste trabalho, a autora confirma a dificuldade em definir e delimitar um deverbal e afirma, após recorrer a várias fontes especializadas:

Nesse contexto, está, conseqüentemente, vetando a possibilidade da formação de um deverbal através do acréscimo de uma partícula sufixal ao verbo “derivador”. Reafirmando tal postura, no mesmo dicionário, trata esse autor, ao falar de derivação, da denominação prefixal e sufixal, mas em nenhum momento se refere a essa segunda como sendo “deverbalizante” (p.8).

Apesar de não encontrar na literatura uma referência convincente e sistemática, ela propõe uma nova classificação que considera mais ampla em relação às definições encontradas. Deverbais, então, segundo Oliveira abarca os seguintes processos de formação de palavra: o que chama de derivação progressiva – formam substantivos de verbos a partir de afixos – que, por sua vez, inclui a derivação sufixal, prefixal e parassintética. Também inclui a derivação regressiva e o que denomina derivação por participio.

Conclui o trabalho com muitas ressalvas e deixa em aberto possíveis sugestões.

1.1.3.2.3. Considerações sobre participios e nomes deverbais no Foro Real (GOMES, 2005)

Esse trabalho é uma tentativa de se estabelecerem classes de palavras a partir de traços distintivos nas estruturas deverbais retiradas do Glossário do “Foro Real – Afonso X”, parte

integrante da edição de José de Azevedo Ferreira de 1987, com um foco específico nas estruturas formadas por participios. A análise dos dados foi feita segundo a *Gramática Descritiva do Português* (Perini, 1997), e da *Gramática da Língua Portuguesa* de Mira Mateus *et al.* (2003).

Esta pesquisa foca, sobretudo, as estruturas nominais formadas a partir de participios, propondo uma distinção entre substantivos, adjetivos e participios, ou seja, como se dá a passagem de um elemento com traços verbais para um elemento com traços nominais nesse texto. A distinção entre os tipos de participios é feita através do contexto sintático.

Esse trabalho é semelhante ao anterior no que tange ao problema de classificação, entretanto, aborda outros tipos de deverbais sob e, sobretudo, a análise é realizado sob o ponto de vista sintático.

Após reportar verbetes, conceitos gramaticais e trabalhos acadêmicos, procederemos a uma análise comparativa com os principais trabalhos acadêmicos citados. Tal análise objetiva a constatação das contribuições e das lacunas em relação ao que já foi realizado na área em questão e, nesse contexto, justificar e apresentar nossas contribuições com esta pesquisa. A Tabela 3 ilustra os trabalhos estudados até agora com o intuito de realizar uma comparação. As pesquisas estão organizadas de acordo com as letras abaixo:

- A- Representação sintático-semântica do verbal em português
- B- Formações de Palavras no Português Brasileiro
- C- Polissemia sistemática em substantivos deverbais
- D- Um modelo teórico de formação de palavras e sua aplicação aos deverbais do português
- E- Os sufixos *-ção* e *-mento* na construção de nomes de ação e de processo: contribuições às práticas lexicográficas
- F- Deverbais em um texto português do século XIV: Considerações sobre o étimo
- G- O estudo de deverbais na carta de Pero Vaz de Caminha
- H- Considerações sobre participios e nomes deverbais no Foro Real
- I- Este trabalho

Os seguintes campos aparecem na Tabela 3: uso de *corpus*; tipo de *corpus*; tamanho do *corpus*; uso de recursos computacionais; período de análise; dados atestados por dicionários; tema do trabalho; e abordagem teórica.

Tabela 3: Comparativo entre os trabalhos estudados

	Uso de <i>corpus</i>	Tipo do <i>corpus</i>	Tamanho do <i>corpus</i>	Uso de recursos comp.	Período de análise	Atestação de dicionário	Tema do trabalho	Abordagem
A	Sim	contemporâneo	34 obras	Não	1955-1975	Sim	Deverbiais de ação -	Sintático - semântica
B	Sim	contemporâneo	42 jornais	Não	1988	Não	Formação de palavras	Morfologia
C	Não	-	-	-	-	-	Deverbiais de ação	Semântica
D	-	-	-	-	-	-	Deverbiais	Fonologia e Semântica
E	Sim	contemporâneo	-	Não		Sim	Deverbiais construídos por sufixação -mento e -ção	Lexicografia
F	Sim	Histórico	1 texto	Não	XIV	Sim	Deverbiais	Morfologia e Etimologia
G	Sim	Histórico	1 texto	Não	XVI	Sim	Deverbiais	Morfologia e Etimologia
H	Sim	Histórico	1 texto	Não		Sim	Deverbiais	Morfologia e Etimologia
I	Sim	Histórico	7 milhões de palavras	Sim	XVI, XVII, XVIII e parte do XIX	Indiferente	Deverbiais de ação	Morfologia/ Lexicologia

Para uma breve análise dos dados da Tabela 3, elegemos apenas os trabalhos que utilizam *corpora* e abarcam todos os processos de formação de deverbais segundo a perspectiva da morfologia derivacional. Portanto, somente A, B e I.

Todos se utilizam de *corpus*, contudo apresentam metodologias diferentes para a compilação do *corpus* e extração dos deverbais. Tanto A quanto B não têm o apoio de ferramentas computacionais para tais etapas, por conseguinte, o gasto temporal para se obter os dados - atualmente aspectos como tempo de compilação, de organização e de busca dos dados são determinantes na viabilidade de realização de qualquer pesquisa – é muito grande.

A e B realizam suas análises lexicais sincronicamente, ou seja, analisam o sistema linguístico em funcionamento num determinado momento, sem a perspectiva histórica. I, por sua vez, apesar de partir de regras disponíveis no léxico atual utilizando um modelo sincrônico, por se tratar de um *corpus* histórico, os dados a serem analisados são diacrônicos.

Em relação ao tema, B reporta todos os processos de formação de palavras do português que envolvem os novos vocábulos, por isso acaba tratando as formações deverbais, todavia, de maneira superficial, já que o intuito do trabalho é dar conta de todas as unidades lexicais novas e não explorar somente a categoria verbal. Em A, apesar da autora tratar, exclusivamente, dos deverbais de ação, a análise é sob a perspectiva sintático-semântica segundo a teoria transformacionalista. E, finalmente I – este trabalho – foca a construção dos deverbais de ação segundo o modelo SILEX que atualmente pode ser considerado quase uma teoria geral do léxico, por pretender dar conta da estrutura e do funcionamento do léxico, incluindo aspectos formais e semânticos.

Sobre a atestação por dicionário, para B é condição fundamental o fato da palavra não estar dicionarizada. Tal procedimento justifica-se pelo fato de o objetivo ser justamente descobrir palavras novas. Em I o fato de uma palavra ser atestada ou não por dicionários é indiferente, pois o que interessa é a produtividade dos mecanismos de construção dos deverbais, pautada na competência dos falantes que é anterior a qualquer intervenção social.²⁷

²⁷ Somos capazes de interpretar e produzir muitas construções lexicais sem saber se elas estão ou não incorporadas no léxico, devido a um conjunto de regras internalizadas. Contudo, na língua em uso, muitas dessas unidades lexicais nem chegam a existir, pois através do princípio da economia lexical, se já temos uma palavra que circula na língua para expressar determinado fenômeno, não necessitamos de outra com a mesma função (por exemplo, se já utilizamos a palavra *aprimoramento* por que deveríamos utilizar também a palavra *aprimoragem*? Este é um verbal possível, mas não necessariamente em uso). Os aspectos sociais que influenciam para que algumas palavras sejam incorporadas e outras não, são inúmeros, mas não nos compete neste trabalho analisá-los.

1.1.4. Deverbais no Portal da Língua Portuguesa

Esta Seção faz uma breve referência a um site de estudos da língua portuguesa cujos responsáveis têm a mesma concepção sobre deverbais que é apresentada neste trabalho.

Enquanto no Brasil os estudos sobre deverbais são recentes e estão em desenvolvimento, em Portugal as pesquisas acerca dos nomes de ação está avançada, haja vista a existência de um dicionário de deverbais on-line disponível para qualquer usuário.

O dicionário está alocado no Portal da Língua Portuguesa (<http://dev.portaldalinguaportuguesa.org/index.php?action=main>) que, além do dicionário, disponibiliza outros recursos como MorDebe (uma base de dados morfológica), Dicionário de Gentílicos e Topônimos e Dicionário de Estrangeirismos de uma forma imediata, acessível e simples.

Esse dicionário de nomes deverbais fornece uma lista dos nomes deverbais da língua portuguesa. Os nomes deverbais listados nesse dicionário são os nomes deverbais registrados na MorDebe, a qual tem mais de 13 mil verbos dos quais apenas 30% têm um nome deverbal registrado. A justificativa para essa pequena porcentagem funda-se nas seguintes razões fundamentais:

Quando o verbo é derivado do nome, nem sempre faz sentido o uso de um nome deverbal dele derivado: o verbo *cedilhar* significa "pôr uma cedilha numa letra" e pelo menos o efeito de *cedilhar* é simplesmente uma cedilha. O próprio significado de alguns verbos implica que só muito dificilmente possam ser utilizados num contexto nominal, como acontece com *verter*. Muitas vezes o verbo já é tão pouco frequente que a dicionarização da forma nominal não se justifica, como parece suceder com *encorpamento*. Por último, o nome deverbal não é dicionarizado por ser completamente previsível, ou por mero lapso do lexicógrafo²⁸.

Nesse dicionário *on-line* só são encontradas as formas deverbais atestadas por dicionários, daí uma quantidade reduzida de nomes de ação. Há uma grande diferença nas possibilidades de construção de palavras em uma língua e aquelas reconhecidas pelos dicionários.

²⁸ Em conversa com Margarita Correia

1.1.5. Considerações

Com essa breve pesquisa feita acerca dos deverbais de ação no Brasil, verificamos que o objetivo deste trabalho é inédito, tanto em relação à teoria utilizada e ao *corpus*, quanto ao período investigado. Além disso, é o único trabalho que propõe a construção de uma ferramenta de auxílio à pesquisa e investigação.

Após a contextualização da área de pesquisa e a exposição do local que ocupamos nesse cenário, consideramos importante definir a nomenclatura adotada neste trabalho para esta categoria, a qual apresenta variação não só no que tange à conceitualização, mas também à denominação. Por isso, a próxima Seção é uma tentativa de nomear os deverbais visando maior transparência terminológica.

2. O modelo SILEX

2.1. Histórico

O modelo SILEX surgiu no final da década de 80, período este em que os estudos sobre morfologia, sobretudo morfologia derivacional, despontaram na França. Havia a necessidade da criação de uma gramática das palavras construídas: regularidades, esclarecimento de regras, princípios e exceções.

Inicialmente, o modelo SILEX foi considerado um modelo de morfologia derivacional, ramo da morfologia que trata da formação de palavras. Após alguns anos de pesquisa, o modelo foi se aprimorando e se assumindo como um modelo de morfologia construcional, por não lidar apenas com a construção de palavras por derivação, mas também por outros mecanismos linguísticos como a desflexionação e a lexicalização. Na sua versão inicial o SILEX apresentava-se como um modelo de descrição das palavras derivadas, evidenciando questões formais relacionadas às mesmas. Era considerado um modelo bastante rígido por ter como principais aspectos a obrigatoriedade da unicidade categorial e a obrigatoriedade da unicidade semântica das regras de formação de palavra. Nesse sentido, o modelo não permitia o tratamento da polissemia, sendo consideradas formas homônimas. Tais características revelaram-se pouco adequadas devido a um tratamento muito específico, não alcançando uma visão macro capaz de provar as regularidades das unidades lexicais.

No início da década de 90, precisamente em 1991, apesar de ser ainda um modelo de morfologia derivacional, passou a lidar com questões de significado das palavras construídas, permitindo o tratamento da polissemia. Isso só foi possível devido à consideração de três aspectos intervenientes na construção do significado de uma unidade derivada:

- O significado dado pela regra de formação de palavra;
- O significado herdado da base;
- O significado específico do operador morfológico envolvido;

o que permitiu melhor entendimento da maneira como se estratificam os diferentes significados previsíveis que uma mesma unidade derivada pode apresentar .

Na terceira e mais recente versão, o modelo, enfim, assume-se como um modelo de morfologia construcional e não apenas derivacional, permitindo o tratamento da semântica e

da referência das palavras construídas morfológica ou não morfológicamente construídas. Pode-se dizer que este é um modelo que se aproxima cada vez mais de uma teoria geral do léxico.

2.2. Princípios de um modelo associativo e estratificado

2.2.1. Um modelo associativo

O modelo SILEX aproxima-se da teoria gerativa por ser um modelo que visa à descrição da competência dos falantes de uma língua e também por conceber que as unidades lexicais existem desde que existam regras que permitam construí-las a partir dos constituintes linguísticos existentes em dada língua. Contudo, distancia-se da teoria gerativa por este ser um modelo associativo – o significado é construído ao mesmo tempo que sua estrutura morfológica. Ao passo que a teoria gerativa dá mais atenção à forma do que ao significado das palavras construídas.

Nossa escolha teórica fundamental é a associatividade: ela repousa sobre a idéia de que o significado de uma palavra construída é constituído ao mesmo tempo em que sua estrutura morfológica, e de modo composicional em relação a esta, e que a representação gramatical deve refletir esta construção simultânea da estrutura e do significado. Esta posição opõe-se radicalmente à concepção dissociativa predominante na gramática gerativa (cf. n.2), segundo a qual, em paralelo aos tratamentos ortodoxos dos enunciados gerados sintaticamente, a estrutura está em primeiro lugar, e o significado é afetado nas estruturas por modelos interpretativos. (CORBIN, 1991 p.9)

A escolha da associatividade entre as relações de forma e significado visa à redução de aparentes distorções, que na maioria dos casos, são consideradas como naturais. Essa visão ‘naturalista’ configura um tratamento atomístico do léxico e implica na impossibilidade de uma sistematização do mesmo.

Segundo o modelo, a maioria das distorções entre forma e significado são aparentes e, portanto, redutíveis. Entretanto, a observação superficial das palavras complexas leva-nos a crer que forma e estrutura nem sempre apresentam uma relação harmônica. Muitos dos teóricos que propõem a dissociação baseiam-se em dois aspectos: palavras que são aparentemente formadas da mesma maneira e não parecem manter a mesma relação semântica com seus constituintes; e palavras que têm o mesmo significado, mas são morfológicamente

diferentes.

Para exemplificar o primeiro caso podemos citar a palavra *mesinha*. É uma palavra construída por meio da base *mesa* (substantivo) adjunta ao sufixo *-inha* (indicador de diminutivo) que designa *pequena mesa*. Por outro lado, a palavra *carrinho* construída sobre a base *carro* adjunta ao sufixo *-inho*, apesar de ser formada da mesma maneira e com o mesmo constituinte sufixal, pode, além de significar um carro pequeno, significar um brinquedo e uma jogada no futebol. Observemos as frases abaixo:

- a) O antigo Ford Ka era um *carrinho*, o novo está mais espaçoso.
- b) João pediu ao tio que lhe desse um *carrinho* de ferro de natal.
- c) O *carrinho* foi por trás e, por isso, o jogador foi expulso.

Em c), *carrinho* denomina uma categoria referencial diversa que, aparentemente, não tem relação nenhuma com *carro* (sm. veículo de transporte terrestre).

No segundo caso, no qual o significado permanece, porém a morfologia é distinta, podemos citar palavras como *vela* (palavra simples), *barco a vela* (palavra formada de maneira composicional) e *veleiro* (palavra construída), as quais designam barco a vela. Ambos os casos demonstram aparentes distorções entre a forma e o significado, contudo segundo Corbin:

Os sistemas morfológicos e os significados que eles ajudam a construir aparecem em disposição cruzada: de uma parte, a uma mesma forma afixal parecem corresponder diversas estruturas e diversos significados (...); de outra parte, a um mesmo significado parece corresponder vários processos morfológicos. (...) e parece mais confortável de adotar um procedimento realista, aquele que dissocia a descrição das estruturas da do sentido, que esforçar-se por pôr ordem naquilo que parece heteróclito. (1991, p.10)

A autora propõe que antes de realizar uma análise não superficial do léxico é preciso livrar-se de falsas evidências observáveis trazidas pela gramática. Nesse sentido, o modelo compromete-se a responder adequadamente aos principais argumentos empíricos expostos pelos modelos concorrentes a favor da dissociatividade.

2.2.1.1. Distorções aparentes no léxico de palavras construídas: significado e forma

O modelo defende que a grande maioria das distorções são redutíveis, desde que seja realizada uma análise precisa e hierarquizada do significado, ou uma análise mais abstrata da estrutura.

Quando se fala de uma distorção entre a forma e o significado de uma palavra construída, é possível medir o desvio em relação ao seu significado derivacionalmente previsível. Isto pode ser descrito como o resultado da ação semântica conjugada à regra de construção de palavra (RCP) utilizada, do processo morfológico escolhido e da base (CORBIN, 1991 p.11). Por exemplo, o significado derivacionalmente previsível de *carrinho* (movimento utilizado no futebol para tirar a bola dos pés do jogador do time adversário) não é imediatamente dedutível do significado de *carro* e do valor diminutivo que se atribui geralmente ao sufixo *-inho(a)* *pequeno carro*, pois operações semânticas podem intervir em diversas etapas no processo de formação da palavra construída, tornando o significado derivacionalmente previsível opaco. Assim, a parte metafórica do significado de *carrinho*, nesse caso, está relacionada à função do carro e (o jogador realiza um movimento semelhante ao de um carro ao se chocar com uma pessoa) pode ser descrita como resultado da aplicação de uma regra semântica sobre a base *carro* da palavra construída, previamente a sufixação de *-inho*. As operações semânticas que se aplicam ao processo construcional de caráter metafórico, geralmente estão relacionadas à forma ou à função do objeto a que se referem, no caso de *carrinho*, à função.

Toda palavra, simples ou construída possui uma referência extralingüística, o que permite que um mesmo significado previsível seja suscetível de se comportar diferentemente em vários domínios referenciais, criando assim um efeito de heterogeneidade superficial. Nesse sentido, confirmamos que as idiosincrasias são aparentes e o que ocorre no léxico são adaptações pragmáticas do significado previsível.

Em relação à forma observável de uma palavra construída, assim como o significado, não é o resultado de uma simples concatenação de elementos, mas da combinação de operações complexas, isto quer dizer que a forma representada pela estrutura morfológica pode estar disfarçada, transformada pela ação de dispositivos pós-derivacionais como a alomorfia, o truncamento e a integração paradigmática. Vejamos, abaixo, um pouco sobre esses dispositivos:

1) A alomorfa, explicada de maneira simplificada, é a mudança de uma forma para outra sem alteração de sentido na palavra. É, também, uma maneira sincrônica de lidar com um conjunto de fenômenos historicamente heterogêneos (empréstimos, evolução fonética), mas que se pode reduzir, sincronicamente, a uma certa regularidade. Por exemplo, o prefixo *in-* que se une a palavras para negar seu significado *infeliz*, *indubitável*. Contudo, também, encontramos *imoral*, *ilegal*. Ambos prefixos *i-* e *in-* têm o mesmo sentido e constituem o paradigma de construção de adjetivos de negação do significado. O que houve foi que o prefixo *in-* sofreu uma alomorfa. Esse processo de mutação constante na língua ocorre por conta de uma das suas principais características: o dinamismo da língua.

2) Truncamento é a ausência superficial de segmentos de uma palavra cuja existência é predita segundo a sua estrutura morfológica. Por exemplo, um truncamento fonológico acontece com a palavra *analista*, a qual é um deverbal de agente:

Analisar → base (*analys-*) + (*-ista*) = *analista** → operação de truncamento → *analista*.

O truncamento pode ser do tipo morfológico:

Vírus → base (*virus*) + (*-al*) = *virusal** → operação de truncamento → *viral*.

3) A integração paradigmática permite explicar a presença superficial de segmentos afixados não necessários à interpretação semântica. Assim, na análise de exemplos como *antigrippal*²⁹:

Antigripal → *anti* + *gripe* = *antigripe* → operação de integração paradigmática → *antigripal*. A terminação *-al* de *antigripal* não funciona como um verdadeiro sufixo, mas como uma marca de integração da palavra na categoria adjetival.

Os casos de distorções aparentes entre a forma e o significado das palavras formadas são, portanto, todos redutíveis, caso se trate de palavras formadas, com a condição de que se veja a palavra formada não como uma simples concatenação formal e semântica de elementos, mas como o resultado da combinação de operações hierarquizadas de natureza derivacional, semântica e fonológica. Os argumentos que justificam o dissociativismo pelas distorções empíricas observáveis não são, portanto, válidos. (Corbin p 15 1991)

²⁹ *Antigrippal*: *antigripal*

2.2.1.1.1. Distorções aparentes ao sistema derivacional

Outro ponto tomado para justificar o dissociativismo relaciona-se ao sistema derivacional – o mesmo processo de construção de palavra pode dar origem a palavras semanticamente diferentes, por exemplo, o sufixo *-(t/d)ura*³⁰ que pode construir tanto substantivos de qualidade (*doçura*) quanto substantivos de ação (*feitura*). Nesse caso, não se pode observar apenas um elemento do processo de constituição da palavra. O significado atribuído à palavra construída está intrinsecamente relacionado à base a qual um sufixo une-se.

Corbin também se posiciona sobre a questão do tratamento não unificado dos significados diferentes, atribuindo mais uma vez, à uma análise superficial.

(...) a análise, fundada exclusivamente sobre a observação da forma superficial das palavras, despreza a história derivacional destas: uma mesma forma pode realmente encobrir homônimos estruturais, pois ela representa o resultado da aplicação seja de operações derivacionais diferentes(...)seja de mesmas operações em uma ordem diferente (Corbin, 1991 p. 15).

Como exemplo, podemos citar batida “ato ou efeito de bater” ou batida “bebida a base de frutas e leite condensado”.

2.2.1.1.2. A concorrência aparente de muitos processos morfológicos

Os substantivos de ação, os substantivos de agente, os adjetivos de relação, os verbos de mudança, todas essas categorias podem ser construídas por mais de um processo morfológico. Segundo Corbin, geralmente, processos semanticamente concorrentes atuam sobre bases distintas. Por exemplo, o substantivo *desenvolvimento* é um deverbal de ação formado por sufixação; *revolta* e *desvio* também são deverbais de ação, porém formados por conversão³¹. Não se encontra atestado em dicionários o substantivo *desenvolva*^o ou *desenvolve*^o – que seria a forma de um possível deverbal construído por conversão. Também não se encontram o deverbal *revoltamento*^o, ou *revoltação*^o ou *desviamento*^o e *desviação*^o construído por sufixação.

Contudo, nos deparamos com palavras como *enterro* e *enterramento*. Ambas são

³⁰ Sufixo latino herdado do tema verbal do particípio passado.

³¹ Os processos morfológicos de construção de deverbais de ação serão descritos no próximo Capítulo.

formadas pelo mesmo processo morfológico e apresentam o mesmo significado derivacionalmente previsível pela estrutura. No entanto, *enterramento* apesar de existir na língua é muito pouco usado. Nesse caso, podemos dizer que uma forma de mesmo sentido sobrepõe a outra. Caracterizando um descarte natural da língua quando há de fato a concorrência semântica entre duas unidades lexicais.

No modelo SILEX cada tipo derivacional corresponde a uma regra de construção de palavras, que define o significado fundamental comum a todas as palavras formadas pela mesma regra (SPcr: significado previsível construído pela regra), e os processos morfológicos associáveis a este tipo derivacional, reagrupados em um paradigma morfológico associado à RCP. Porém, cada processo morfológico tem suas propriedades e é submetido a contrações específicas que formam o que no derivado será o significado previsível específico no processo morfológico (SPspm), que é suscetível de ocorrer de modos diferentes em função do significado da base (SPhb: significado predizível herdado da base).

Dessa maneira percebemos a presença de diversas variáveis na construção do significado de uma dada palavra construída. A descrição desse processo construcional do significado dá-se de maneira estratificada e sempre está associada à estrutura morfológica. Nessa perspectiva, deve-se levar em conta a especificidade, na maioria das vezes omitida, destas palavras: sua característica construída.

2.2.1.2. Uma definição não natural da noção palavra construída

Muitas das aparentes distorções lexicais podem ser esclarecidas através da compreensão do que constitui uma palavra construída: além de apresentar uma estrutura complexa, ser predizível à medida que estrutura morfológica e significado são construídos conjuntamente e, por isso, ser a área de aplicação das RCPs, uma palavra é analisada como construída se satisfizer as três condições seguintes:

1) Os constituintes de uma palavra construída devem ser todos categorizáveis e associados a um significado de maneira reprodutível. Por exemplo, na palavra *predizer* podemos isolar o prefixo *pre*³² e a base *-dizer*, mas não podemos isolar prefixo e base em

³² Prefixo que traduz as idéias de <<antes, diante de; antecedência; antecipação; superioridade; intensidade>>; do latim *prae-*, do advérbio *prae*, <<antes, à frente de, diante>>: preceder, predispor (...) (Machado, 1967 p.1868)

*precoce*³³. O prefixo *pre-* – existente na língua portuguesa, de origem latina – designa algo anterior, porém *-coce* não é uma base conhecida em português; não aparece nem como forma autônoma, nem associada a outro morfema; e não se pode a ele atribuir nenhum significado ou categoria fixos. Portanto, *precoce* é uma palavra complexa estrutural e semanticamente, mas não é uma palavra construída devido a um de seus constituintes não ser categorizável e interpretável no português.

Além das palavras complexas não construídas, existem as palavras simples que não apresentam estrutura interna compósita, apesar de apresentarem, muitas vezes formas homógrafas que podem confundir o analista. É o caso da palavra *cimento*. A forma *-mento* é um dos possíveis sufixos formadores de deverbais de ação no português. No entanto, em *cimento*, a terminação *-mento* não é um sufixo, e sim, uma terminação que coincide ortograficamente com o sufixo. A palavra *cimento* é de origem latina *caementu* que significava “*pedra de alvenaria*”³⁴. Portanto, não é uma palavra construída em português. Se quiséssemos analisar a suposta base (caso fosse um deverbais) da palavra simples *cimento*, teríamos: a suposta base *ci-* a qual não aparece nem como forma autônoma, nem associada a outro morfema *ci-* *-ção* (a união a outro sufixo não constitui uma unidade lexical reconhecida no português) ou *des*³⁵- *-ci* (a união a um prefixo não constitui uma unidade lexical reconhecida no português); e não pode a ele ser atribuído nenhum significado ou categoria fixos. Então, constatamos que *ci-* não é uma base e *-mento* não é um sufixo do português. Esta é, então, uma palavra de estrutura simples cujo significado é “Material de propriedades aglomerantes, usado em construções e trabalhos de alvenaria, que misturado à água forma uma pasta de consistência plástica e endurece muito à medida que seca”³⁶.

2) Seu significado previsível e sua estrutura morfológica devem ser composicionais, um em relação ao outro. Seu significado previsível e sua abstrusa estrutura morfológica devem ser composicionais, um em relação ao outro. (...) Além disso, ela permite eliminar do conjunto de palavras formadas, palavras como *peuplier*, complexa estruturalmente, mas não semanticamente (CORBIN, p. 17, 1991).

3) As eventuais distorções entre sua estrutura morfológica abstrusa e sua forma

³³ *Adj.* Do latim *praecoce-*, <<precoce (em relação a frutos e plantas); *fig.*, que chega antes de tempo; apressado, prematuro, precoce >>. A acentuação parece revelar que se trata de palavra entrada indiretamente talvez pelo fr. *Precoce*. Séc. XVIII, segundo Morais (Dicionário da língua portuguesa, 1813)

³⁴ Daí o significado de “alicerse”. No lat. tardio já aparece no sentido de argamassa, pelo hábito de os pedreiros romanos juntarem lascas de mármore à massa. (Nascentes, 1966 p.176)

³⁵ Prefixo indicativo das idéias de <<negação, contrário, oposição, separação, afastamento, divisão, supressão, >>. (...) (Machado, 1967 p. 786).

³⁶ http://aulete.uol.com.br/site.php?mdl=aulete_digital&op=loadVerbete&pesquisa=1&palavra=cimento

aparente ou entre seu significado previsível e seu significado lexicalizado devem ser explicáveis por mecanismos regulares permitindo transitar de um a outro. Esta condição permite assinalar como excepcionais as palavras cujas propriedades observáveis (exceto aquelas que pertencem à adaptação pragmática do significado linguístico previsível, (cf. §2.2.1.1) apresentam em relação às propriedades composicionalmente previsíveis correspondentes das anomalias (as ‘idiossincrasias’), não pertencendo aos mecanismos regulares que são as regras semânticas, operando em diversos níveis da linhagem derivacional, ou as regras formais fonológicas ou pós-derivacionais (alomorfia, truncação, integração paradigmática). (CORBIN p.17, 1991)

2.2.2. Um modelo estratificado

O objetivo do modelo SILEX é representar o conjunto de fenômenos de ordem derivacional; tal representação tenta, por um lado, hierarquizar as regularidades de diversos tipos, e situar as regularidades em seu adequado lugar, as sub-regularidades e as irregularidades; além disso, também, situar estes fenômenos em relação a outros componentes da gramática, e principalmente aos componentes semântico e fonológico.

Ser estratificado implica em que as diversas operações necessárias para construir a forma e significado de uma palavra são ajustadas às diferentes variáveis que intervêm, e sua ordem é tida para reproduzir a hierarquia que vai do mais regular ao menos regular, o que encobre normalmente o intervalo que separa o previsível do observável. As relações entre o componente derivacional e o componente semântico concernem à intervenção na construção do significado de uma palavra construída por regras semânticas independentes dos fatores derivacionais que intervêm na construção do significado previsível (RCM, processo morfológico em relação com a base). O componente fonológico intervém, por intermédio das regras de alomorfia e truncamento, na construção da forma superficial da palavra.

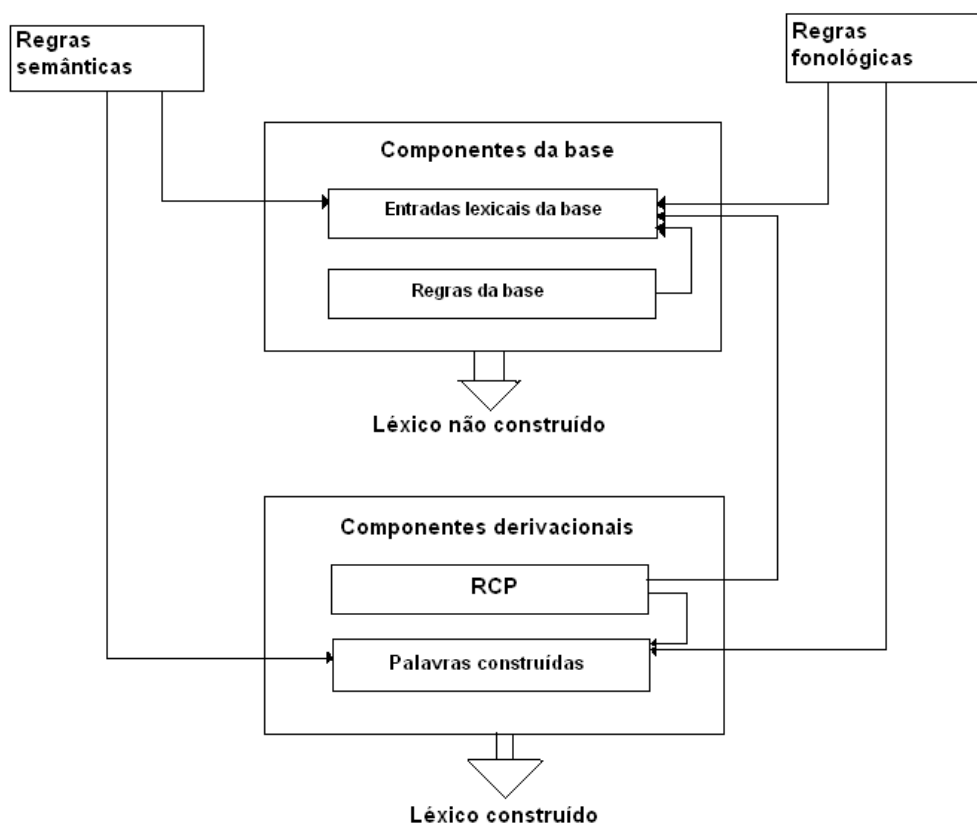


Figura 1: Arquitetura dos componentes atuantes no processo de construção de palavras

A arquitetura da Figura 1 apresenta dois Componentes: o Componente de base (1ª parte do sistema no qual regras semânticas e fonológicas operam) e o Componente derivacional (2ª parte do sistema no qual operam as RCPs e, também, novamente, podem intervir regras semânticas e fonológicas). Ambos serão descritos mais detalhadamente nas Seções 2.2.2.1 e 2.2.2.2.

2.2.2.1. O Componente de Base

O Componente de Base, de caráter idiossincrático, contém a lista das entradas lexicais de base dotadas de todas as suas propriedades e regras. A lista das entradas lexicais de base é constituída de tudo o que, no léxico, não é construído derivacionalmente:

1) Lista de morfemas lexicais não complexos, isto é morfemas primitivos, pertencentes a categorias maiores e autônomas e próprios para servir de base a palavras construídas (*varrer*³⁷, *vaso*³⁸, *árvore*³⁹, etc.).

³⁷ *Varrer*: v. Do latim *verrere* – serve de base para a construção do deverbal *varredura*.

³⁸ *Vaso*: s. Do latim *vāsu* – serve de base para a construção do dessubstantival *vasinho*.

³⁹ *Árvore*: s. Do latim *arbore* – serve de base para a construção do dessubstantival *arvorizinha*.

2) lista de palavras chamadas “transcategorizadas”, isto é, aquelas que provém de outro componente da gramática, seja pelo congelamento de grupos sintáticos, eventualmente submetidos a regras semânticas de metáfora, metonímia, etc. (*fora-da-lei* N, *zé ninguém* N), seja pela passagem de uma categoria verbal submetida a uma categoria nominal ou adjetival (passagem do infinitivo à categoria nominal, em certas condições, e os participípios, presente e passado, à categoria adjetival) (CORBIN, p. 20, 1991);

3) Lista de palavras complexas não-construídas (*arvoredo*⁴⁰, *precoce*⁴¹, *vasilha*⁴², *varredura*⁴³ etc.).

4) Lista de afixos, na qual pode aparecer afixos homônimos, caso as propriedades destes não forem tratáveis de modo unificado (por exemplo, *-agem 1* associada à regra que forma substantivos de ação sobre bases verbais (*lavagem*), e *-agem 2*, associada àquela que forma substantivos “coletivos” sobre bases nominais (*plumagem*).

Portanto, é neste componente que estão alocadas as entradas lexicais que abarcam todas as unidades lexicais não construídas e que serão combinadas no Componente Derivacional. É nesse estrato que estão localizados os afixos e as bases verbais.

No nível do componente de base podem intervir regras externas, semânticas e fonológicas. As regras semânticas permitem tratar da polissemia das entradas lexicais de base. As regras fonológicas dão conta, por exemplo, das relações alomórficas entre duas entradas lexicais semanticamente ligadas. À exceção dos afixos, as entradas lexicais figurantes neste componente podem aparecer nas estruturas sintáticas, lá compreendidas, potencialmente, as entradas não-autônomas. A saída do componente de base constitui o léxico não construído (CORBIN, p.20, 1991).

2.2.2.2. O componente derivacional

O Componente Derivacional, que é o Componente gerativo por excelência, comporta a lista das RCPs e gera as palavras construídas possíveis da língua, dotadas de todas suas propriedades previsíveis. As RCPs aplicam-se, sob a reserva de que todas as condições de aplicação sejam respeitadas, nas entradas lexicais de base pertencentes a categorias maiores, e a seus próprios produtos, já que a base de uma palavra construída pode ser uma palavra

⁴⁰ Palavra de estrutura compósita de origem latina importada para o português.

⁴¹ Palavra de estrutura compósita de origem latina importada para o português.

⁴² Palavra de estrutura compósita de origem latina importada para o português.

⁴³ Palavra de estrutura compósita de origem latina importada para o português.

construída (*regionalização* N, é formada sobre *regionaliz(ar)* V , ela mesma constituída sobre *regional* A, constituída por sua vez, sobre região N).

- A estrutura morfológica das palavras construídas é formada de acordo com esquemas de disposição categorial que são próprios a uma língua ou a um grupo de línguas dado.

Dado que a uma RCP podem ser associados muitos processos morfológicos de natureza diferente (sufixação, prefixação, conversão, etc.), uma mesma RCP é capaz de construir várias estruturas. A maioria das RCPs obedece a um princípio de unicidade categorial que lhes impõe empregar-se somente em uma categoria lexical de base para construir somente uma categoria de palavras formadas.

O SPcr das palavras construídas, determinada pelos SPspm e SPhb (ver Seção 2.2.1.1.2.), constitui o significado previsível das palavras construídas. É provável que no interior de um grupo de línguas, haja uma certa homogeneidade dos SPcr, e que se possa, assim, delimitar a noção de palavra construída semanticamente possível nestas línguas. O emprego referencial das palavras construídas tem por resultado que, ao significado linguisticamente previsível de uma palavra X, se acrescentem propriedades referenciais, que podem ser transmitidas às palavras construídas a partir de X.

Esta “interface” entre a língua e o mundo não foi representada no esquema acima; ela poderia figurar nele sob a forma de “operações referenciais”, possíveis de se aplicar tanto às palavras construídas como às palavras não construídas.

2.3. Considerações

Na Figura 2, apresentamos um quadro contendo um resumo do que foi reportado acima.

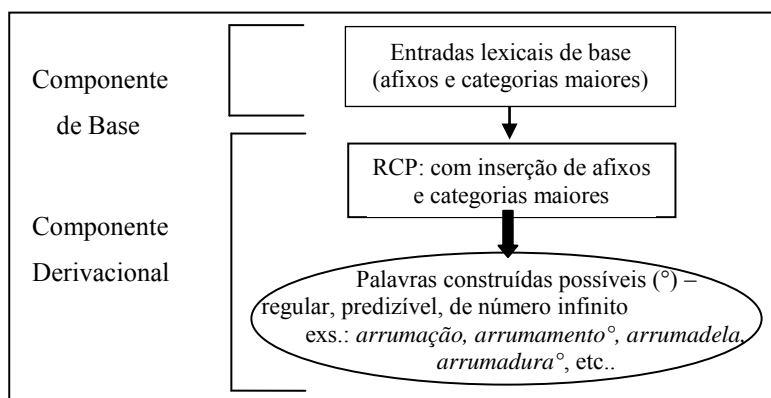


Figura 2: Arquitetura resumida dos componentes atuantes no processo de construção de palavras

A Figura 2 mostra uma parte do modelo de morfologia construcional que retrata o Componente Lexical e os estratos que participam do processo de formação das palavras construídas. A seta fina significa o que é submetido às regras e a seta grossa aponta o *output* das regras e operações linguísticas.

Para obtermos uma palavra construída possível no léxico, categorias maiores e afixos são submetidos a uma série de operações linguísticas. Segundo Corbin (1997, p.7) “Pode-se definir uma unidade lexical construída como uma unidade lexical cujo sentido, a forma e/ou a estrutura resultante de uma ou muitas operações linguísticas. Essa propriedade se marca por uma motivação recíproca da forma (e/ou da estrutura) e do sentido.”.

A escolha desse modelo é fundamental para a construção das regras de recuperação de deverbais em um *corpus* histórico, já que este é um modelo hipergerativo que considera todas as possibilidades de construção e não somente as atestadas em dicionários. Então, podemos encontrar formas que não usamos e que podem parecer estranhas ao homem de hoje. É justamente esse desprendimento de fatores de ordem extralinguística que nos ajuda a descobrir e conhecer outras formas deverbais que já existiram no período do *corpus* deste trabalho.

É importante salientarmos que a teoria construcional será empregada aqui somente em seu aspecto descritivo visando à funcionalidade do modelo que permite a sustentabilidade do sistema EXTRADEV.

Para este trabalho utilizaremos apenas o Componente de Base e o Componente derivacional que é o componente hipergerativo. Tal escolha justifica-se por nosso objetivo não levar em conta o que é ou não atestado por dicionários. O Componente convencional considera aspectos extralinguísticos e o nosso sistema trabalha com as possibilidades de construção de palavras no léxico, portanto, esse nível não é relevante para nosso objetivo. Em relação ao Componente Pós-derivacional, duas problemáticas podem vir à superfície em relação aos dados obtidos: geração de muitas unidades agramaticais; e perda de deverbais que deveriam passar pelo componente de correção – pós derivacional. Porém, em relação à hipergeratividade de unidades lexicais agramaticais, elas serão barradas no *corpus*, ou seja, só será extraído do *corpus* aquilo que existir, portanto os dados não terão unidades não válidas. Em relação às perdas, não teremos nos dados extraídos alguns deverbais que sofrem ou sofreram alterações de truncamento ou integração paradigmática (por ex.: coleccionar → coleccionação* → coleção). Porém, acredita-se que a perda seja pequena, já que é uma minoria do léxico que está suscetível a processos como esses. Em trabalhos futuros, com um

refinamento no algoritmo de busca de deverbais do sistema EXTRADEV, poder-se-á recuperar essas unidades que são representadas e sistematizadas pelo modelo SILEX no Componente pós-derivacional.

Um sistema estratificado, com a hierarquização dos Componentes Lexicais e o conhecimento das operações que podem intervir sobre cada Componente, torna muito mais claro o processo de formação das palavras construídas. E, por conseguinte, permite que se construam algoritmos de busca de deverbais de alta cobertura e muito mais precisos do que os buscadores comuns.

3. Deverbais de ação: denominação, significado, sintaxe e estrutura

Neste Capítulo, primeiramente, propomos uma terminologia para a categoria dos deverbais e então procedemos à descrição dos tipos de estruturas morfológicas passíveis de constituírem deverbais de ação guiadas pela RCP responsável. Sendo necessário para isso a apresentação de alguns conceitos relacionados ao modelo de análise escolhido.

3. 1. Denominação

Após a investigação em gramáticas, dicionários e trabalhos acadêmicos sobre os deverbais, decidimos utilizar a nomenclatura “deverbal de ação”, não somente verbal, tampouco, nome de ação, visando uma terminologia caracterizada pela transparência. Ou seja, é uma tentativa de fugir de termos incompletos, ou opacos, que acabam focando apenas em alguns dos aspectos que são relevantes para designação do termo e, desse modo, podem contribuir para a obscuridade terminológica.

Como pudemos observar a gramática normativa e mesmo os próprios linguístas não utilizam uma terminologia que representa os aspectos relevantes dessa categoria. A utilização do termo ora verbal, ora nomes que vem de verbos podem refletir, justamente, o foco estrutural que geralmente se dá para as palavras construídas. Além de refletir somente a parte estrutural da unidade lexical, há mais de uma classe gramatical construída a partir de verbos, tornando-se uma definição muito abrangente e, portanto, insuficiente para a especificidade que a categoria requer.

A palavra verbal significa aquilo que vem do verbo. A parte verbal do verbal é a base. Entretanto, existem outras categorias que também são formadas a partir de bases verbais, por exemplo, a base verbal que dá origem a outros verbos. Em seguida podemos observar a RCP que gera verbos a partir de bases verbais:

V → V
saltar → saltitar

Ou a RCP que gera adjetivos a partir de bases verbais:

V → Adj
 vibrar → vibrante;
 acreditar → acreditável
 contar → contável

E, finalmente, substantivos que são construídos a partir de bases verbais, e estes são os mais produtivos em relação à base verbal:

V → N1 (designam ação)
 N2 (designam agente)

Desse modo, a utilização do termo deverbal pode fazer referência tanto a adjetivos, substantivos e ao próprio verbo. Outra maneira encontrada na literatura de denominar os deverbais é por “nomes de ação”. Nesse caso, ocorre o oposto, o foco está apenas na questão semântica, deixando de lado a parte estrutural. Da mesma maneira que “deverbal” apenas é um termo carente de informações em relação à especificidade do objeto que designa, adotar somente “nomes de ação” também teria o mesmo efeito de incompletude. Existem nomes denominais que ostentam significados próximos de nomes de ação. Tome-se como exemplo os nomes em -ada, que podem ser parafraseados por “golpe desferido com N (nome de base)”:

laço → laçada
 braço → braçada
 pata → patada
 vassoura → vassourada

A adoção do termo “deverbais de ação” reflete um dos fundamentos principais do modelo SILEX - o caráter associativo. A terminologia é coerente com a teoria adotada, a qual não prioriza a parte estrutural, mas sim considera significado e forma no mesmo nível.

Então teríamos:

verbo → substantivo = deverbal de ação

A Figura 3 ilustra as classes construídas a partir dos verbos, focando nos substantivos deverbais e nos sentidos que assumem. A linha laranja faz referência exatamente aos deverbais de ação.

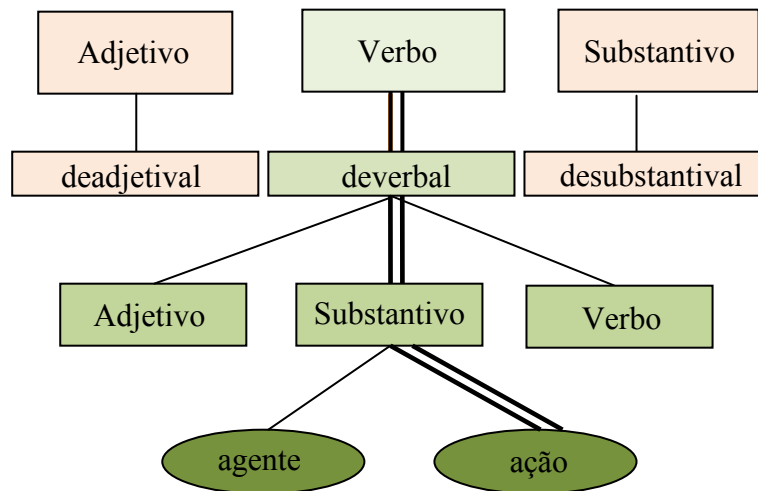


Figura 3: Estrutura das classes gramaticais que funcionam de base para processos construcionais.

A Figura 3 apresenta um esquema das classes gramaticais que servem de bases para os processos de derivação (no topo); no nível 2 os nomes dados às unidades lexicais formadas a partir das classes base; abaixo, as classes construídas; e, finalmente, no último nível, os significados que podem assumir tais unidades. O traçado duplo delimita o caminho na estrutura para os deverbais de ação, que têm na base o verbo – portanto deverbais – e formam substantivos que denotam ação.

Após justificarmos a escolha da denominação da categoria que elegemos para este trabalho, iniciaremos uma breve descrição⁴⁵ sobre os deverbais de ação do ponto de vista formal, semântico e da estrutura argumental.

⁴⁵Não nos compete neste trabalho uma descrição aprofundada e minuciosa dos aspectos formais, semânticos e estruturais, mas sim, uma descrição suficiente para construirmos o sistema computacional.

3.2. O significado dos deverbais de ação

Os deverbais de ação são formas nominais que expressam algum acontecimento. Podem ser usados como outra forma de construir o discurso, preserva-se o significado e muda-se a estrutura argumental ou, também, para nomear processos e atividades. Uma maneira de verificar se o nome é um deverbais de ação é parafraseá-lo por “ato ou efeito de Vb”, ou seja, *combinação* é um deverbais de ação porque pode ser parafraseado por ato ou efeito de *combinar*.

Deverbais de ação = Ato ou efeito de Vb

Apesar de todos os deverbais serem parafraseados pela frase acima, mantendo o sentido de acontecimento, eles são polissêmicos como já foi mencionado. Esse fato implica existir outros sentidos que estão relacionados aos complementos dos verbos e a fatores extralinguísticos. Para exemplificarmos esse fenômeno, podemos citar o verbo *parar*:

a) A **parada** do motorista foi brusca.

A sentença acima denota “ato ou efeito de parar”

b) Vou descer na **parada** sete.

Nessa frase, *parada* assume o sentido do argumento do verbo “local onde se para”.

c) Posso saber qual é a **parada**?

A palavra *parada*, nesse contexto, assume o sentido de uma situação de entrave, um acontecimento a ser resolvido de difícil solução. Vejamos as acepções encontradas para o substantivo *parada* no dicionário Michaelis⁴⁶:

pa.ra.da

sf (parar+ada1) 1 Ato ou efeito de parar. 2 Lugar onde se pára, especialmente os pontos de bonde, ônibus, trem etc. 3 Demora, pausa. 4 Quantia que se aposta ou se arrisca de cada vez no jogo. 5 Reunião ou passagem de tropas para revista ou exercício. 6 Esgr Ato de se defender de um golpe. 7 Aventura. 8 Empresa ou situação arriscada. 9 Conversa fiada,

⁴⁶<http://michaelis.uol.com.br/moderno/portugues/index.php?lingua=portugues-portugues&palavra=parada>

fanfarronada. 10 Reg (Sul) Quantia pela qual se contrata uma carreira de animais ou rinha de galos. 11 Golpe de capoeiragem. sm Pessoa fanfarrona. Parada dura: situação ou pessoa difícil de vencer. Parada morta, Reg (Sul): a parada de corrida de cavalos ou briga de galos que um dos contratantes é obrigado a pagar, mesmo quando, por doença ou qualquer outro motivo, não faça correr seu cavalo ou brigar seu galo. Comprar a parada, Reg (Rio Grande do Sul): tomar as dores por outrem. Não enjeitar a parada: aceitar o desafio. Topar a parada: o mesmo que não enjeitar a parada.

Sobre uma mesma base é possível construir vários deverbais, potencialmente polissêmicos. Como a língua tem tendência a anular a sinonímia, porque contraria o princípio de economia do sistema, um (ou dois) dos deverbais assumem a designação da ação, enquanto os restantes vão sofrer especializações semânticas, deixando o seu significado básico de estar presente no léxico disponível aos falantes ou registrado nos dicionários.

Podemos citar outro exemplo de polissemia verbal. No caso do verbo *gravar* podemos derivar: *gravação*, *gravamento*, *gravura*, etc. A construção dessas palavras faz parte da mesma Regra de Construção de Palavra (RCP⁴⁷) de deverbais de ação, entretanto, *gravura*, apesar de significar *Ato ou efeito de gravar*, é usada mais frequentemente com a acepção de “produto da ação de gravar”⁴⁸:

1-A *gravura* foi descolada pelos meninos.

2-A *gravação* foi elogiada pelos diretores.

As frases acima mostram que em ambas o verbal ocupa posição de sujeito e é seguido de verbo de ligação mais a forma participípio do verbo e o agente da passiva. Apesar de apresentarem a mesma estrutura sintática, no enunciado 1, temos um substantivo concreto que expressa produto ou resultado da ação e, no exemplo 2, temos um substantivo abstrato que expressa ação. Ambos são substantivos derivados de verbos, entretanto, o primeiro é um substantivo concreto e não abstrato que denota um objeto. Esse é um caso de polissemia que

⁴⁷ Os deverbais de ação pertencem a mesma Regra de Construção de Palavra porque são construídos a partir dos mesmos constituintes designando o mesmo sentido (ou seja, por exemplo os deverbais de ação construídos por sufixação são formados a partir do tema, mais um conjunto determinado de sufixos, já os deverbais de agente apresentam a mesma estrutura formal, porém, um conjunto de sufixos diferentes e um sentido resultante distinto, portanto pertencem a outra RCP), essas duas características determinam uma RCP.

⁴⁸ Em casos de sinonímia, um dos sinônimos tende a especializar-se quer semanticamente (assumindo significados específicos que o seu sinônimo não ostenta), quer do ponto de vista do uso (por exemplo, beijo e osculo, ou gripe H1N1 – mais científico por oposição a gripe suína – mais popular, ou encefalite espongiiforme dos bovinos por oposição a doença das vacas loucas).

só pode ser solucionado a partir da frase na qual está inserido deverbais.

3.3. Comportamento sintático dos deverbais de ação

Sujeito e complemento⁴⁹ são as principais possibilidades que assumem os deverbais na estrutura sintática, porém eles também aparecem no interior do adjunto adverbial.

Sujeito:

*A **distinção** dos processos foi camuflada.*

*Era bem conhecida a sua **obstinação** por Maria.*

Complemento:

*Ele não sentiu **arrependimento** por perder o jogo.*

*Conte a eles sobre a **aplicação** ilícita de nosso gerente.*

Adjunto Adverbial:

*Na **chegada** ao estádio, ele se deu conta de que não havia ingresso.*

*Em meio ao **fechamento** dos portões, Júlia não percebeu o roubo da carteira .*

Após alguns exemplos dos papéis que os deverbais assumem na estrutura sintática, iniciaremos nossa análise da estrutura argumental a partir do exemplo abaixo:

*A **combinação** da roupa foi feita por Maria.*

O substantivo abstrato *combinação* nomeia a ação de combinar realizada por Maria. A ação expressa na forma nominal é herdada do verbo *combinar* do qual *combinação* é palavra derivada por sufixação⁵⁰. A mesma frase pode ser escrita substituindo-se o nome pela forma verbal, ocasionando mudanças sintáticas e preservando o sentido:

*Maria **combinou** a roupa.*

⁴⁹ Adotamos os termos sujeito e predicado por razões didáticas, já que este e aquele são terminologias mais populares.

⁵⁰ Nas próximas seções discorreremos sobre os processos de formação de palavras, nos quais também está inserida a sufixação.

O ato de combinar a roupa continua sendo realizado por Maria, entretanto, no lugar de um nome temos um verbo e no lugar do complemento nominal (*da roupa*) temos o objeto direto (*a roupa*). O sujeito agora é agente e não mais agente da passiva. Observemos o que diz Basílio (p.54) sobre a necessidade sintática de um deverbais na frase: “Algumas das principais motivações para a formação destes substantivos deverbais são (i) a necessidade de expressar noções verbais em estruturas que sintaticamente exigem a ocorrência de um substantivo.”

Vejamos alguns exemplos:

*Os artistas **combinaram** peças de diferentes culturas para o evento.*

*O público apreciou a **combinação** das peças no evento.*

“i) a necessidade de referência genérica ao evento representado pelo verbo ou seu efeito, sem especificações de argumentos e categorias gramaticais, como em (3)”:

Esta casa precisa de uma arrumação.

“(iii) a necessidade de referência a uma instância específica do evento representado pelo verbo”:

Ela deu uma chorada e conseguiu o prêmio.

E “(iv) a necessidade de referência a frases e orações através do uso da forma nominalizada do verbo”:

A empresa Salvatore fabricou mais de vinte mil amostras do novo produto no mês de junho. A fabricação exige uma série de procedimentos...

A escolha de um deverbais na frase (e não o próprio verbo) enfatiza a ação no discurso e não quem a realiza. Mas os efeitos discursivos do uso de deverbais não serão abordados neste trabalho. Apenas nos compete descrevê-los para compreendermos de uma maneira geral os principais aspectos morfossemânticos desta categoria de palavras.

3.4. Estrutura

Os deverbais de ação, segundo o modelo adotado, podem assumir dois tipos de estrutura: palavras complexas contruídas e palavras complexas não-contruídas.

a) palavras complexas não-construídas são unidades lexicais de estrutura compósita, porém a concordância entre estrutura e significado é parcial. Ou seja, são palavras que não se construíram em português. Muitas entraram nesta língua como resultado de importações diversas, ocorridas em várias épocas e a partir de línguas diferentes, apresentando estruturas simples ou complexa, na sua língua de origem.

b) palavras construídas são unidades lexicais de estrutura compósita que apresentam concordância mais ou menos imediata entre a estrutura morfológica e o significado exibido pela palavra (CORREIA, p.100, 2004). Contudo, é possível estabelecer as seguintes distinções em relação às palavras construídas de acordo com a tipologia sufixal: existem estruturas, que do ponto de vista sincrônico, são analisáveis na língua, por exibirem uma base reconhecível e uma forma sufixal que é possível encontrar num número significativo de outras estruturas semelhantes (construindo um paradigma), porém não foram construídas na determinada língua. Vejamos, abaixo, os tipos de sufixo:

1) Sufixos fósseis:

- surgem em estruturas analisáveis em português; são portadores de significado relacional característico do português; estabelecem com suas bases uma relação semântica que é produtível em outras estruturas semelhantes; são integráveis em paradigmas sufixais próprios de RCP do Português; são, portanto, integráveis como sufixos do português; porém, nenhuma das palavras que os exibem foi espontaneamente construída em português, muito embora algumas delas possam resultar de replicas deliberadas de estruturas de outras línguas, por exemplo, do latim ao nível do discursos científico-técnicos (CORREIA, p. 106, 2004)

2) Sufixos internacionais

- surgem em estruturas analisáveis em português; são portadores de significado relacional característico de sufixos do português; estabelecem com suas bases uma relação semântica que é produtível em outras estruturas semelhantes; integráveis em paradigmas sufixais próprios de RCP do Português; os seus cognatos portugueses -se disponíveis para a construção de novas unidades; são, portanto, integráveis como sufixos do português; as estruturas que os exibem podem ou não ter sido construídas em português.

3) Sufixos autóctones

Por outro lado, existem casos que um determinado sufixo é efetivamente autóctone,

isto é, todas as unidades lexicais que o constituem foram construídas na língua em estudo, neste caso, o português. Podemos citar os sufixos formadores de nomes de qualidade *-ez* (ex.: *lucidez*, De *lúcido*. Em 1873, D. V.⁵¹) e *-eza* (ex.: *esperteza*, De *esperto*. Em 1813, Morais⁵²).

Em uma breve consulta a dicionários de etimologia constatamos que boa parte dos deverbais de ação investigados apresentam estrutura importada integralmente do latim como, por exemplo, a palavra *composição* (**COMPOSIÇÃO** XIII. Do lat. *compositio –onis*); *criação* (**criAÇÃO** XIV Do lat. *criatio –onis*). Ambas foram construídas no latim e importadas para língua portuguesa. Da mesma maneira comporta-se o sufixo *-(t/d)ura* também pode ser encaixado da mesma forma em palavras como *ligadura*: do latim *ligatura*; *ligamento* do latim: *ligamentum* (CUNHA, 1982). Esses sufixos são classificados como sufixos internacionais segundo a descrição acima. Ex.: *motivação* segundo Dicionário Etimológico Nova Fronteira (p. 535, 1982) tem sua origem datada em 1899, ou seja, uma palavra construída no português com um sufixo internacional. Além disso, apesar de origem latina, apresentam sufixos cognatos em diversas línguas e comportamentos semelhantes em todas as línguas nas quais surgem – línguas românicas e em inglês (Do inglês: *motivation*; Do espanhol *motivación*; do português *motivação*). A partir dos dados obtidos neste trabalho, voltaremos a discutir as estruturas dos deverbais no Capítulo 4 (Análise dos dados).

Se adotarmos uma postura sincronista na descrição das estruturas de palavras de uma língua todos os derivados referidos em b) serão tratados ao mesmo nível como palavras construídas. Porém, além de descrever a competência derivacional dos falantes, também descrevemos uma faixa do léxico da língua portuguesa e este léxico apresenta palavras que entraram nele, em épocas, modos e proveniências diferentes. Neste trabalho, os dados da história da língua, e da história das palavras são de extrema relevância e por isso, considerados. Esta é a postura assumida aqui, em coerência com a postura atual do modelo SILEX.

Nesse cenário, atentamos que somente as palavras construídas com sufixos autóctones, sobre bases autóctones, construídas dentro da própria língua, não geram problemas significativos dada sua vernaculidade e transparência (*escassez* → *escasso* (S) + *-ez* (sufixo) → o fato de ser escasso⁵³).

Dada a explanação sobre a tipologia de sufixos, uma breve análise sobre alguns

⁵¹ (MACHADO, p. 1440, 1967)

⁵² (MACHADO, p. 942, 1947)

⁵³ Para saber mais sobre morfologia construcional do português, ler “Denominação e Construção de Palavra” (CORREIA, 2004).

deverbais de ação, podemos delimitar – até o momento – as seguintes estruturas de deverbais de ação:

- substantivos derivados por sufixação;
tema + sufixo (*cria__ção*) ;
tema do particípio passado + sufixo (*feit__ura*);
- substantivos resultantes de conversão;
radical + desinência de gênero (dg) (*pesquis__a*);
- substantivos resultantes de desflexionação;
forma desflexionada do infinitivo ou particípio. (*brigar/parada*)
- palavras complexas não construídas;
palavras formadas no latim (*geração*, s. Do lat. *generatione*, Séc. XIII), ou construídas no português com sufixos internacionais. (*mortificação* s. De *mortificar*, não do Lat. *mortificatione-*, séc. XVI⁵⁴; *molhamento*, s. De *molhar*. Séc. XV⁵⁵)

A presente classificação é provisória, será esmiuçada no Capítulo 4. Mas neste momento ela permite prosseguir a organização e análise dos dados.

3.4.1. Sufixos formadores de deverbais de ação

Após a revisão da literatura (Capítulo 1) constatamos que não há unanimidade na lista de sufixos dados como formadores de deverbais. Observemos, na Tabela 4, os sufixos deverbais considerados de cada autor:

A – Toledo (1976)

B – Sandmamm (1989)

C – Bechara (1992)

D – Celso Cunha (1970)

E – Este trabalho

⁵⁴ (MACHADO, p. 1611, 1967)

⁵⁵ (MACHADO, p. 1611, 1596)

Tabela 4: Sufixos escolhidos para este trabalho.

A	Ança	são ção	mento	encia	ura	agem	--	--	--	--	--	--	--	--	
B	--	ção	mento	--	dura	--	ada	deira	--	--	--	--	--	--	
C	Ança Ancia	ção	mento	ença ncia	ura dura tura	agem	ada ida	--	ame	ata	ário	--	--	--	
D	Ança ância	são ção	mento	ença ência	(t)jura (d)jura (s)jura	--	--	--	--	--	--	tório	dório		
E	nça	ção	mento	ncia	(t/d)jura	gem	--	eira	me	ta	nço	nsia	ria	são	(d)ela

A partir da investigação dos sufixos deverbais na literatura especializada, decidimos utilizar todos os que foram encontrados. Tal decisão baseou-se na tentativa de recuperar todos os deverbais que eram utilizados nos séculos XVI, XVII, XVIII e parte do XIX. Para tentar justificar essa inconsistência no conjunto de sufixos deverbais de ação, podemos citar dois argumentos principais: baixa ocorrência na língua em uso de determinados sufixos; maior ocorrência de determinado sufixo com sentido diferente de ato ou efeito de verbo base⁵⁶.

Muitos trabalhos atuais citam apenas os mais recorrentes, isso significa que esses excluídos podem ter sido usados com maior frequência em outro período - ainda não sabemos – porém se algum estudioso do português chegou a citá-los, espera-se que esses sufixos tenham sido representativos em algum momento da história da língua. Nesse sentido, devemos procurar deverbais de ação com todos os sufixos citados para que não percamos possíveis ocorrências que tiveram algum destaque no passado.

A Tabela 5 apresenta os sufixos formadores de deverbais de ação de origem latina e seus respectivos produtos no português brasileiro.

⁵⁶ Por exemplo, o sufixo *-gem* que tem alta produtividade para desubstantivais. Ex: plumagem, folhagem, etc.

Tabela 5: Sufixos formadores de deverbais de origem latina.

Latim	Português
<i>-mentum</i>	<i>-mento</i> – alisamento, apaziguamento
<i>-tio, -tione</i> <i>-sio -sione</i>	<i>-ção</i> – pregação, animação <i>-são</i> – expansão, expulsão
<i>-tura</i>	<i>-(t/d)</i> – ligadura, atadura
<i>-me</i> ⁵⁷	<i>-ame</i> – ligame, azedume
<i>-ta</i>	<i>-ta</i> – berrata, passeata
<i>-ariu</i>	<i>-eira</i> – bebedeira,
<i>-one</i>	<i>-ão</i> – puxão, arranhão
<i>-antia</i> <i>-entia</i>	<i>-ança</i> – esperança, lembrança <i>-ância</i> – concordância, alternância <i>-ença</i> – parecença, fervença <i>-ência</i> – experiência, exigência

Não foi encontrado nenhum tipo de informação etimológica para o sufixo *-(d)ela*.

Outra questão a ser tratada é em relação à fronteira dos sufixos e dos radicais. Vejamos o que diz Câmara (1976) sobre a questão sufixal:

(...) A primeira é a variabilidade do limite entre o que se considera sufixo e o radical. Na história da língua os sufixos se ampliam ou se reduzem, incorporando um fonema do radical ou destacando de si o que era um seu fonema inicial. (...) mesmo na descrição atual da língua um dado sufixo pode apresentar variações a esse respeito na maneira por que entram em palavras derivadas.

Diante dessa variação, a posição adotada nesta pesquisa foi delimitar as fronteiras isolando o radical, vt e sufixo, para alguns sufixos e não deixar a vogal temática unida ao sufixo para outros. Tal decisão foi tomada por duas razões: a primeira, para não perdermos ocorrências, e a segunda, para evitarmos a extração de ambigüidades. Desse modo nosso conjunto de sufixos é um pouco diferente dos demais, como mostramos me seguida: *-mento/-mento*, *-ação/-ção*, *-ansia/-nsia*, *-agem/-gem*, *-ancia/-ência/-ncia*, *-(d/t)ura / -(d/t)ura*, *-dela/-dela*, *-ão/-ão*, *-anç/-nç-*, *-são/-são*, *-deir/-deir-*, *-ame - -ume/-ume -ame*, *-ata/-ata*.

Embora alguns sufixos ocorram sempre com a mesma vogal temática por se unir a

⁵⁷ O sufixo *-me* aparece com a forma de *-ame* (CUNHA, p.228,), ou seja, adjunto a vogal temática, porém no exerto é citada a troca de vogal temática por 'u' *-ume*. Essa variação – que não é exclusiva deste sufixo- contribui para a decisão de separar as vogais temáticas do sufixo como fora adotado neste trabalho.

bases apenas de determinada conjugação, existem outros que podem apresentar variação na vt devido a dependência em relação a que conjugação o verbo base ao qual se une pertence. Nesse sentido, adotou-se um forma sufixal mais flexível (supracitada), a qual permite que ocorrências não frequentes e, por isso, desconsideradas apareçam. Visamos não perder as ocorrências que apresentam variação na vt.

Seguem, abaixo, as tabelas com as possíveis estruturas deverbais de ação (6 a 8).

Tabela 6: Estrutura dos deverbais de ação possíveis a partir do verbo puxar.

Verbo	Tema(bv)	Sufixo	Deverbal de ação
<i>Puxar</i>	<i>Pux(a)-</i>	<i>-mento</i>	<i>puxamento</i> ^o
		<i>-ção</i>	<i>puxação</i> ^o
		<i>-ncia</i>	<i>puxância</i> ^o
		<i>-gem</i>	<i>puxagem</i> ^o
		<i>-(t/d)ura</i>	<i>puxadura</i> ^o
		<i>-dela</i>	<i>puxadela</i> ^o
		<i>-aria</i>	<i>puxaria</i> ^o
		<i>-ão</i>	<i>puxão</i> ^o
		<i>-nç(a/o)</i>	<i>puxanço</i> ^o / <i>puxança</i> ^o
		<i>-são</i>	<i>puxasão</i> [*]
		<i>-deira</i>	<i>puxadeira</i> ^o
		<i>-ata</i>	<i>puxata</i> ^o

Participam da RCP de deverbais de ação todos esses sufixos embora alguns sejam muitos mais recorrentes que outros e, muitos, participem e sejam mais produtivos em outras RCPs.

Tabela 7: Estrutura dos deverbais possíveis a partir do verbo pesquisar

Verbo	Tema(bv)	Desinência gênero	Deverbal de ação
<i>Pesquisar</i>	<i>pesquis-</i>	<i>-a</i>	<i>Pesquisa</i>
		<i>-o</i>	<i>Pesquiso</i> ^o
		<i>-e</i>	<i>Pesquise</i> ^o

Tabela 8: Estrutura dos deverbais de ação possíveis a partir do verbo parar.

Verbo	Deverbal de ação
<i>Parar</i>	<i>Parar</i>
<i>Parada</i>	<i>Parada</i>

No caso acima (Tabela 8) não há alteração formal. A forma verbal é idêntica a forma nominal.

3.5. Mecanismos de construção de deverbais de ação

Os processos morfológicos de construção de palavra em geral são divididos em derivação e composição. Para os deverbais somente a derivação nos interessa. A derivação, segundo a Gramática Tradicional (GT), abarca os seguintes processos de formação de palavras: prefixação, sufixação, derivação regressiva, derivação imprópria.

Por sua vez, os deverbais de ação são formados pelos processos:

- 1) Sufixação
- 2) Derivação regressiva
- 3) Derivação imprópria

Podem ocorrer algumas distinções nas denominações e definições dos processos de construção de palavra da GT em relação a outras teorias linguísticas e até entre os próprios linguistas. As diferenças são sutis do ponto de vista teórico e, como veremos em seguida, os autores divergem há décadas e, a cada dia surgem novas análises para tais inconsistências no estudo da formação de palavra.

Consultamos algumas das obras mais renomadas a fim de conhecer como tais processos são compreendidos e em quais aspectos apresentam diferenças em relação ao adotado neste trabalho. As seguintes obras foram consultadas: Compêndio de Gramática

Histórica Portuguesa de Dr. Joaquim José Nunes (1945); Gramática da Língua Portuguesa de Maria Helena Mira Mateus et.al. (1983); Gramática Histórica de Said Ali (1923); História e Estrutura da Língua Portuguesa Matoso Câmara (1976); Formação lexical de Sandmann (1991); Margarida Basílio (2004); e a obra mais recente Correia e Lemos (2005) que segue o mesmo modelo que será utilizado neste mestrado – o SILEX.

Segundo o modelo SILEX, existe uma RCP para a construção de deverbais de ação, na qual se incluem dois processos, a sufixação e a conversão (a também chamada derivação regressiva na gramática tradicional). Além desses processos derivacionais, os deverbais podem, também, resultar da desflexionação de formas de particípio passado ou de infinitivo (por exemplo, *a partida, a chegada, o falar e o olhar*).

Dentre os processos de formação denominados acima, dois deles coincidem com a nomenclatura adotada pela GT – sufixação e conversão – entretanto, apenas um coincide com a definição – o processo de sufixação. Os três processos são detalhados abaixo:

a) Sufixação

Tanto o modelo SILEX quanto a GT e a gramática histórica estão de acordo que a derivação sufixal é a junção do sufixo ao tema do verbo.

De acordo com Nunes (1945):

O processo de formação vocabular mais produtivo é o popular que se deu por duas maneiras. A primeira é a atribuição a palavras já existentes papel diferente do que até então representavam, ou também através da adição de elementos novos para modificar a primeira idéia e exemplifica: “assim estrada, vagar, por exemplo, de formas verbais tornaram-se substantivos, de velho e alvo fêz-se velhice, alvura;

O que difere é como entendemos cada um desses constituintes e o caráter produtivo deste processo. Para a GT o sufixo tem apenas função gramatical e se aglutina a um radical para originar uma nova palavra definindo sua categoria. Na concepção do modelo construcional, o sufixo não tem apenas função gramatical: ele ajuda a construir o significado da palavra derivada; não se une a um radical ou tema de maneira aleatória; há uma RCP específica que une forma, sentido, levando em conta também aspectos fonológicos e semânticos para as palavras derivadas, enquanto a GT pauta-se apenas em aspectos morfossintáticos. Cunha e Cintra (2001, p.83) logo na introdução do Capítulo sobre derivação

trazem uma citação do dicionário de lingüística de Dubois (1973): “Chama-se FORMAÇÃO DE PALAVRAS o conjunto de processos morfossintáticos que permitem a criação de unidades novas com base em morfemas lexicais. Utilizam-se assim, para formar as palavras, os afixos de derivação ou os procedimentos de composição”.

Para Nunes (1945, p. 378) a sufixação é denominada “Derivação própria”, a qual ele atribui a grande riqueza da língua:

(...) processo pelo qual se criam palavras novas, adicionando aos radicais existentes certos elementos, que podem constar de uma ou mais sílabas, chamados sufixos os quais vão modificar-lhes a significação. É a êle que a língua deve a sua principal riqueza; por êle foi em certa maneira compensada a perda de grande número de raízes que ele acusa, quando a comparamos com o latim; a sua força ostenta-se ainda tão fecunda, como no período da formação.

Em seguida, definiremos algumas variáveis para facilitar a explanação dos processos de construção de deverbais de ação por sufixação.

V= verbo; **r**= radical; **vt**= vogal temática; **t**= tema⁵⁸; **t** = **r** + **vt** ; **sufix**= sufixo; **Da**= deverbais de ação;

$$V \rightarrow r + vt + sufix = Da$$

$$partição \rightarrow part- + -i- + -ção = partição$$

Ou

$$V \rightarrow t + sufix = Da$$

$$Arrumar \rightarrow arruma- + -ção = arrumação$$

A GT e a Gramática Histórica partem da observação do léxico, simplificando o processo de derivação a uma simples junção do morfema ao radical. Podemos classificar como uma análise muito superficial se considerarmos todos os aspectos linguísticos que atuam num processo de derivação.

⁵⁸ O variável tema armazena tanto o tema a partir do infinitivo verbal, quanto a partir do participio passado.

b) Conversão

Chamamos de conversão o que a GT designa como processo de derivação regressiva. Segundo o modelo SILEX no processo de conversão a base verbal – no caso o radical – converte-se à forma nominal e une-se a uma vogal que funciona como uma marcadora de gênero. E, portanto, não há um processo de redução, preserva-se o radical do verbo que não se une a nenhum sufixo, mas sim a uma vogal designadora de gênero (-e, -a, -o).

A conversão é um dos processos de formação de palavra mais polêmicos no que tange a derivação. Tanto linguístas quanto gramáticos apresentam diversos pontos de vistas em relação a definição de qual, de fato, é o termo derivante e outra questão controversa é a caracterização da vogal final que se une ao radical convertido em nome. Vejamos como Sandmann (1991) posiciona-se a respeito:

A existência em português de fatos como buscar – (a) busca; implantar – (o) implante; manejar – (o) manejo, de um lado; e esfolar – (a) esfolo, esfoladura, esfolamento; descascar – (a) descasca, (o) descasque, descascação, descascadura, descarcamento, de outro, levou os gramáticos ou lingüistas a chamar o processo de formação de “busca”, “implante”, “manejo”, “esfolo”, “descasca”, “descarque”, de derivação regressiva, eis que, diferentemente de “esfoladura”, e “esfolamento”, “descascação”, “descascadura” e “descascamento”, em que há a adição dos morfemas sufixais -ção, -dura, -mento, nesses substantivos deverbais haveria a supressão de sufixo. (p.74)

O fato de não haver no processo de formação dessas palavras o acréscimo de sufixos como os citados acima, não quer dizer que algo foi reduzido, pois a base se mantém em ambos e é somente ela que é o *input* na construção dos deverbais. Porém, ocorre uma junção de morfemas distintos. De acordo com nossa regra, para deverbais de ação formados por conversão temos:

V= verbo; **r** = radical; **Da** = deverbais de ação; **dg** = desinência de gênero

$$V \rightarrow r + dg = Da$$

$$pesquisar \rightarrow pesquis- + -a = pesquisa$$

Para Sandmann, o deverbal formado por conversão é composto pelo radical mais o sufixo zero. E combate aqueles que defendem que a desinência de gênero (-e, -a e -o) é na verdade a vogal temática do verbo e exemplifica:

(...) a vogal final das palavras formadas por derivação regressiva é um tema à parte. Dentre as oito palavras de nosso cópús, todas derivadas de verbos com sufixo -ar, quatro terminam em -o (afrouxo, entorno, sufoco, tateio), três em -e (aporte, desarme, despite) e uma em -a (desossa). As razões para essa escolha não são em geral claras. No caso de tateio, derivado de tatear, podemos-nos basear no modelo dos verbos terminados com o sufixo -ear, e que tem derivações regressivas correspondentes em -o: passear → passeio, ratear → rateio, recear → receio. Nas outras palavras a razão para a escolha da vogal final é mais difícil de constatar. Talvez seja a semelhança fonética com outras palavras formadas por derivação regressiva já existentes na língua. **De qualquer maneira a escolha não tem nada a ver com a vogal do sufixo verbal.** Isso foi mostrado pelas oito palavras formadas por derivação regressiva de nosso cópús.

Diferentemente de Sandmann, Matoso Câmara afirma:

Há um outro padrão, que se desenvolveu amplamente em latim vulgar pelo modelo de algumas formas que já existiam em latim clássico, e consiste na formação de um substantivo com o radical verbal e a vogal de tema em -a, -o, -e. Ex: consolar: consolo; comprar: compra; atacar: ataca. (1976, p. 223).

Para o autor o fato de a vogal temática ser alterada para -o é um caso de alomorfa que ocorre com os verbos da primeira conjugação. Justificada pelo fato de não existir uma vogal temática -o.

Correia e Lemos (2005, p.34) estão em consonância com a visão adotada neste trabalho:

Observando este fenómeno de outra perspectiva, o que acontece, de facto, é que o radical verbal (fug(ir)V, atac(ar)V, ou us(ar)V) assume a categoria de substantivo, adquirindo, apenas, uma vogal, que é a sua desinência (o seu marcador) de género. Por isso abordagens mais recentes deste fenómeno defendem que, em rigor, a derivação regressiva não se verifica que estamos, sim, perante casos de conversão.

Já Basílio (2004, p.44) destina a maior parte de suas reflexões a determinação da direcionalidade do processo (verbo → nome | nome → verbo) e não faz nenhuma consideração sobre o constituinte que se une ao radical verbal: “A derivação regressiva com o apoio na vogal –a (...)”.

Os linguístas acordam que nesse processo ocorra uma conversão, mudança do radical verbal para nominal, entretanto não há um consenso em relação a vogal que se une ao radical – Desinência de gênero? Vogal temática? Neste trabalho, seguiremos a visão do modelo SILEX que considera desinência de gênero as vogais finais adjuntas ao radical.

d) Desflexionação

Por último, detalhamos o processo que concebemos como desflexionação e a GT denomina como derivação imprópria ou conversão. Segundo Cunha e Cintra (2001 p. 104), a conversão existe quando uma palavra muda de categoria gramatical sem ocorrer qualquer mudança formal.

olhar → *olhar*
(V) (Subst.)

Segundo o modelo SILEX, a desflexionação é um processo de construção não-morfológico de palavras, pois não há nenhuma operação derivacional na mudança categorial de verbos para nomes. “A 'desflexionação' (“déflexivation”), que consiste na passagem para a componente lexical de uma forma flexionada de uma unidade lexical (exs.: jantarN, torradoAdj ou avermelhadoAdj) (CORREIA, 2004)”.

Essas formas flexionadas podem ser formas de infinitivo ou resultante de participio passado que passam a funcionar como unidades lexicais (*largada, parada, subida, investigada*) e de infinitivos (*olhar, jantar, despertar, esperar*). As desinências comumente conhecidas como de infinitivo e de participio, neste caso, perdem o caráter flexional e passam a ser constituintes de deverbais de ação.

3.6. A RCP de deverbais de ação

Apresentados os mecanismos e estruturas dos deverbais de ação, definimos qual o paradigma dos deverbais de ação, ou seja, podemos descrever qual é a regra de formação de palavra para deverbais de ação. A RCP de deverbais de ação – permite a construção de nomes de ação, deverbais, cujos produtos são marcados como [+abstrato] e são construídos a partir dos processos de sufixação, conversão e desflexionação. Observemos o quadro que resume a Regra de Construção de Palavra para os deverbais de ação:

1- Sufixação = Tema verbal + conjunto de sufixos = deverbal de ação				
Ex.:	<i>lava-</i>	→	<i>-gem</i> <i>-nsia</i>	<i>lavagem</i> <i>lavansia</i>
			<i>-ção</i> <i>(-t/d)ura</i>	<i>lavação</i> <i>lavadura</i>
			<i>-mento</i> <i>-deira</i>	<i>lavamento</i> <i>lavadeira</i>
			<i>-dela</i> <i>-ria</i>	<i>lavadela</i> <i>lavaria</i>
			<i>-ncia</i>	<i>lavância</i>
			<i>-nço/nça</i>	<i>lavanço/a</i>
			<i>-ata</i>	<i>lavata</i>
			<i>-ame/ume</i>	<i>lavame/lavume</i>
2- Conversão = Radical verbal + desinência de gênero = deverbal de ação				
Ex.:	<i>lav-</i>	→	<i>-o</i>	<i>lavo</i>
			<i>-a</i>	<i>lava</i>
			<i>-e</i>	<i>lave</i>
3- Desflexionação = verbo → nome				
Ex.:	<i>lavar</i>			<i>lavar</i>
	<i>lavado</i>			<i>lavado</i>

1 - Dada uma base verbal X, ela pode combinar-se com o conjunto y (vide Tabela 5) de sufixos formando um deverbal;

2 - O deverbal de ação pode ser resultante do processo de conversão de bases verbais;

3- O deverbal de ação pode ser resultante do processo de desflexionação de formas verbais no infinitivo ou participio.

3.6.1.Considerações sobre RCP de deverbais de ação e de agente

O estabelecimento do número de RCPs para um dado produto não é consenso para os estudiosos do modelo como podemos ver nos trabalhos de G. M. Rio Torto (1992a) e CORREIA (2004). Existem propostas e adoções de determinados critérios que justificam o ponto de vista de cada pesquisador. Neste trabalho seguimos os critérios definidos por Correia (2004).

Tanto os deverbais de ação, quanto os deverbais de agente assemelham-se em relação a categoria da base – ambos são construídos a partir de verbos. Entretanto, unem-se a conjuntos de sufixos distintos. Nesse cenário, alguns pesquisadores trabalham com a ideia de tratá-los na mesma RCP. Contudo, apesar de termos uma RCP marcada categorialmente, segundo Correia (2004, p.162) uma solução desse tipo não garante identificar as especificidades dos diferentes processos morfológicos e semânticos que estão envolvidos. A diferença neste caso, vai além do conjunto de sufixos participantes. Apesar de ambos carregarem o sentido da base, os substantivos designativos são diferentes: concretos e abstratos. Portanto, teríamos sufixos, significados e subclasses distintas. Além disso, a RCP de deverbais de ação é muito mais produtiva que a de agente, por apresentar um conjunto de sufixos muito superior aos de formadores de agente. Nesse sentido a existência de duas RCPs respeita tais diferenças e acreditamos ajudar na compreensão das mesmas.

Após a investigação da área, definição dos deverbais em relação a aspectos formais, semânticos e sintáticos e compreensão e descrição dos processos de construção de deverbais de ação, iniciaremos a descrição do Projeto EXTRADEV que compreende as seguintes seções: Seleção do *corpus* (Seção 4.2); Sistemas de linguagens (Seção 4.3); Estudo Piloto (Seção 4.4); As regras (Seção 4.5); e a Avaliação (Seção 4.6).

4. EXTRADEV: um extrator de deverbais de ação históricos e suas variantes gráficas

A construção deste sistema motivou-se por não existir nenhuma ferramenta computacional que realize a recuperação automática de palavras de estrutura composta considerando todas as partes que a compõe e não apenas uma parte dela. A busca parcial (geralmente realizadas por sufixos ou prefixos) não garante a recuperação de uma categoria determinada. Ou seja, além de recuperar as categorias requeridas, muitas outras que apresentam estrutura morfológica semelhante podem ser recuperadas. Isso se dá devido a dois motivos: a estrutura morfológica desse tipo de palavra apresenta, muitas vezes, sufixos ou prefixos que correspondem a terminações de palavras que na verdade, são palavras simples. Outro fato que complica também a recuperação exclusiva do que se deseja é o caráter produtivo de alguns sufixos, como por exemplo *-gem* que uni-se tanto a bases verbais quanto a bases nominais.

4.1. O Projeto EXTRADEV

O sistema EXTRADEV consiste basicamente em um sistema de detecção e extração de deverbais de ação de maneira semi-automática. Esse sistema se beneficia de duas bases de conhecimento externas: o dicionário Unitex-PB (Muniz, 2004; Muniz et al, 2005) no formato DELAF que é usado pelo UNITEX (ver Seção 4.3.3) e a lista de variação de grafia gerada pelo sistema SIACONF (Giusti et al., 2007). A partir do dicionário DELAF, extraímos a lista de bases verbais para composição das regras; e a lista gerada pelo SIACONF foi determinante na identificação das variantes gráficas para os deverbais de ação presentes no *corpus*.

Apresentamos na Figura 4 a arquitetura do EXTRADEV em alto nível, compreendendo os módulos de geração de deverbais históricos com e sem variação de grafia, mais os conhecimentos utilizados e arquivos de entrada e saída. Nesta figura, são omitidos alguns processos de tratamento de texto, que são explicitados nas Figuras 6 e 7.

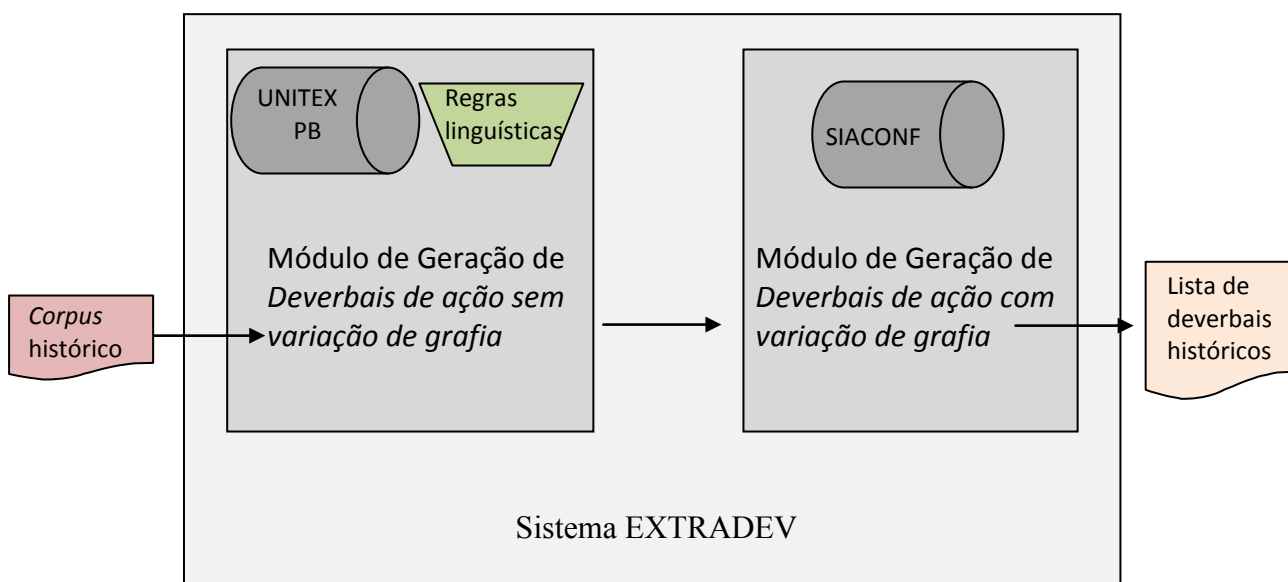


Figura 4: Visão geral do sistema EXTRADEV

- Legenda⁵⁹:
- Documento (lista)
 - Processo
 - Documento (*corpus*)
 - Regras manuais
 - Recurso/Sistema externo

Além dos recursos provindos do SIACONF e UNITEX-PB, foram necessários alguns tratamentos de textos, também chamados de pré-processamento de texto, que visam à adequação do formato do documento para que ele seja processado. Além dos *scripts*⁶⁰ e comandos utilizados para a preparação dos textos no formato adequado, há um *script* que foi denominado como o principal e que é responsável por executar as regras de detecção dos deverbais de ação. O *script* que contém a função de executar as regras faz parte do que chamamos de 1º módulo do sistema EXTRADEV que dividimos em duas partes – *Deverbais sem variação de grafia* e *Deverbais com variação de grafia*. Primeiramente, discorreremos sobre o primeiro módulo o qual contém 2 etapas. Observemos a arquitetura da primeira etapa na Figura 5.

⁵⁹ Esta legenda deve ser utilizada também para as Figuras 5 a 8.

⁶⁰ Conjunto de comandos e parâmetros escritos numa determinada linguagem de programação para a execução automática de tarefas. (<http://www.dicweb.com/ss.htm>)

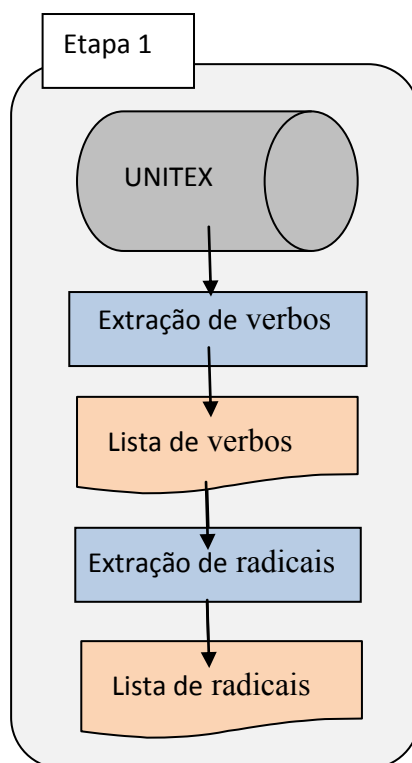


Figura 5: Arquitetura representativa da etapa 1 do módulo Deverbais de ação sem variação de grafia do sistema EXTRADEV

Etapa 1 – preparação da lista de radicais verbais. Essa etapa é dividida em duas fases; uma seleciona somente as linhas do dicionário cujas palavras são classificadas como verbos. Em seguida, apenas a coluna com a forma canônica do verbo é escrita em um documento txt. Essa lista, por sua vez, é submetida ao *script* que extrai apenas os radicais. Abaixo, são descritos, brevemente, os métodos utilizados para a obtenção das listas:

a) Extração de verbos: extrai somente os verbos do Dicionário DELAF.

Operação realizada via *comando de processamento de texto*⁶¹.

Algoritmo: extrai todas as linhas que contêm a etiqueta ‘V’ (grep ‘ V ’); imprime coluna com a forma canônica do verbo e escreve em um documento txt (awk ‘{print \$1}’ > lista_verbos.txt).

b) Extração de radicais: extrai as desinências verbais que indicam a qual conjugação o verbo pertence (-ar, -er, -ir).

Operação realizada via *script*.

Algoritmo: procura em cada palavra as terminações -ar, -er, -ir⁶² e substitui por nada.

Etapa 2 – Extração de deverbais de ação sem variação de grafia. O que chamamos de

⁶¹ Todos os comandos de processamento de texto serão descritos na Seção 4.2.2.

⁶² Também procuramos pela terminação -or porque, apesar de não pertencer a nenhuma conjugação, do ponto de vista computacional, essa é uma terminação que pode ocorrer nos verbos (ex.: por, decompor, etc..).

Etapa 2 é a parte principal do sistema EXTRADEV. É nesta fase que geramos todos os deverbais possíveis a partir dos radicais provindos da etapa anterior, juntamente com os constituintes linguísticos que participam da regra de construção dos deverbais de cada processo. Gerados os deverbais, eles são, então, buscados no *corpus* de entrada (neste caso, no *corpus* histórico do DHPB) e caso a forma exista, é escrita em um arquivo de saída .txt, constituindo uma lista de deverbais históricos sem variação de grafia. Em seguida, podemos observar a arquitetura representativa da etapa 2, na Figura 6.

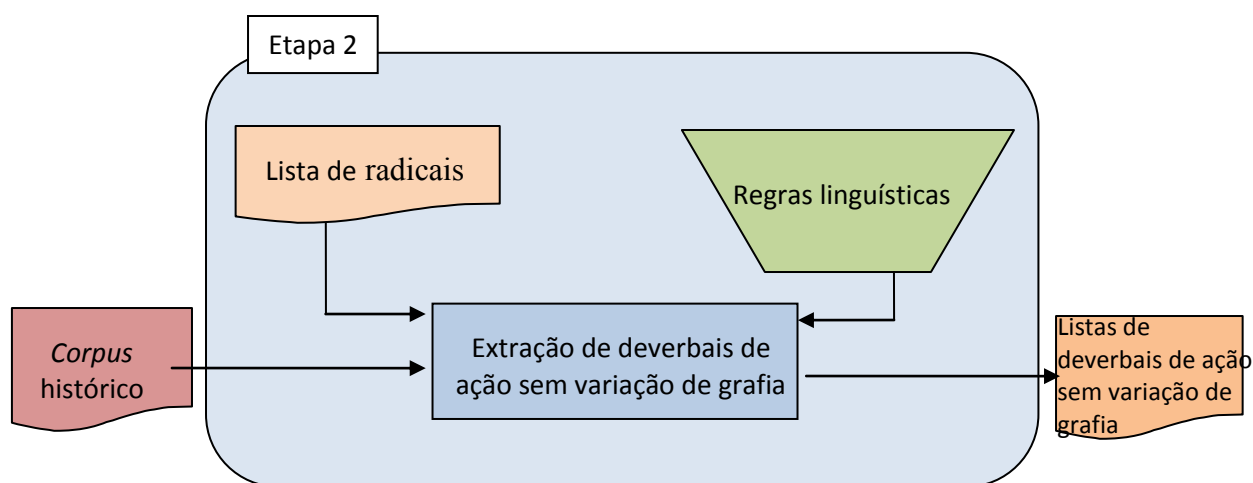


Figura 6: Arquitetura representativa da Etapa 2 do sistema EXTRADEV.

a) Extração de deverbais de ação: procura no *corpus* de entrada os deverbais de ação através das regras linguísticas no formato de expressão regular.

Operação realizada via *script*⁶³.

Algoritmo: concatena as variáveis de acordo com cada processo de construção de deverbais de ação (este processo é dirigido pelas regras elaboradas a partir do estudo dos deverbais segundo o modelo SILEX).

Com a lista de deverbais de ação sem variação de grafia gerada, termina o que chamamos de primeiro módulo do EXTRADEV. A saída desse primeiro módulo é a entrada do segundo módulo do sistema. A lista gerada na primeira fase foi confrontada com a lista de agrupamentos de variações de grafia geradas a partir do SIACONF (a lista do SIACONF tem um formato que precisou de um tratamento antes de ter sido confrontada com a lista de

⁶³ Para visualizar o script em python basta consultar o Apêndice A.

deverbais sem variação de grafia; ver Figura 7). E, então, para todos os deverbais de ação existentes no *corpus* foram recuperadas as variantes gráficas e, dessa maneira, a lista final aumentou consideravelmente⁶⁴. Na Figura 9, podemos observar a primeira parte – preparação de lista – da terceira etapa que pertence ao segundo módulo do EXTRADEV *Deverbais de ação com variação de grafia*.

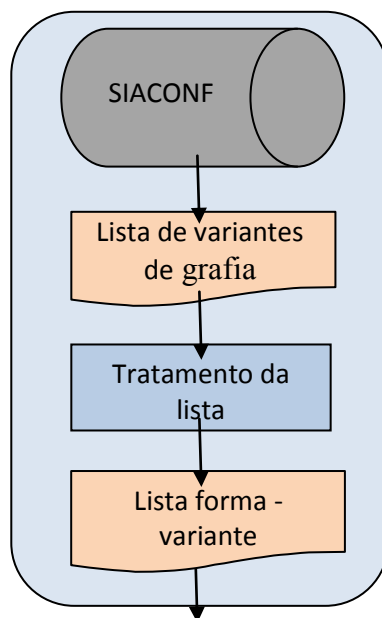


Figura 7: Arquitetura da etapa 3 do módulo de extração de variação de grafia do sistema EXTRADEV

Etapa 3 – Módulo extração dos deverbais de ação que apresentam variação de grafia

a) Tratamento da lista SIACONF: a lista é reformatada segundo o formato da lista resultante do módulo anterior.

Operação realizada via *script*.

b) Extração dos deverbais de ação com variação de grafia (ver figura 8): procura na lista SIACONF as respectivas variações dos deverbais extraídos no módulo I.

Algoritmo: se o deverbais da lista de entrada existe na lista do SIACONF, todas as palavras relacionadas a ele são escritas num arquivo .txt.

Vejamos a arquitetura na Figura 8.

⁶⁴ Os dados exatos são apresentados na Seção 4.5.

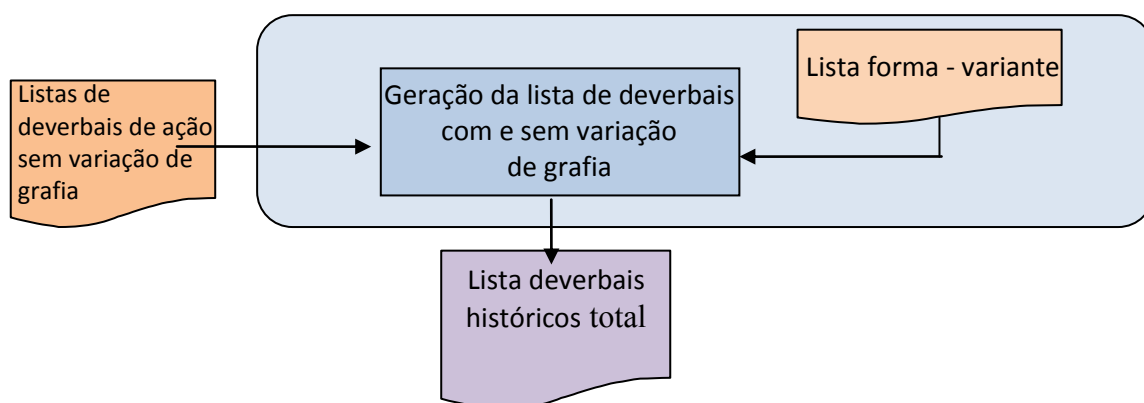


Figura 8: Arquitetura do módulo do sistema EXTRADEV de extração de verbais históricos.

O sistema EXTRADEV pode ser reutilizado para recuperação de outras palavras construídas tanto em textos do português histórico (utiliza sistema EXTRADEV completo), como em textos do português atual (utiliza sistema EXTRADEV módulo I). No presente momento, o sistema é indicado para categorias verbais, pois já existem mais de 15.000 bases verbais disponíveis sendo necessário apenas alterar as regras linguísticas de acordo com os processos de formação para a categoria almejada (excluindo ou adicionando processos de formação ou apenas mudando os sufixos como é o caso dos verbais de agente⁶⁵). Contudo, a reutilização do sistema EXTRADEV para outros cenários de pesquisa é um assunto para trabalhos futuros.

⁶⁵ Para a construção de verbais de agente bastaria excluir a conversão e a desflexionação (já que não são processos formadores de verbais de agente) e trocar os sufixos de ação (-*ção*, -*mento*, -*ncia*) pelos de agente (-*dor* - *vender* - *vendedor* -*dor*; *trabalhar* - *trabalhador*; e -*sta* - *equilibrar* - *equilibrista*; *manobrar* - *manobrista*).

4.2. Seleção do Corpus

O *corpus* utilizado para a presente pesquisa foi construído para o projeto Dicionário Histórico do Português Brasileiro dos séculos XVI, XVII e XVIII (DHPB), projeto desenvolvido no âmbito do projeto Institutos do Milênio, apoiado pelo CNPQ (Conselho Nacional de Desenvolvimento Científico e Tecnológico).

O *corpus* possui 2.458 textos⁶⁶, aproximadamente 7,5 milhões de formas simples⁶⁷ no formato de codificação UTF-16; esses textos foram escritos por brasileiros e portugueses (que viveram no Brasil por um longo tempo) no período de 1500 -1808. A Tabela 10 mostra a distribuição dos textos, sentenças, formas simples e tamanho de acordo com o século.

Tabela 9: Distribuição do corpus por século

Dados	XVI	XVII	XVIII	XIX
Textos (%)	6.24	26.39	59.78	7.59
Sentenças (%)	6.30	18.32	64.34	11.04
Formas simples (%)	7.60	20.18	62.57	9.65
Megabytes (%)	7.23	19.95	63.09	9.73

Há uma pequena parcela de textos do começo do Século XIX (chegada da família real no Brasil e início da imprensa brasileira): 123 textos, total de palavras 624.322 e 24.822 sentenças. O *corpus* contém, ainda, 188 textos de século desconhecido, perfazendo um total de 641.861 palavras e 25.837 sentenças. É necessário ressaltar que o número de sentenças é uma aproximação otimista para cima feita a partir do número de pontos finais “.”.

Podemos observar que a maioria dos textos concentra-se no século XVIII, devido a fatores de ordem de preservação (já que dentro o período analisado neste trabalho, o século XVIII é o mais próximo aos dias de hoje, logo a probabilidade de estarem em melhor estado de conservação é muito maior) e também da necessidade de documentação em decorrência das diversas mudanças e avanços que ocorreram ao longo desse século, e, não mais apenas como manuscritos de mera descrição destinados aos portugueses nos séculos anteriores, mas sim, de um novo país que se desenvolvia. Segundo Pinto (1988, p. 12):

⁶⁶ Todos os textos foram selecionados a partir de pesquisas bibliográficas em arquivos públicos e bibliotecas brasileiras e portuguesas.

⁶⁷ É o número total de palavras no corpus que são compostas de letras pertencentes ao alfabeto Português Histórico, criado no âmbito do projeto DHPB.

Foram eficazes e duradouros os efeitos da revolução pedagógica pombalina. Os estudos científicos passaram a dispor de laboratórios especializados, observatórios, hospitais e instituições auxiliares – hortos e museus responsáveis pela mudança de perspectiva de um ensino tipicamente mnemônico para outro que se fundamentava nas práticas experimentais. Tudo em consonância com o espírito do século: desprezar as especulações absurdas por um saber de cunho racionalista.

Além disso, é neste século que são descobertas jazidas de ouro, fundam-se mais colégios, os jesuítas são expulsos do Brasil, a sede do governo muda para o Rio de Janeiro, novos impostos são instituídos, etc.

Não existem muitas informações sobre gêneros nas fontes, apenas uma parcela muito pequena dos textos está classificada; são exemplos deles: cartas de missionários jesuítas, documentos da Inquisição, relatos de exploradores brasileiros, etc. O processo de identificação dos gêneros textuais é muito moroso e muitos textos trazem características de vários gêneros, diferindo assim da tipologia textual adotada pelo português contemporâneo.

Os textos provêm de manuscritos (manualmente digitados), documentos originais impressos e arquivos PDF composto de imagens (processados por OCR - *Optical Character Recognition* – Reconhecimento Óptico de Caractere).

4.3. Sistemas e Linguagens

4.3.1. Linguagem PYTHON

Segundo seu criador Guido van Rossum, Python é uma linguagem de programação poderosa e fácil de aprender. Ela possui estruturas de dados de alto nível e uma simples, mas eficiente, abordagem da programação orientada a objetos. Uma de suas grandes vantagens é a simplicidade de seu código que permite ao programador se ater mais à solução do problema, em vez de concentrar-se na linguagem propriamente dita.

Outra grande vantagem é que Python é um software livre. Isto significa distribuição de cópias, acesso ao código fonte - alteração, modificação e redistribuição - e utilização de pedaços dele em outros programas. Python é portátil para diversas arquiteturas, isto significa que um programa escrito para Linux pode facilmente rodar em Linux, Windows, etc.. Tudo isso devido ao Python ser uma linguagem interpretada, ou seja, não necessita ser compilada. O próprio interpretador lê e interpreta o código de maneira rápida, embora um código compilado possa ser mais rápido, em geral, do que um interpretado.

Python suporta tanto a programação orientada à função quanto a programação orientada a objetos. Em linguagens de programação orientadas à função o programa é construído em torno de funções que nada mais são que pedaços de código reutilizável. Em linguagens de programação orientadas a objetos o programa é construído em torno de objetos que combinam dados e funcionalidade. Python tem uma maneira poderosa, porém simples, de implementar a orientação a objetos, especialmente quando comparada a importantes linguagens como C++ ou Java.

A biblioteca padrão do Python é bastante vasta. Ela pode ajudar a fazer sistemas diversos como os que usam expressões regulares, banco de dados, navegadores, CGI⁶⁸, ftp⁶⁹, email, XML⁷⁰, XML-RPC⁷¹, HTML⁷², arquivos WAV⁷³, criptografia, GUI (interfaces

⁶⁸ CGI: Imagem em computação gráfica (www.wikipédia.org/wiki/CGI)

⁶⁹ ftp: Protocolo de Transferência de Arquivos (www.wikipédia.org/wiki/ftp)

⁷⁰ XML: Extensible Markup Language (www.wikipédia.org/wiki/XML)

⁷¹ RCP: Chamada Remota de Procedimento ou acrônimo de Remote Procedure Call (para mais informações consulte: <http://pt.wikipedia.org/wiki/XML-RPC>)

⁷² HTML: Linguagem de Marcação de Hipertexto (www.wikipédia.org/wiki/HTML)

⁷³ WAVEform audio format (www.wikipédia.org/wiki/Wav)

gráficas), entre outras coisas. Para a construção do script⁷⁴ de detecção e recuperação de deverbais utilizaremos o módulos *string*⁷⁵ e expressões regulares (*RE*⁷⁶) - os mais usados para tratamento de língua natural. O módulo RE é apresentado na Seção 4.3.1.1.

4.3.1.1. Módulo RE

Além de suportar qualquer sintaxe relacionada às expressões regulares, essa biblioteca possui métodos comuns ao módulo *strings* que estão listados na sequência:

<code>re.search(s, [,pos⁷⁷ [,endpos⁷⁸]])</code>	<code>#casa a re no começo da linha</code>
<code>re.match(s, [,pos [,endpos]])</code>	<code>#casa a re em qualquer lugar da linha</code>
<code>re.split(s)</code>	<code>#quebra no padrão desejado</code>
<code>re.findall</code>	<code>#encontra todos os padrões</code>
<code>re.sub(repl, s)</code>	<code>#substitui todos os padrões por repl</code>

Quando o padrão de busca é encontrado, um objeto é retornado com informações sobre a localização do padrão. Apesar das funcionalidades serem semelhantes, há algumas pequenas diferenças que podem ser determinantes para a escolha do método a ser usado. Por exemplo, quando não sabemos quantos padrões vamos encontrar, é interessante usarmos *findall*⁷⁹.

Observemos na Figura 9 um exemplo de como seria o código para a detecção de números de RG (o código está em negrito e em itálico os comentários sobre os passos do *script*):

⁷⁴ É uma descrição geral de qualquer programa escrito em linguagem interpretada, ou seja, não compilada.

⁷⁵ String é uma cadeia de caracteres ordenados segundo um conjunto pré-determinado. Por exemplo, uma palavra se encaixa nessa definição, portanto uma unidade lexical é uma string.

⁷⁶ Do inglês *Regular Expression*.

⁷⁷ Posição inicial para procurar padrão

⁷⁸ Posição final para procurar padrão

⁷⁹ Do inglês *find= encontrar e all= tudo*. Quando queremos **encontrar** padrões em textos, por exemplo, podemos utilizar "*findall*", já quando queremos **quebrar** palavras podemos utilizar "*split*", etc.

```

# importando o método re
import re

#guardando na variável 'pat' o padrão a ser detectado
pat = 'RG\s?[0-9]{2}\.[0-9]{3}\.[0-9]{3}(\-[0-9])?'

#compilando o padrão

compPat = re.compile(pat)

#procurando o padrão para cada linha no texto escolhido

for line in x:

findPat = compPat.findall(line)

#condição de existência do padrão – se o padrão é verdadeiro...

if findPat != []

```

Figura 9: Exemplo de um código em linguagem Python.

Para montar as expressões regulares⁸⁰ de acordo com o padrão requerido pelo programador, é necessário dominar alguns caracteres especiais. A Tabela 11 apresenta os principais metacaracteres.

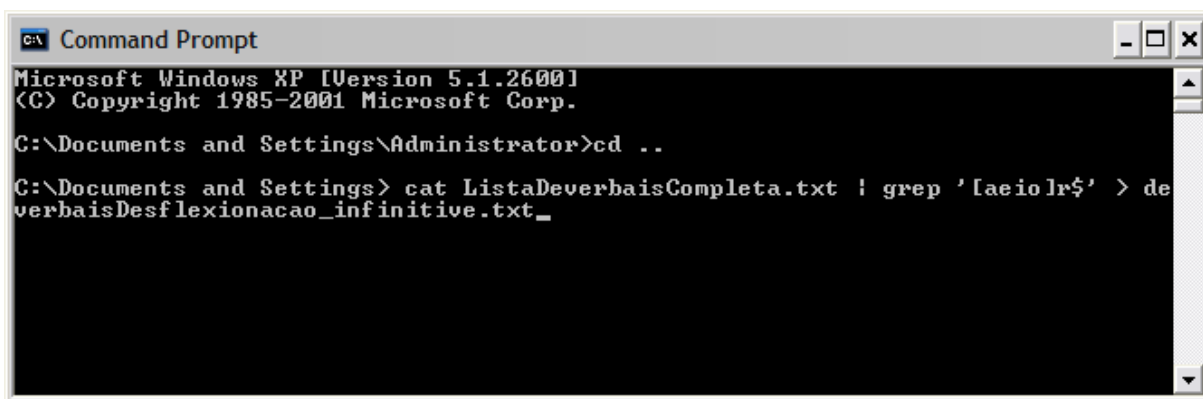
Tabela 10: Metacaracteres para expressões regulares.

Meta	Função
^	casa qualquer caractere no começo da linha
\$	casa qualquer caractere no final da linha
\w	qualquer letra
\d	qualquer dígito
\s	Espaço
[]	Lista
{}	quantificador mínimo e máximo
[^]	lista negada
()	grupo
	alternativo 'ou isso' 'ou aquilo'
?	quantificador: 1 vez ou nada
*	quantificador: nada ou infinitas vezes
+	quantificador: 1 vez ou infinitas
.	casa qualquer caractere

⁸⁰ Para compreender mais sobre expressões regulares recomendamos a leitura de Expressões Regulares – uma abordagem divertida (Jargas, 2006).

4.3.2. Comandos para processamento de texto

Chamamos de ferramentas alguns comandos disponíveis⁸¹ em sistemas operacionais Linux⁸². São mais simples que as outras linguagens de programação, porque não necessitam de toda a sintaxe requerida para criação de *scripts* e sobretudo, já são pré interpretados pelo sistema operacional Linux. Basta digitar os comandos na linha de comando. A Figura 10 mostra a interface de um *prompt*⁸³.



```
Microsoft Windows XP [Version 5.1.2600]
(C) Copyright 1985-2001 Microsoft Corp.

C:\Documents and Settings\Administrator>cd ..

C:\Documents and Settings> cat ListaDeverbaisCompleta.txt | grep '[aeio]r$' > de
verbaisDesflexionacao_infinite.txt_
```

Figura 10: Interface do prompt (também chamado linha de comando) com o comando `grep`.

Os comandos utilizados neste trabalho estão listados abaixo:

awk - (Alfred V. Aho, Peter J. Weinberger e Brian W. Kernighan)⁸⁴

awk é considerado uma linguagem de programação a parte. Não nos comprometemos em descrevê-lo, pois apenas utilizamos uma (que podemos ver no exemplo que segue) de suas muitas funcionalidades. Usamos o comando de extração de colunas.

Ex.: `cat verbosUnitex.txt | awk '{print$1}'`

grep – global regular expression print

Utilizado para buscar padrões através de expressões regulares.

ex.: `grep ' V' dicionárioUnitex.txt`

Com o comando acima é possível recuperar todas as linhas do documento que contenham a letra maiúscula ' V' antecedida de espaço.

⁸¹ Também podem ser utilizados em windows, basta instalar as bibliotecas.

⁸² Sistema operacional, multitarefa, para computadores pessoais (PCs) desenvolvido pelo finlandês Linus Torvalds, em 1991. Disponibilizado gratuitamente na Internet, passou a receber a colaboração de outros programadores, o que tem contribuído para seu aperfeiçoamento. (<http://www.dicweb.com/pp.htm>)

⁸³ Ing.] (Pronto, preparado). 1. Em sistemas operacionais baseados em comando, um ou mais símbolos, que indicam o local, a partir do qual, o usuário deve digitar uma instrução. No MS-DOS, por exemplo, é uma letra que representa a unidade de disco, seguida do sinal “maior que”, ex. C>. 2. Mensagem breve que surge na tela do computador, indicando que um programa aguarda a entrada de dados por parte do usuário. (<http://www.dicweb.com/pp.htm>)

⁸⁴ O nome awk vem das iniciais dos seus criadores.

wc – word count (conta palavra)

Utilizado para realizar contagem de linhas ou palavras.

ex.: `cat listaDeverbais.txt | wc`

cat – cat (pegar)

Este comando sempre deve ser utilizado para exibir determinado arquivo no prompt ou quando se deseja utilizá-lo com outro comando.

ex.: `cat corpus_utf8_100.txt`

sort – sort (ordenar)

é capaz de ordenar arquivos e palavras por ordem alfabética ou por frequência.

ex.: `cat verbosUnitex.txt | sort >`

tr – replace

com essas duas letrinhas podemos transformas textos em listas, palavras em nada, substituir letras maiúsculas por minúsculas.

ex.: `cat corpusDeverbal.txt | tr ‘ ’ ‘\n’`

uniq – elimina palavras, símbolos repetidos ou espaços em branco.

ex.: `cat corpusDeverbal.txt | sort | uniq -c | sort -nr >`

Esses comandos abarcam funcionalidades bastante úteis para linguístas. São extremamente rápidos e não exigem conhecimento especializado para sua utilização.

4.3.3. UNITEX

No desenvolvimento da parte computacional foi utilizada a ferramenta computacional UNITEX, a qual contém uma série de recursos para tratamento de *corpus* de língua escrita. Essa ferramenta consiste em um conjunto de programas para processamento de *corpus* linguístico composto por uma interface gráfica em Java e diversos programas desenvolvidos em C. Dentre os recursos oferecidos estão um concordanciador, um contador de frequência, grafos e dicionários de línguas que quando aplicado a um *corpus* geram três listas: desconhecidas do dicionário aplicado, lexias simples e lexias complexas. A versão 1.2 da ferramenta provê suporte para mais de 14 idiomas (incluindo o Português); atualmente foi lançada uma nova versão do programa, a versão 2.0 beta que está sendo usada neste projeto. Entretanto, pode-se adicionar facilmente suporte a qualquer idioma graças ao uso do padrão Unicode (Unicode, 2003) para codificação de texto. O suporte ao idioma português contemporâneo é particularmente bom graças ao trabalho UNITEX-PB desenvolvido em um mestrado do ICMC (Muniz, 2004; Muniz et al, 2005).

O software UNITEX é uma implementação livre do programa Intex, ambos criados no laboratório francês LADL (Laboratoire d'Automatique Documentaire et Linguistique). Desta

forma, as funcionalidades fornecidas por estas ferramentas são bem semelhantes. Os dicionários UNITEX se baseiam no formalismo DELA (Dictionnaire Electronique du LADL) também desenvolvido no laboratório LADL.

Um arquivo de texto não formatado (formato txt) com codificação Unicode é convertido para uma forma pré-processada após ser aberto pela primeira vez no Unitex. Os arquivos pré-processados geralmente possuem a extensão “.snt”. Durante o processo de conversão, o texto original é segmentado em sentenças e unidades lexicais (*tokens*). Além disso, repetições desnecessárias de caracteres de separação, tais como espaços, quebras de linha e tabulações, são removidas e formas não ambíguas do texto são normalizadas para simplificar operações de busca, sendo que as normalizações são definidas pelo usuário. Como exemplo, a palavra “daí” é normalizada em “de aí”. É importante notar que normalização não pode ocorrer para palavras ambíguas tais como a forma gráfica “desse” que pode significar “de esse” ou uma conjugação do verbo “dar”.

Para este trabalho, somente utilizaremos a *Token List*, por apresentar a frequência de cada *token* contido no *corpus*, independentemente de serem reconhecidos ou não pelo dicionário da ferramenta. Por ser este um *corpus* histórico, nem todas as unidades lexicais coincidem com o português contemporâneo.

4.3.4. SIACONF

SIACONF (Sistema de Apoio a Contagem de Frequência em Corpus) (Giusti et al., 2007) é um sistema de tratamento de variação de grafia, desenvolvido no escopo do projeto DHPB, e baseado na metodologia do trabalho de Hirohashi (2004) e Menegatti (2002).

O sistema SIACONF aplica uma série de regras de transformação para uma lista de palavras extraídas de um corpus histórico, visando, principalmente, agrupar diferentes variações de grafia através dessa técnica que é capaz de estabelecer uma relação entre diferentes grafias. A expectativa é que essa relação seja capaz de mostrar as possíveis variações de grafia para qualquer palavra num determinado *corpus*.

O sistema SIACONF processa um *corpus* a partir de uma lista de regras iniciais que oferece três tipos de detalhamento: (a) agrupamentos que incluem variação de grafia da mesma palavra; (b) informações das regras aplicadas; (c) e uma lista das palavras não processadas.

O agrupamento utilizado no sistema SIACONF difere das técnicas de normalização de Hirohashi (2005). O SIACONF não tenta encontrar uma ortografia atual/moderna de uma variante no *corpus*, entretanto isso ocorre na maioria das vezes. Por exemplo, as palavras “chaõ” e “chaão” são agrupadas junto à variação “xam”, a qual não existe mais no português brasileiro contemporâneo. Por exemplo, usando as regras de transformação do SIACONF, as variantes de “chão” foram encontradas: chaõ, xão, cham, chaão, and xam. A Tabela 12 mostra exemplos de variantes detectadas para as palavras “apelido”, “mais”, “não” e “vila” acompanhadas de suas respectivas frequências no *corpus*. A primeira linha de cada exemplo contém a grafia obtida pelas regras de transformação e a soma das frequências de todas as suas variantes.

Tabela 11: Variantes detectadas para as palavras “apelido”, “mais”, “não” e “vila”

apelido (90)	nam (37,100)
appellido (48)	não (33,684)
apelido (30)	naõ (2,652)
appelido (7)	nam (439)
apellido (5)	nao (325)"
mais (23053)	vila (5,218)
mais (22,918)	villa (4,073)
majs (67)	vila (1,113)
maes (38)	vyla (13)
mays (30)	vjlla (9)
	vylla (9)
	vjla (1)

O objetivo do SIACONF é que os agrupamentos reduzam o impacto da variação de grafia na contagem da frequência e que o agrupamento permita o estudo da variação de grafia no corpus compilado. Desse modo, todas as variantes são encontradas e agrupadas facilitando tanto situações de normalização para a obtenção de uma frequência correta das palavras, quanto fonte de pesquisa das diversidade de grafias de uma mesma palavra.

Regras de transformação adotadas nessa técnica usam expressões regulares. A regra de transformação é uma tripla (C1,C2,S), na qual C1 e C2 são expressões regulares e S é uma string. C1 determina o critério de cobertura da regra, i.e., das formas que serão processadas pela regra. C2 determina a *substring* em cada W_i , o qual será substituído por S. Por exemplo, a regra “(e[ao] e ei)” é aplicada como segue:

1. C1 é testada em todas as formas do *corpus* e restringe a aplicação da regra para aqueles que contêm a *substring* “ea” ou “eo”, por exemplo aldea. Ou seja, C1 armazena a

variação de grafia a ser encontrada na palavra, neste caso, todas as palavras que contenham as *substrings* *ea* e *eo*.

2. C2 determina a *substring* que será substituída, por exemplo, a letra *e* em *aldeia*.

3. S determina a substituição *ei*, usada para generalizar a nova forma, por exemplo: *aldeia*.

Depois da aplicação de diferentes regras, várias grafias G_i produzem uma nova grafia H . Assim é possível inferir que grafias G_i são variantes da mesma palavra. Por exemplo, as regras (ll, ll, l) e (y y l) podem ser aplicadas para as grafias “vyla” e “villa”, respectivamente, resultando uma nova forma de grafar “vila”. Assim, eles são provavelmente variantes da mesma palavra. Além disso, podemos observar que mais de uma regra pode ser aplicada a uma forma dada, como é mostrado na Tabela 13.

Tabela 12: Agrupamento de “não” e “naõ” em torno da forma “nam”

Palavras	Regras aplicadas	Variantes geradas
NAÕ	[óôôôô] . o [[^] r][aã]o\$ [aã]o am	"nao" "nam"
NÃO	[[^] r][aã]o\$ [aã]o am	"nam"

Durante esse processo, todas as regras são aplicadas às formas únicas no *corpus*, gerando um conjunto de novas formas H_i . Cada nova forma representa um agrupamento de variações da mesma forma. Atualmente, são usadas 51 regras de transformações, as quais podem ser divididas em seis grupos:

- Regras para formas que caíram em desuso. Por exemplo: substituição de “y” por “i”. “Y” e “i” o som é o mesmo em português. Entretanto, “y” foi substituído por “i” em todas as palavras exceto nas palavras estrangeiras e nomes próprios. Outras regras são mostradas na Tabela 14.

Tabela 13: Regras do SIACONF para formas que caíram em desuso.

ee ee é	[áàää] . a	[ýÿÿÿ] . y	^ha ha a
ph ph f	[éèêê] . e	gu[ao] gu g	^he he e
pt pt t	[îïî] . i	dh dh d	^hi hi i
th th t	[óòôô] . o	v\$ v u	^ho ho o
ff s	[úùûû] . u	[^r][aã]o\$ [aã]o	^hu hu u
g[ei] g j		am	

• Regras para consoantes duplas. Por exemplo: substituição de “ff” por “f”. Outras regras são mostradas na Tabela 15.

Tabela 14: Regras SIACONF para consoantes duplas.

pp pp p	mm mm m	gg gg g	ll ll l
tt tt t	bb bb b	vv vv v	uu uu u
nn nn n	dd dd d	zz zz z	cc cc c

• Regras geradas de acordo com a norma ortográfica. No português, o sons de “m” e “n” são os mesmos quando precedem uma consoante. Contudo, “m” precede somente antes de “b” e “p”, e “n” precede todas outras consoantes. Elas são mostradas na Tabela 16.

Tabela 15: Regras do SIACONF geradas de acordo com a norma ortográfica.

j[bcdfghklmnpqrstvwxyz] j i	mpt mpt nt
m[cd fghjklqrstwxz] m n	n[pb] n m
mn mn n	ct ct t
mpt mp n	

- Regras baseadas na frequência, formuladas para tratar padrões recorrentes de variação de grafia. Por exemplo: substituição de “chr” por “cr”, como em Christo (Christ). Outras regras são mostradas na Tabela 17.

Tabela 16: Regras do SIACONF baseadas na frequência.

ch ch x	.acem\$ c ss	aes\$ aes ais
---------	--------------	---------------

- Regras lexicalizadas: regras para específicas palavras. Por exemplo substituição de “o” por “u” em “Deos” .

- Regras automáticas, baseadas nos estudos de Hirohashi (2005) de técnicas de aprendizado automático no corpus Tycho-Brahe . Não é possível usar as mesmas técnicas no HDBP porque o corpus HDBP não tem o mesmo nível de anotação que o corpus Tycho-Brahe. Um exemplo é a substituição de “z” por “s” no sufixo “zente”, como em “presente”. Outras regras são mostradas na Tabela 18.

Tabela 17: Regras do SIACONF baseadas em aprendizado automático .

ozo\$ z s	serviss serviss service	preciz preciz precis
-----------	-------------------------	----------------------

Depois da aplicação dessas regras no corpus, foram identificadas 76.754 formas variantes em 31.069 agrupamentos de palavras. O relatório das palavras não processadas geradas pelo SIACONF é útil para desenvolver novas regras. Nesse relatório é possível encontrar palavras com alta frequência no corpus que não são agrupadas em nenhuma regra.

Após reportar todas as ferramentas e linguagens aqui utilizadas, procederemos ao relato do estudo piloto realizado com vistas a avaliar a dificuldade e alcançabilidade de construir o recurso computacional. Abaixo descrevemos todas as etapas pelas quais passamos até chegarmos à proposta final: o sistema EXTRADEV.

4.4. Estudo Piloto

Um dos objetivos deste trabalho foi a construção de um sistema computacional que realizasse a extração de deverbais de ação automaticamente de um dado *corpus*. A idéia

inicial vislumbrava o estudo do padrão de construção de 50 verbos apenas – os mais frequentes do *corpus* – criar regras hipergerativas para detecção dos deverbais de ação baseadas nos seus processos de formação e obter todos os deverbais de ação para essa amostra de verbos, considerando suas variantes. Contudo, após esse estudo piloto, o trabalho cresceu e sofreu algumas modificações. Apesar das mudanças, o estudo piloto foi um bom exercício para chegar ao sistema EXTRADEV. Nesse sentido, reporto abaixo os procedimentos necessários para a construção desse piloto.

I -Seleção dos 50 verbos mais frequentes no *corpus* com auxílio da *Token List* da ferramenta UNITEX;

II -Extração de todas as ocorrências (candidatas a deverbais) nominais com o auxílio também da *Token List* para uma amostra de 5 verbos. Essa etapa visava conhecer as possíveis variações de grafia que poderíamos encontrar nas formas deverbais.

III -Construção das regras de acordo com os processos de formação de deverbais segundo o modelo SILEX;

IV -Construção do script em Python para a geração dos deverbais (ver Apêndice A);
Nas próximas seções, segue a descrição de todas as atividades supracitadas.

4.4.1. Etapa I: Seleção dos 50 verbos

O critério para seleção dos verbos foi baseado na frequência. Os 50 primeiros foram extraídos com a ajuda da ferramenta de processamento de texto Unitex. Como já descrito na Seção 3.3, o Unitex oferece uma lista chamada *Token List* que elenca todos os *tokens* (espaços em branco, sinais de pontuação e unidades lexicais) presentes no *corpus*, justapostos ao número de vezes que ocorrem.

O usuário pode visualizar a *Token List* por ordem alfabética ou frequência. No nosso caso, como nos interessa os verbos mais recorrentes no *corpus*, adotamos o critério de frequência. Entretanto, não basta apenas olhar os verbos que estão no topo da lista, pois estes aparecem exatamente como estão nos textos, ou seja, as formas verbais podem estar flexionadas (*sou, era, serei, etc.*) ou não (*ser*). Não há nenhum tipo de automatização que realiza a contagem das formas flexionadas e as relaciona com o lema. Em decorrência disso, a tarefa manual torna-se mais morosa.

A primeira aparição do verbo *ser* é na forma *é* apresentando 17.328 instâncias; em seguida, temos *são* com 11.844 instâncias; *ser* com 11.440 instâncias; e assim por diante. Para

obtermos a frequência total teríamos que somar cada nova forma encontrada na lista, considerando todos os tempos, pessoas e modos; essa tarefa exigiria um gasto de tempo muito grande. A Figura 11 mostra a *Token List*, ordenada segundo critério de frequência, destacando as ocorrências das várias formas flexionadas do verbo ser: *é*, *são* e *ser* respectivamente.

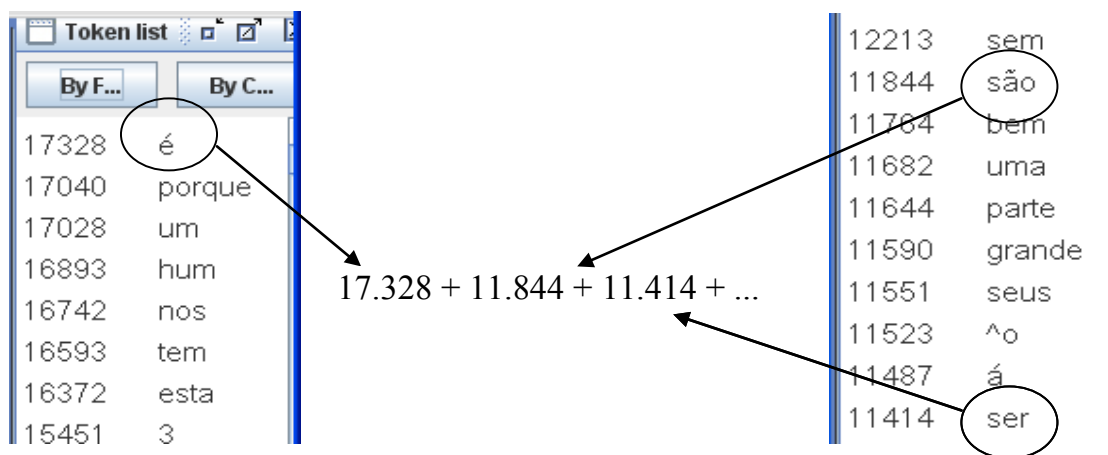


Figura 11: Token List e a frequência das formas do verbo ser

Desse modo, recorreremos a outro recurso da ferramenta Unitex que possibilita a obtenção do total de formas encontradas no *corpus* para um único verbo. Utilizamos, então, o sistema de busca com o uso da seguinte expressão: <ser.V>. Basta digitarmos o verbo no infinitivo seguido de ponto, seguido da letra V (de verbo). A Figura 12 mostra a interface de busca *Local Pattern* e a janela com os resultados da busca.

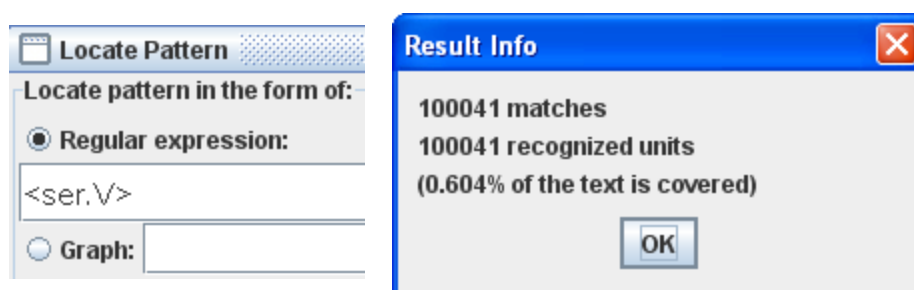


Figura 12: Interface de busca e a janela dos resultados obtidos a partir da expressão <ser.V>

O número de ocorrências pode ser um valor aproximado, já que muitos verbos também podem ser classificados como nomes, ou seja, a mesma unidade lexical pode pertencer a categorias gramaticais distintas. Podemos citar o verbo de maior ocorrência no *corpus* – o verbo *ser* com 100.041 instâncias. Esse verbo assume função de verbo de ligação (*Joana é doente*), verbo auxiliar (*A janta é feita com muito amor*), e, muitas vezes, como o verbo principal de uma locução verbal (*Maria queria ser a melhor secretária da empresa*) e

essa mesma forma também pode representar o substantivo *ser* (Aquele *ser que não é capaz de amar, não sabe viver*). Neste último caso, ocorre a desflexionação do verbo *ser* que assume a forma nominal. É possível encontrar para cada verbo uma forma desflexionada, não só a partir de verbos no infinitivo, como nesse caso, mas também no particípio como vimos no Capítulo 3; devido a esse fato, aumenta o percentual de ruído em nossos dados. Nas Figuras 13 e 14, podemos observar dois exemplos do verbo *ser* na forma verbal e nominal encontrados no *corpus*.

Todo o mais tempo gastey en confessar e algumas vezes pregar; isto fiz forçando-me muyto, por a intenção do Padre **ser** esta, e de Ramos para cá o faço. (...) {324 P. ANTÓNIO PIRES - PP. E II. DE COIMBRA - A00_0006.txt,N}

Figura 13: A forma ser como verbo

A razão natural d'este dito é fundada n'aquella certa philosophia, com que disse Aristoteles, que aos pais ninguem póde pagar o que deve, porque lhes devemos o ser e a vida. {A00_1056 - 254 PADRE ANTONIO VIEIRA,N}

Figura 14: A forma ser como substantivo

No primeiro excerto a forma *ser* aparece como verbo no infinitivo; e, no segundo, aparece na forma substantiva, ou podemos dizer também, na forma de um deverbal de ação.

Dentre as 11.665 ocorrências do verbo *ser*, podemos encontrar formas *ser* que não são verbos. Para sabermos exatamente qual o número de formas nominais, seria necessária uma desambiguação a partir da análise do contexto. A ferramenta Unitex oferece o concordanciador para análise de contexto. Contudo, tal desambiguação categorial não será tratada aqui, pois a obtenção do número aproximado já é suficiente para elegermos os mais frequentes, além disso, sabemos que a desambiguação não é um processo simples - principalmente por lidarmos com um *corpus* histórico - e poderia ser tema de outro trabalho.

Nesse cenário, após a verificação dos verbos mais frequentes, nas formas flexionadas ou não, usamos o buscador e ordenamos, de acordo com a frequência, todos os 50 verbos como está explicitado na Tabela 19.

Tabela 18: Os 50 verbos mais frequentes no corpus total

1- SER = 100.041	26- PARECER = 6.484
2- TER = 50.878	27- MATAR = 5.959
3- FAZER = 43.898	28- FALTAR = 4.947

4- DIZER= 43.840	29- DEIXAR = 5.930
5- COMER = 41.594	30- SERVIR = 5.744
6- IR = 35.164	31- TIRAR= 5.501
7- DAR = 28.272	32- LEVAR= 5.171
8- ESTAR = 23.049	33- FALTAR = 4.947
9- HAVER= 21.616	34-PEDIR= 4.787
10- PODER = 20.096	35- PAGAR = 4.354
11- VER = 18.886	36- MORRER = 4.203
12- PARTIR = 18.319	37-SEGUIR= 3.846
13- MANDAR = 13.370	38- ESCREVER = 3.797
14- ENTRAR= 13.097	39 -VIVER= 3.736
15- FICAR = 12.455	40- CONHECER= 3.532
16- VIR= 12.084	41- MOSTRAR = 3.448
17- ACHAR = 10.371	42- TRABALHAR =3.392
18- PASSAR= 9.421	43 -PERGUNTAR=3.220
19- CHEGAR = 8.870	44 -REFERIR = 3.208
20- PÔR = 8.719	45 - CORRER = 3.165
21- DEVER = 8.495	46 -OBRIGAR=3061
22- CHAMAR = 8.140	47 -PERDER=2.930
23- QUERER = 7.651	48 -GUARDAR= 2.858
24- SABER = 7.466	49 -ESPERAR=2.533
25- TOMAR = 7.209	50 -CONSTAR = 2.472

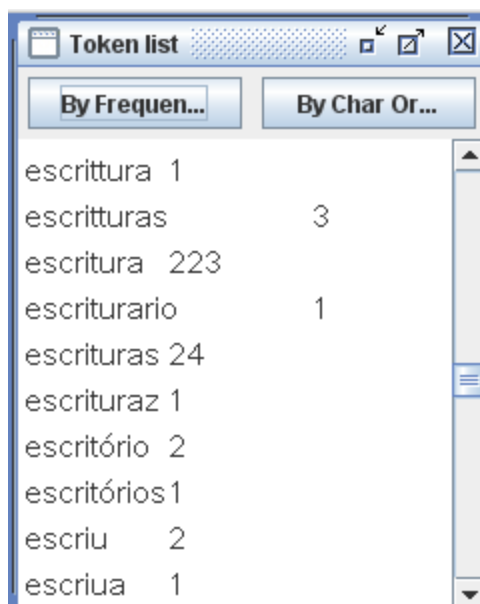
O verbo de maior ocorrência no *corpus* é o *ser* (100.041) com o dobro da frequência do segundo verbo mais recorrente (*ter*) e, quase 50 vezes superior ao último da lista (*constar*); cerca de 80% dos verbos tem frequência abaixo de 20.000; são 10 os que apresentam frequência superior a 20.000; a priori, a alguns deles, não servem de base para estrutura deverbal por conversão e sufixação devido as suas características morfológicas e fonéticas.

4.4.2. Etapa II: Extração das variantes gráficas de deverbais de ação

A principal característica do português histórico é a variabilidade na escrita das palavras, o que multiplica a possibilidade de se escrever a mesma palavra. Nesse contexto, é necessário conhecer as possibilidades de grafia mais comuns para que não percamos ocorrências dos possíveis deverbais de ação⁸⁵.

Novamente, recorremos à *Token List*, porém, nessa etapa, utilizamos a ordenação alfabética. Nosso objetivo foi extrair uma amostra dos deverbais construídos a partir dos verbos mostrados na Tabela 19 com vistas a conhecer as possíveis variações de grafia que essa categoria apresenta, influenciada pelas bases e sufixos (quando existem) característicos da RCP para deverbais de ação.

A Figura 15 mostra uma parte da *Token List* com os deverbais de ação, formados a partir dos verbos eleitos.



The screenshot shows a window titled "Token list" with two buttons: "By Frequen..." and "By Char Or...". The list contains the following items:

escrittura	1
escritturas	3
escritura	223
escriturario	1
escrituras	24
escrituraz	1
escritório	2
escritórios	1
escriu	2
escriua	1

⁸⁵ A decisão de se utilizar o SIACONF para este projeto surgiu após o estudo piloto.

Figura 15: Parte da Token List ordenada alfabeticamente.

Outra característica relevante da *Token List* para extração das variantes é o fato de conter todas as ocorrências de unidades lexicais do *corpus*, mesmo as consideradas agramaticais. O que nos permite a detecção de palavras grafadas de maneiras distintas em relação ao português contemporâneo. Dessa maneira, a lista, além de garantir que não percamos nenhuma ocorrência, é reveladora de maneiras de grafar as palavras possivelmente inimagináveis por falantes do português.

A fim de avaliarmos as possíveis formas de variação gráfica, escolhemos cinco verbos e extraímos todos os deverbais de ação encontrados na lista. São eles: *escrever*, *mandar*, *chegar*, *partir e fazer*.

A Figura 16 contém todas as formas nominais encontradas no *corpus* para o verbo escrever.

Tabela 19: Formas nominais encontradas no corpus DHPB para o verbo escrever

escrip^to	Escriptu	escripturaçaõ	Escrita
escrip^tra	Escriptura	escripturaçoens	Escrittura
Escripta	escripturado	escripturação	Escritturas
Escriptas	escripturados	efcripta	Escritura
Escriptos	Escripturas	efcripto	Escrituras
Escriptr	escripturaçam	efcrittura	Escrituraz
Escltr	Escritoz	efcritura	efcritores
Escritto	efcritos	Escríptas	efcriptura
Escritas	efcritto	Escríptos	efcrita
Escritaz	efcrittos	efcrittena	efcritas
Escrito	efcrituras	efcrittenas	efcrito
Escritos	Escrjtas		

A partir dessa lista geral, dividimos em três partes os deverbais de ação para a análise de variação de grafia: bases comuns, vogais temáticas e sufixos (quando usados) para, também, futuramente, podermos criar a expressão regular adequada. A Figura 17 mostra as variações de grafia possíveis para bases.

Variante gráfica	base
-t- ou -tt-	escrit- escritt-
-s- ou -f-	eScrit- efcrit-
inserção de -p- antes de -t-	escript-
acento agudo na letra -i-	escrípt-
-i- ou -j- ou -y- :	escrijt-

Figura 16: Variações de grafia possíveis de bases para o verbo escrever

As variantes para sufixo serão definidas somente no final do levantamento, pois esse é um grupo fixo, ou seja, para o processo de sufixação a variável **sufixo** guardará sempre o mesmo conjunto. Em seguida procederemos da mesma forma para o verbo *mandar* (Tabela 20).

Tabela 20: Formas nominais do verbo mandar

Mandado	Mandar	Manda	Mamdar
Mandamento	Mando	Mamda	Mandada
Mandato	Mamdo	mamdado	Mandata

As variações possíveis para base no verbo *mandar* são mostradas na Tabela 21.

Tabela 21: Variações de grafia presentes no verbo mandar

Variante gráfica	base
-m- ou -n-	mand- mamd-
-d- ou -dd-	mandd- mamdd-

Não há alterações na grafia da vogal temática -a-.

As variações possíveis para a base no verbo *chegar* são mostradas na Tabela 22.

Tabela 22: Variações de grafia presentes no verbo chegar

Chegar	chegada	chegar	chegádos
Chega	chegado	chego	Xegado
Xegar			

Não há variação de grafia na vogal temática *-a-*.

As variações possíveis para a base do verbo *partir* são mostradas nas Tabelas 23 a 25.

Tabela 23: Formas nominais do verbo partir

Parta	particam	Partir	parte
Parteda	parti	partimento	Parto
Partição	partj	Partte	

Tabela 24: Variações de grafia presentes no verbo partir

Variante gráfica	base
-t- ou -tt-	Part- partt-

Tabela 25: Variações de grafia presentes na vogal temática da 3ª conjugação

Variante gráfica	Vt
-i- ou -j- ou -y- ou -e-	Parti- party- parte- partj-

As variações possíveis para a base do verbo fazer são mostradas nas Tabelas 26 a 28.

Tabela 26: Formas nominais do verbos fazer

Fazer	fazêr	Fazimento	fazenda°
fasendo°	fafenda	partimento	faze
Fazer			

Tabela 27: Variações de grafia presentes no verbo fazer

Variante gráfica	base
-s- ou -z-	fas- faz-
-t- ou -tt- e -i- ou -j- ou -y-	feit- feitt- feyt- fejt-

Tabela 28: Variações de grafia presentes na vogal temática da 2ª conjugação

Variante gráfica	Vt
-i- ou -j- ou -y-	Fasi- fazy- fasj-

A Tabela 29 mostra o levantamento de variantes de grafia para os sufixos que participam do processo de construção dos deverbais de ação.

Tabela 29: Variações de grafia presentes nos sufixos formadores de deverbais de ação.

Sufixos	Variações possíveis
<i>-mento</i>	<i>-mento -memto -mmento</i>
<i>-ção</i>	<i>ção- são- ção- cãõ- caõ-</i>
<i>-ura/-dura/-tura</i>	<i>-ttura</i>
<i>-agem</i>	<i>-ajem -agem</i>
<i>-dela</i>	<i>-della -ddela -dela</i>
-ança	<i>-ansa -anca -amsa -amça -anza - amza</i>
<i>-ência</i>	<i>-encia -ênsia -ensia -ênçia -ençia -emcia -emçia -êmcia, etc.</i>

Após esta investigação podemos deduzir algumas possíveis variações em determinadas letras e determinados contextos. Abaixo, segue uma lista com as considerações de grafia que serão levadas em conta na busca dos deverbais de ação:

- Quando -s- → [s,z,ç,c, f]
- Quando -n- → [m,n]
- Quando -t- → [tt,pt,t]
- Quando -i- → [i,í,y,j,e]
- Quando -u- → [u,v]
- Quando -d- → dd

Desse modo, acrescentamos em nosso conjunto de regras a possível duplicação para outras consoantes como o -cc-, -bb-, -gg-, -vv-, -uu-, -zz-, apesar de não termos encontrado em nossa amostra a duplicação de tais consoantes.

As vogais temáticas são três: *a*, *e*, *i* ou *o*. Abaixo, seguem as possíveis variações de grafia para as vogais temáticas:

quando -a- → [á,a]
 quando -e- → [e,i,y,j,í]
 quando -i- → [e,i,í,j,y]
 quando -o- → [o,u]

Abaixo, como forma de mostrar como as variações de grafia podem ser encaixadas na regra, alguns exemplo:

Base + sufixo – deverbais de ação construídos por sufixação

fa[z,s, f,-	+	[e,i,í,y,j]	+	sufixo (?)
fe[i,y,j,í][tt,pt,t]-	+	[a,á]?	+	sufixo(?)

Base + vt - deverbais de ação construídos por conversão:

fe[i,y,j,í][tt,pt,t]-	+	[a,á,o,u]
-----------------------	---	-----------

4.5. As regras

A partir do estudo da estrutura dos deverbais de ação, de seus processos construcionais – segundo a morfologia construcional –, juntamente com o conhecimento de expressões regulares, tornou-se possível a elaboração de regras de recuperação destes deverbais.

Por exemplo, no processo de sufixação participam a base, a vogal temática – que pode ter três formas dependendo da conjugação a que pertence – mais o sufixo, portanto, uma variável para cada constituinte (base + vt_1|vt_2| vt_3 + sufixo). Já as regras do processo de conversão requerem apenas duas variáveis: uma para armazenar a base e a outra desinência de gênero (base + dg). O processo de desflexionação, por ter como *entrada* tanto a forma infinitiva quanto a forma do particípio passado exige variáveis que armazenem as três conjugações que são marcadas na desinência final da forma infinitiva e as formas finais do particípio passado que são *-ado*, *-ido* (base + conj_1| conj_2|conj_3 e base + part_1| part_2). Portanto, cinco variáveis, além da base.

É necessário reforçar aqui que em relação aos particípios irregulares, salvo *escrit-* de *escrever* e *dit-* de *dizer*⁸⁶, foram excluídas de nossas regras quaisquer formas que apresentassem alteração da representação fonológica da base e/ ou ausência da vogal temática por razões de viabilizar as implementações das regras computacionais. Exemplos de deverbais que ficaram de fora: *rotação* de *rodar*, *impressão* de *imprimir*, *decisão* de *decidir*, etc.

4.5.1. Os constituintes linguísticos x ambiguidade

Na Seção 1.2.1.1. decidimos separar a vogal temática para não perdemos ocorrências. Se utilizássemos o sufixo *-ação*, nunca recuperaríamos do *corpus* deverbais como *partição*, *expedição*. Apesar do sufixo *-ção* unir-se quase sempre a bases de verbos de primeira conjugação – portanto muito mais produtivo para verbos terminados em *-ar* – ele pode ocorrer com verbos de segunda conjugação. Sendo nosso intuito a alta cobertura, pesamos a complexidade computacional de deixar a vogal temática flexível e avaliamos como viável do ponto de vista computacional.

No caso dos sufixos *-me* e *-ta*, procedemos de maneira diferente com intuito de evitar ambiguidades. Por eles serem de tamanho menor e por isso passível de maior ambiguidade (*cometa*, de *comer*; *Sarita*, de *sarar*), decidimos juntar a vogal temática ao sufixo. No caso de

⁸⁶ Por serem considerados de alta frequência foram inseridos manualmente na variável base do programa.

-ata, possivelmente essa escolha não traria perda, já que não existe nenhuma ocorrência na literatura de *-eta ou -ita* como construtores de deverbais. Já o sufixo *-me* aparece na literatura adjunto a 'a' e 'u', por isso, duplicamos o sufixo: *-ame* e *-ume*.

Os deverbais de ação construídos sobre bases participiais, tiveram o 't' e o 'd' incorporados ao sufixo, por uma questão de diminuir as ambiguidades também.

Observemos as regras e suas respectivas variáveis a seguir:

I. Processo de sufixação:

base = lista de radicais

vt_1= [a]; vt_2= [e]; vt2= [i]

sufix= (mento|ção|são|dura|tura|ncia|nsia|dela|nço|nça|eira|ata|ame|ume|dela|gem)

II. Processo de conversão:

base = lista de radicais

dg_1= [a]; dg_2= [e]; dg_3= [o]

III. Processo de desflexionação:

base = lista de radicais

infinitivo: conj1= ar; conj2= er; conj3= [io]r

particípio: part=[ai]d[ao]

As formas geradas a partir das regras são:

Sufixação:

base	+	sufixo	→	deverbal
lav-a		ção	→	lavação°
		mento	→	lavamento
		são	→	lavasão*
		dela	→	lavadela°
		deira	→	lavadeira°
		ncia	→	lavância°
		dura	→	lavadura°
		gem	→	lavagem°
		aria	→	lavaria°
		nça/o	→	lavança° avanço°
		ume/ame	→	lavame° lavume°
		ata	→	lavata°
		nsia	→	lavansia*

Conversão:

base	+	dg	→	deverbal de ação
lav-		a	→	lava°
		e	→	lave°
		o	→	lavo°

Desflexionação:

base	+	dg	→	deverbal de ação
lav-		-ar	→	lavar°
		-ado	→	lavado°
		-ido	→	lavido*
		-ida	→	lavida*
		-ada	→	lavada°

Para cada verbo existem 13 deverbais de ação possíveis por sufixação, três por conversão e cinco por desflexionação, portanto um total de 21 deverbais de ação possíveis

para cada verbo. A partir de aproximadamente 14.000 bases verbais disponíveis, são gerados mais de 350.000 deverbais de ação.

Quanto mais padronizadas são as regras, mais simples é a implementação do ponto de vista computacional e, por conseguinte, mais legível é o código, facilitando possíveis manutenções que possam ser necessárias. Entretanto, para aqueles que não dominam expressões regulares, a compreensão não é trivial.

A idéia inicial era incluir as variantes de grafia nas próprias regras como ilustra a Tabela 30.

Tabela 30: Regras de recuperação automática de deverbais de ação com variação de grafia.

Fazer	Por sufixação:
faz- / feit-	Radical 1: (feit feitt feyt feytt fejt fejtt) + vt1 + sufix
	Radical 2: fa[zs] + vt2 + sufixo
	Por conversão:
	Radical2: fa[zs] + dg
	Por desflexionação:
	Infinitivo – fa[sz]er
	Participio1 – Fe[iyj]tt?[ao]
	Participio2 – fa[sz][iyj]d[ao]
	Participio3 – fa[sz]enda

Após o estudo piloto, constatamos que o sistema poderia ter maior cobertura se tivesse o apoio do SIACONF. Ganhamos regras mais confiáveis para detecção da variação de grafia (afinal o SIACONF foi criado exatamente para detectar e agrupar as variantes das palavras), aproveitamos um conhecimento já realizado. Além disso, ganhamos regras mais “limpas”, por não precisarmos incluir as variações de grafia na regra, e sobretudo, pela inviabilidade de construí-las manualmente agora que passamos de 50 verbos para mais de 14.000.

Outra vantagem em acrescentarmos o SIACONF – muito relevante – neste trabalho, está relacionada à usabilidade do EXTRADEV. Considerando que o SIACONF nos fornece todas as variações de grafia (baseadas nas regras que têm atualmente) a partir de uma entrada, pudemos manter as regras de detecção de deverbais de ação sem a variação de grafia, o que habilitou a extração de deverbais em *corpus* do português contemporâneo, já que a grafia dos deverbais de ação segue o padrão contemporâneo. Desse modo, o EXTRADEV passa a ter

dupla funcionalidade: pode detectar deverbais tanto em *corpora* contemporâneos, quanto históricos. Amplia-se a utilidade do sistema e pode-se contribuir com mais pesquisadores de diferentes vertentes teóricas.

4.6. Avaliações do EXTRADEV

As regras para geração de deverbais foram submetidas ao *corpus* DHPB através do sistema EXTRADEV que, como saída, gerou uma lista de deverbais históricos com e sem variação de grafia. Essa lista serviu para realizar a validação de quão preciso e abrangente é o algoritmo criado. As métricas de precisão (*precision*) e abrangência (*recall*) são muito utilizadas na área de PLN e são originárias da área de Recuperação da Informação (RI). Precisão é uma medida de fidelidade e abrangência é de completude (Baeza-Yates, Ribeiro-Neto, 1999).

Em RI, precisão é definida como o número de documentos relevantes recuperados por uma busca dividido pelo número total de documentos recuperados por aquela busca. Já abrangência é definida pelo número de documentos relevantes recuperados por uma busca dividido pelo número total de documentos relevantes que deveriam ter sido recuperados.

Para o cenário de PLN, em tarefas de classificação, precisão para uma classe é o número de verdadeiros positivos (isto é, o número de itens corretamente rotulados como pertencentes à classe positiva) dividido pelo número total de elementos rotulados como pertencentes à classe positiva (isto é, a soma de verdadeiros positivos e falso positivos, que são itens incorretamente rotulados como pertencentes à classe). Abrangência, neste contexto de classificação, é definida como o número de verdadeiros positivos dividido pelo número total de elementos que realmente pertencem à classe positivo (isto é, a soma dos verdadeiros positivos e falsos negativos, que são elementos que não foram rotulados como pertencendo à classe dos positivos mas que deveriam ter sido).

A avaliação foi realizada manualmente. Os seguintes aspectos foram considerados:

- a) todas as unidades lexicais da lista são possíveis deverbais de ação?
 - a.1) foram encontradas unidades lexicais agramaticais?
 - a.2) há palavras não construídas (ou não derivadas) na lista?

As respostas para essas perguntas se baseiam em dois fatores: regras hipergerativas e fatores de restrição da base. Discorreremos a seguir sobre eles.

4.6.1. Regras hipergerativas

A vantagem do modelo SILEX, como já foi mencionada, é gerar todas as possibilidades de construção para deverbais segundo uma determinada RCP. Essa característica deixa a regra com alta cobertura e, portanto, acreditamos não perder nenhuma ocorrência, já que não sabemos dentre as regras disponíveis para construção de deverbais de ação quais eram as mais usadas, quais os sufixos mais utilizados, etc. Nesse contexto, o modelo é ideal, entretanto, como em toda regra gerativa podem existir problemas – conhecidos como ‘ruídos’ na computação, ou seja, quando o dado não condiz com o que objetivamos. Desse modo, a seguinte pergunta vem à superfície: há perda em precisão devido ao caráter hipergerativo do modelo? A resposta é não. Primeiramente, as regras disponíveis para construção de deverbais são altamente produtivas, ou seja, é natural que para cada verbo existam 25 possibilidades de deverbais. Fator este que não influi na qualidade das regras, pois elas são bem delimitadas por seus constituintes, principalmente na sufixação. O que pode ocorrer é a existência de formas homógrafas que pertencem a outras categorias.

No paradigma de construção de deverbais há um conjunto determinado de constituintes que **podem** construir deverbais. No entanto, sabemos que para uma determinada base, geralmente, apenas um sufixo está sempre ligado a ela e a essa combinação é atribuído concomitantemente um sentido. Desse modo dentre as 25 possibilidades uma, ou duas, talvez três coexistam em determinado período, então, dentre as outras 22 formas geradas, pode existir algum homógrafo de outras unidades lexicais e serem recuperadas palavras que não são deverbais. Esse fato não desmotivou em nada a utilização do algoritmo hipergerativo, pelo contrário, é justamente tal característica que habilita que novas estruturas morfológicas de deverbais de ação sejam descobertas.

Previmos que alguns deverbais gerados pela regra poderiam ter maior número de homografias como o próprio verbo, adjetivos e substantivos. Para evitá-los ou tentar diminuir seus efeitos em sistemas computacionais, muitas vezes, temos de recorrer a outros tratamentos como gramáticas de desambiguação, por exemplo. Os deverbais de ação resultantes do processo de desflexionação são os mais difíceis de garantir sua categoria, já que eles são sofrem alteração formal em relação ao verbo. Vejamos alguns exemplos nos excertos abaixo:

- a) Desflexionação de infinitivo e participio.

- *feito*

(...)a todos os que as quisessem povoar divulgando-se primeiro a notícia da bondade, e fertilidade do terreno, para serem inumeráveis as famílias, que concorreriam; e

*muito mais tendo-lhes já lá os víveres, que dissemos, e alojamentos **feitos**; e promessa de quantas terras podés sem cultivar.*

*(...)onde tomou delle posse em nome de El Rey com muita solemnidade de actos, que mandou fazer muito bem notados, e com este **feito** se tornou mui satisfeito a Pernambuco, e dahi depois de concluidos os negocios de seu officio outra vez para a Bahia(...)*

No primeiro excerto *feito* ocorre como verbo, já no segundo *feito* nomeia a ação de fazer, portanto, assume a forma nominal.

- *mandar*

Parece-me cousa muy conveniente mandar S. A. algumas molheres, que lá tem pouco remedio de casamento, a estas partes, ainda que fossem erradas, porque casaram todas muy bem, com tanto que nom sejam taes que de todo tenham perdida a vergonha a Deus e ao mundo.

*Mas amasiaram-se tão mal os Padres com estes capitães e tinham cada dia tanta porfia e diferenças sôbre o **mandar** dêstes índios (...)*

O mesmo comportamento do excerto acima acontece com *mandar*, primeiro aparece como verbo e depois como deverbais.

Para sabermos ao certo quais de fato são deverbais de ação teríamos de realizar uma análise contextual – detectar padrões de ocorrência, como por exemplo: são acompanhados de determinantes: [a, esta, uma, um, o] parada, olhar, saída, etc. Esse refinamento poderá ser implementado em trabalhos futuros.

Outro fenômeno semelhante que pode ocorrer é que um deverbais pode ser exatamente idêntico a uma palavra não construída de origem não verbal. Contudo, esses casos têm baixa frequência, portanto não são muito representativos na análise dos dados.

- b) Coração – substantivo concreto, órgão que bombeia o sangue. É um dos primeiros significados que nos vem à mente, entretanto, esse poderia ser também um deverbais de ação construído por sufixação a partir do verbo *corar* adjunto ao sufixo -ção → *cora-* -ção.

Alguém por favor veja se a carne está corada.

*A **coração** da carne ainda não está boa, mamãe.*

Parece estranho ao homem de hoje, mas não podemos assegurar que isso também não ocorria nos séculos XVI, XVII, ou XVIII. Apenas no sentido acima, essa palavra seria um deverbais de ação, pois no sentido concreto – um órgão vital – essa não é uma palavra

construída, uma vez que não tem uma base reconhecida, não ao menos no português. Provavelmente é oriunda do latim e inseriu-se no português já nesta forma.

Outro caso semelhante aos citados é quando um sufixo coincide com uma desinência verbal, como, por exemplo, o sufixo *-ria* que é terminação característica das 1ª e 3ª pessoas do singular do futuro do pretérito. Nesses casos é possível que exista, na frequência encontrada de um deverbais construído com esse sufixo, muitos mais verbos do que deverbais (esse fato será retomado na Seção sobre os sufixos mais frequentes do português histórico). Mas não podemos afirmar, afinal, trata-se de um *corpus* histórico e podem ocorrer fenômenos linguísticos não previstos. Esses, portanto são os casos previstos que podem contribuir para um aumento no número de deverbais gerados.

Para esses casos e, muitos outros que envolvem diversas aplicações computacionais relacionados à língua portuguesa, deve-se recorrer ao contexto, ou seja, analisar a frase onde ele ocorre. Esse procedimento pode ser feito manualmente, mas também dependendo da quantidade de dados e do objetivo da pesquisa, a construção de uma gramática de desambiguação é um dos recursos usados por cientistas da linguagem.

Após a explanação de alguns casos de ambiguidade, é importante reforçar (e responder as perguntas feitas no início deste Capítulo) que do ponto de vista da disponibilidade de regras para produzir e interpretar as formas construídas na língua, todos os deverbais de ação presentes na lista são possíveis, (não necessariamente em uso na língua). Segundo Basílio (2006, p. 11): “o potencial de atuação dos processos de formação de palavras não é igual a formação concreta de novos itens”. Em consonância está Corbin:

No Componente Derivacional operam as RCPs, cujo output corresponde às ‘palavras construídas possíveis’ na língua, regulares, predizíveis e, dado que podem servir de base a novas RCPs, de número infinito, embora nem todas elas sejam atestadas no léxico atual (CORBIN, 1987, p. 418 apud SANTOS, 2006 p.71).

Em relação a encontrarmos palavras agramaticais na lista de deverbais históricos, por mais que algumas regras tenham gerado algum fenômeno como esse (De *descrever* → *descrevição**), para uma unidade lexical agramatical estar na lista, é necessário que ela ocorra no *corpus*. Sendo agramatical, não estaria no *corpus* de pesquisa, portanto, toda a lista têm unidades lexicais reconhecidas no português histórico.

4.6.2. Restrições da base

Sobre os verbos monossílabos átonos é inviável a construção de deverbais de ação por conversão. Há uma restrição de caráter morfoprosódico que admite existir um radical mínimo para geração de deverbais por conversão. Por exemplo, os verbos *dar* e *ser*, retiradas as terminações equivalentes às conjugações (-ar, -er, -ir), restam apenas as consoantes iniciais *d-* e *s-* que podem servir de base para o processo construcional da sufixação e para o processo de desflexionação *dação*^o, *damento*^o, *ser*^o, etc. Porém, a existência de apenas uma consoante como radical inviabiliza o processo de conversão.

Verbo →	extração do radical	→	processo de conversão (base + dg)
<i>dar</i>	<i>d-</i>	<i>d-</i> + [<i>a e o</i>]	→ <i>da, do, de</i>
<i>ser</i>	<i>s-</i>	<i>s-</i> + [<i>a o e</i>]	→ <i>se, so, sa</i>

Segundo (RODRIGUES, 2004) existem algumas restrições e condições sobre as bases para que elas sejam produtivas, uma delas é chamada de condição morfoprosódica para radical mínimo:

Uma das ausências mais salientes do conjunto de postverbais do português é a de substantivos formados a partir de verbos que apresentam uma estrutura fonético-fonológica de extensão curta. Quando nos referimos a uma estrutura fonético-fonológica de extensão curta, pensamos em verbos que, no infinitivo, apresentam por exemplo uma estrutura composta por (Consoante) vogal Consoante. Falamos, pois, de verbos como *ver*, *ler*, *dar*, *ser*, *ter*, *rir*, *vir*, *ir*. (...) Na formação dos postverbais, o verbo derivante ocorre sob a forma do radical (Rodrigues, 2001). Depois de o radical verbal sofrer um processo de conversão para substantivo, é-lhe agregado um marcador de classe (-o, -a ou -e) (Rodrigues (em publicação)). Com base neste processo, procedemos as hipotéticas construções que se revelaram impossíveis (1). a. *ver* > **va*, **ve*, **vo*; b. *ler* > *la**, *le**, *lo**; c. *dar* > *da**, *de**, *do** (...); f. *rir* > **ra*, **re*, **ro* (...) Damos exemplos de verbos das três conjugações do português, pois todas elas se apresentam disponíveis para a formação de postverbais, e apresentamos as hipotéticas construções com todos os marcadores de classe para cada um dos verbos, pois para cada uma das conjugações verbais há postverbais com os três marcadores de classe.

Nessas condições foram excluídas as bases verbais dos verbos monossílabos átonos para construção de deverbais por conversão.

5. Análise dos Resultados do Sistema EXTRADEV

O EXTRADEV é um sistema de extração semi-automático de deverbais de ação que engloba os três processos, através dos quais, os deverbais são construídos – sufixação, conversão e desflexionação . Esse sistema é capaz de processar mais de 14 mil bases verbais de uma única vez, que, por sua vez, se concatenam a outros constituintes linguísticos formadores de deverbais de ação, já previamente descritos e formalizados nesta dissertação de mestrado. Por esse sistema realizar buscas a partir de regras linguísticas específicas para deverbais de ação, abarcando todos os constituintes (não apenas partes de palavras como os buscadores comuns) formadores dos mesmos em uma regra só, ele demonstra desempenho superior a buscadores como os existentes em processadores de texto como o Word, por exemplo, que não estão preparados para concatenar constituintes e trabalhar com pipeline de expressões regulares (ou seja, várias possibilidades em uma só regra).

No EXTRADEV, usamos o seguinte pipeline para recuperar os deverbais de ação de um *corpus*, usando o processo de sufixação:

(Base) + (vt) + (sufixos)
(Filh- + -a- + mento|ção|são|dela|dura|tura|ata|ncia|nsia|nço|nça|gem|deira|aria|ume|ame)

Nesse exemplo, a Base representa todos os verbos extraídos do dicionário do Unitex que pertencem a primeira conjugação e vt, a vogal temática *-a-*.

5.1. Um comparativo entre o sistema EXTRADEV e buscadores de processadores de texto

Os buscadores presentes em processadores de texto operam com textos em línguas naturais e alguns suportam expressões regulares. Por exemplo, o Word 2003 oferece buscas avançadas que podem ser feitas através de sufixos e prefixos. Portanto, podemos recuperar todas as palavras formadas por determinado sufixo. Para isto, basta selecionar “Find” , “More” , ativar “Match suffix” e escrever o sufixo desejado. A Figura 17 mostra a interface do buscador do Word.

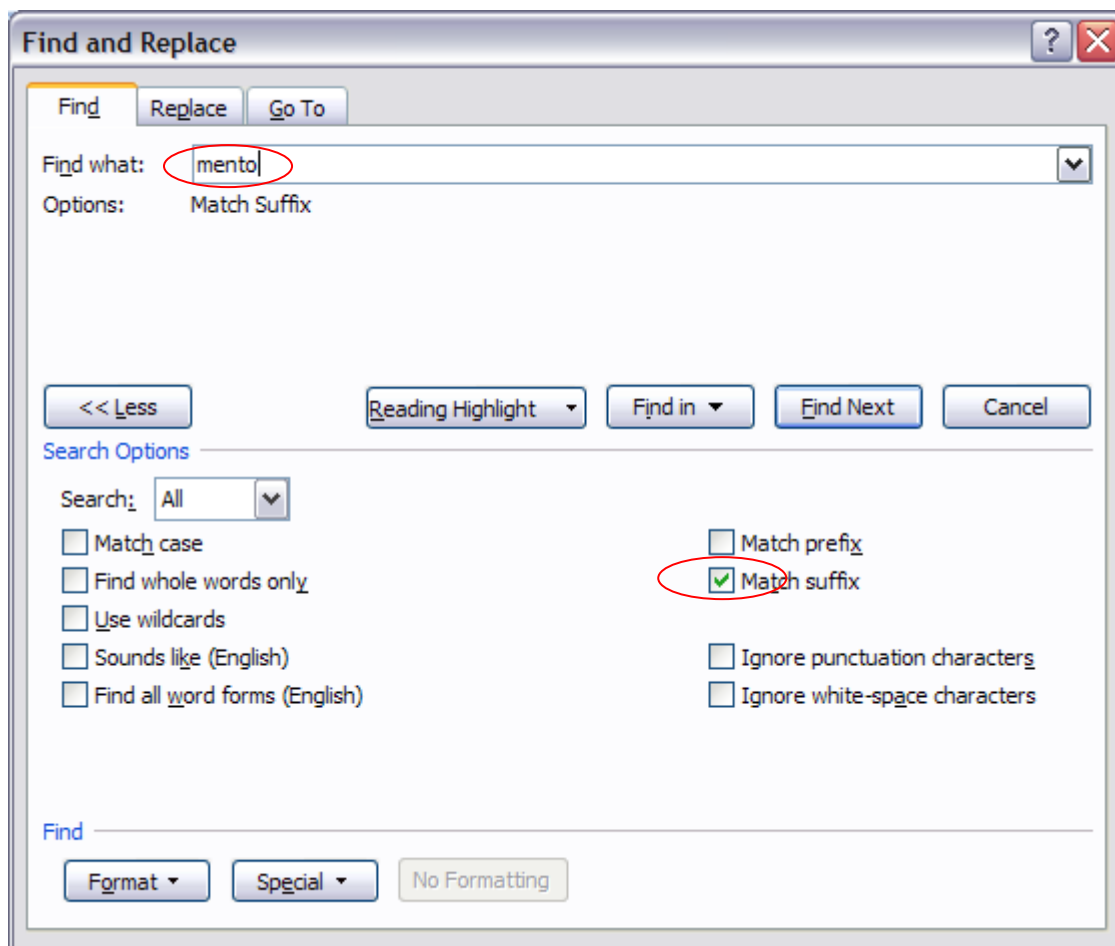


Figura 17: Interface do buscador de palavras do Word.

No entanto, por realizarmos a recuperação baseada apenas em um constituinte da palavra, perdemos muito em **precisão**, pois podemos recuperar muitas unidades lexicais que, embora tenham a terminação requerida, não são sufixos deverbais e, na maioria dos casos, não são palavras construídas. Por exemplo, se colocássemos no buscador o sufixo *-mento* recuperaríamos as palavras *fermento*, *aumento*, *cimento*, *lamento*, etc. (caso elas existam no texto). Se fizéssemos o caminho contrário e, ao invés de realizarmos as buscas por sufixos, partíssemos das bases verbais, poderíamos recuperar todos os sufixos que se juntam a essa base, sufixos que não pertencem a RCP de deverbais de ação e, sobretudo, todas as terminações que não são sufixos, por exemplo as desinências verbais. Por exemplo, se buscássemos pela base *parti-*, poderíamos recuperar no *find* do Word: *partimos*, *partirem*, *particular*, *partiriam*, *partirei*, *participantes*, *partida*, *partimento*, *partido*, *particularmente*,

partindo, particularidades, parti, partirão, partiam, participe, partição e muitas outras ocorrências. Sendo, das 17 ocorrências, apenas dois deverbais.

Esse fato revela a importância de determinarmos todos os constituintes desta categoria e os utilizarmos para realizar buscas precisas e de maneira rápida.

Se utilizássemos o EXTRADEV, obteríamos somente *partimento* e *partição*, ou seja, evitaríamos cerca de 80% de “lixo”, pois o EXTRADEV realiza uma composição de regras (pipeline) que une duas ou mais expressões regulares, combinando o resultado de uma com outra(s).

Além da baixa precisão decorrente da busca baseada em apenas uma parte da palavra, outro inconveniente desse procedimento é que teríamos de colocar base por base no buscador, após a busca, separar as ocorrências distintas de cada sufixo e depois reuni-las em um documento texto. Caso a busca seja para poucos verbos, esse método é exequível. Porém, se tivermos uma quantidade de dados grande, esse método demandaria muito tempo. Nesse aspecto o sistema EXTRADEV também é superior, pois realiza de maneira automática a extração de mais de 14 mil bases verbais, concatenando-as às vogais temáticas e sufixos (ou a desinência de gênero no caso de conversão) e procurando cada possibilidade gerada no *corpus* de entrada de uma única vez.

Na Tabela 27, apresentamos um quadro com o desempenho do buscador Word 2003 em relação ao sistema EXTRADEV, a partir do processo de sufixação.

Tabela 31: quadro comparativo do desempenho do buscador Word e do sistema EXTRADEV

Sufixos	Buscador Word com caracteres especiais	EXTRADEV – total de ocorrências	EXTRADEV – deverbais distintos	Ganho por utilização do Extradev em relação ao total de ocorrências
-mento	21.423	4.568	185	~ 9 vezes mais preciso
-ção	41.857	17.638	569	~ 2,3 vezes mais preciso
-gem	7.646	1.367	51	~ 5,5 vezes mais preciso
-tura	6.158	220	13	~ 28 vezes mais preciso
-ria	29.014	3.389	412	~ 8,5 vezes mais preciso
-dura	1.245	236	46	~ 5,3 vezes mais preciso
-dela	1.133	3	3	~ 377 vezes mais preciso

-me	33.436	194	56	~ 187 vezes mais preciso
-nça	11.240	2.144	50	~ 5,3 vezes mais preciso
-nço	1.361	28	8	~ 48,6 vezes mais preciso
-nsia	96	8	5	12 vezes mais preciso
-ta	120.002	795	56	~ 151 vezes mais preciso

O quadro acima revela que uma busca mais refinada, ou seja, uma busca que é realizada a partir do conhecimento linguístico da categoria requerida é muito mais precisa, que as buscas feitas baseadas em partes de palavras, sobretudo para os sufixos de tamanho menor que são mais passíveis de ambiguidade.

5.2. Dados coletados

Partimos de 14.281 bases verbais para geração de inúmeras construções deverbais disponíveis na língua. Dentre as possibilidades construídas, foram encontrados 1.742.663 ocorrências de possíveis deverbais de ação e 15.633 formas distintas dos mesmos sem variação de grafia. Este número mais as variantes extraídas no segundo módulo do EXTRADEV, somam (6.809 variantes + 15.633 deverbais sem variação de grafia) um total de 22.442 ocorrências de deverbais históricos.

Conforme mencionado acima, o *corpus* de trabalho tem aproximadamente 7.500.000 (sete milhões e meio de palavras), das quais, aproximadamente, 1.749.472⁸⁷ (um milhão setecentos e quarenta e nove mil e quatrocentos e setenta e duas) são formas deverbais de ação possíveis na língua portuguesa, traduzindo em percentagem, são aproximadamente 23% do *corpus* total. Consideramos que a percentagem dos dados é passível de variação para mais ou para menos por duas razões principais: variação de grafia característica do português histórico (ver Seção 4.5.1) e como vimos no Capítulo anterior, pode haver palavras coincidentes com as formas deverbais, mas que não são palavras construídas.

Devido à imensa quantidade de dados gerados e as diversas possibilidades de análise, seguimos dois critérios de análise:

⁸⁷ Nesse número só foram incluídas as formas variantes distintas ocorridas no corpus. Esse número aumentaria se tivéssemos a frequência total de cada uma das variantes encontradas.

I - Baseado em dados de frequência: verificação de quais eram os processos mais recorrentes no português histórico e em relação ao processo de sufixação, quais eram os sufixos de maior ocorrência; e realização de uma análise comparativa com o trabalho de Basílio (1996) em relação à frequência de ocorrências no discurso dos sufixos *-ção* e *-mento*. Para obtenção dessas sublistas (lista de deverbais formados por sufixação, conversão, lista dos deverbais apenas formados pelo sufixo *-gem*, etc.) utilizamos comandos de processamento de texto (no apêndice B é possível observar as sublistas de deverbais de ação construídas com os sufixos *-ção* e *-mento*).

II- Baseado na observação da lista de deverbais históricos final: extração de deverbais de ação que nos despertaram a atenção por apresentarem formas distintas da utilizada no português contemporâneo – ou em relação aos sufixos utilizados ou em relação à mudanças no processo de construção (de sufixação → conversão, coexistência de dois processos para o mesmo verbal de ação, etc.).

Antes de iniciarmos a análise dos dados, discorreremos sobre a dificuldade de obtermos dados de frequência em *corpus* histórico. E, então, logo em seguida, apresentaremos os dados obtidos sobre os diferentes deverbais encontrados.

5.3. Desafios da contagem de palavra em corpus histórico

Para sabermos a frequência exata de um verbal de ação no *corpus*, são necessários dois procedimentos: o primeiro é conhecer todas as variantes do verbal em questão e, em segundo lugar analisar o contexto no qual está inserido o verbal.

Conhecer as variantes possíveis para um determinado verbal implica diretamente na sua frequência. As regras de construção de deverbais de ação seguem as regras de ortografia vigentes no português contemporâneo. Por essa razão, a fim de recuperar variantes para obtenção de um dado mais realístico, utilizamos o sistema SIANCONF (ver Seção 4.3.4). Para ilustrar esse fato, vejamos o verbal abaixo:

1352 povoação povoação povoaçam povoaçãõ povoação povoação

A unidade lexical *povoação* é a mais frequente construída por sufixação e a 237ª (ducentésima trigésima sétima) em relação ao *corpus* total. Esses dados são baseados na frequência da variante *povoação* (apenas uma das seis formas possíveis), ou seja, são dados

aproximados. Por isso é importante saber quais as variantes para uma possível busca e obtenção da frequência (via comando de processamento de textos ou ferramentas computacionais disponíveis). Dessa maneira, as variantes têm as seguintes frequências: *povoaçáo* = 49; *povoação* = 1; *povoaçam* = 12; *povoaçãõ* = 149; *povoaçao* = 24. Portanto temos $1352 + 235 = 1587$. Neste trabalho nos propusemos a recuperar as formas deverbais e suas variantes para que estudiosos tenham um acesso rápido às variantes que um deverbais de ação possa ter. A recuperação da frequência para cada variante será um trabalho futuro.

Outro aspecto relevante da variação de grafia, que prejudica a ocorrência exata de um determinado deverbais, é o fato de existirem mudanças de acento e de letra, as quais podem resultar em outras unidades lexicais reconhecidas pela língua, que podem ser tanto variações da mesma palavra quanto uma unidade lexical distinta.

A variante mais recorrente no corpus é *distancia* sem o acento circunflexo, o que designa no português contemporâneo uma forma verbal no presente do indicativo e na terceira pessoa do singular. Porém, sabemos que no português histórico havia variações para grafar a mesma palavra, portanto, dentre as 1230 ocorrências de *distancia*, podemos encontrar tanto formas verbais quanto nomes deverbais de ação. Nesse sentido, o valor que temos é superestimado, porém não tanto. Analisamos superficialmente 15 excertos que continham a palavra *distancia* e constatamos que todas, apesar da falta do acento, estavam na forma de nome. Isso significa quase 2% apenas de todas as ocorrências dessa palavra no *corpus*; para obtermos um dado real seria necessário nos debruçarmos sobre todas as ocorrências e analisar qual a forma empregada.

Quando ocorrem fenômenos como este, podemos ter o dado de frequência superestimado, ao contrário do aspecto anterior, no qual perdemos dados por não conhecermos todas as possíveis variantes de um mesmo deverbais de ação. Na Tabela 31, apresentamos uma parte da lista de deverbais históricos total (com e sem variantes).

Tabela 32: Parte da lista de deverbais históricos com e sem variação de grafia.

Avexação	Avexação		
affentamento	Assentamento		
arrifcaria ^o	Arriscaria ^o		
annunciação	Annunciação	Anunciação	Anunciação
anniquilação	Aniquilação		
allucinação	Alucinação	Halucinação	

alienação	Alienação		
allegação	Alegação		
agregação	Agregação		
afundação	Afundação		
administrasam	Administrasão		
accommodaria	Acomodaria		
habitação	Habitação	abitação	habitaçam

5.4. Processos construcionais mais recorrentes no português histórico

Antes de iniciarmos a análise dos dados em relação a cada processo, é necessário definir dois conceitos muito importantes: produtividade e frequência. Neste trabalho, compreendemos produtividade relacionada à disponibilidade de regras e à ocorrência de palavras distintas; e frequência relacionada ao número de ocorrências de cada palavra encontrada no *corpus*. Observemos o excerto abaixo:

Com a preocupação de ressaltar a distinção entre produtividade e frequência de aplicação, Corbin (1984) estabelece o conceito de disponibilidade, definindo uma regra como disponível quando passível de utilização pelo falante/ouvinte no exercício normal da criatividade lexical e colocando o conceito de disponibilidade na esfera da competência lexical, enquanto a rentabilidade de uma dada regra se situaria na esfera da Análise do Discurso. Segundo Corbin, a produtividade define formas disponíveis e não formas necessariamente atestadas. Assim, o conceito de produtividade deve ser entendido tão somente como o conjunto de condições estruturais de uma regra na definição de construções morfológicas possíveis. A produtividade de uma regra deve ser distinta das Condições de Produção, que são condições não estruturais de operação e dependem de fatores de ordem paradigmática, discursiva e sócio-cultural. (Basílio, 1999 p.59)

A regra de construção de deverbais de ação é altamente produtiva (até o presente momento, é relativamente pequeno o número de restrições de construção de deverbais conhecidas) haja vista, a quantidade de processos construcionais disponíveis e, em especial, na sufixação, a quantidade de sufixos disponíveis para se adjungir às bases verbais. Contudo,

apesar de termos conhecimento do caráter produtivo da RCP de deverbais de ação, não sabemos ainda como essas formas se distribuem no português dos séculos XVI, XVII e XVIII e início do XIX. Por isso, esta Seção foi dedicada a descrição e análise dos deverbais de ação encontrados no português histórico.

Para realizarmos uma breve análise dos dados encontrados, além de nos basearmos em critérios de frequência e ocorrências de palavras distintas, utilizaremos também, o Dicionário da Língua Portuguesa (MORAES, 1813) e o Dicionário de Etimologia da Língua Portuguesa (CUNHA, 1959) para descrevermos os deverbais do PB e relação a sua estrutura: são palavras construídas, são palavras simples, são palavras construídas em outra língua e importadas para o português, etc.. No entanto, esperamos que os dados levantados neste trabalho sejam aprofundados por outros pesquisadores.

Todas as análises abaixo foram realizadas sobre os dados dos deverbais de ação sem a variação de grafia, por não sabermos a frequência de cada variante. Acreditamos que não foram perdidos muitos dados, pois, como observamos na lista total de deverbais, na grande maioria dos casos, a maior frequência é das que apresentam formas com grafia contemporânea.

5.2.1 Sufixação

Foram encontradas 1.498 formas distintas de deverbais de ação construídas pelo processo de sufixação e 914 formas variantes, somando um total de 2.412 deverbais de ação. Dentre as formações sem variação de grafia, os sufixos distribuem-se conforme ilustra a Figura 18.

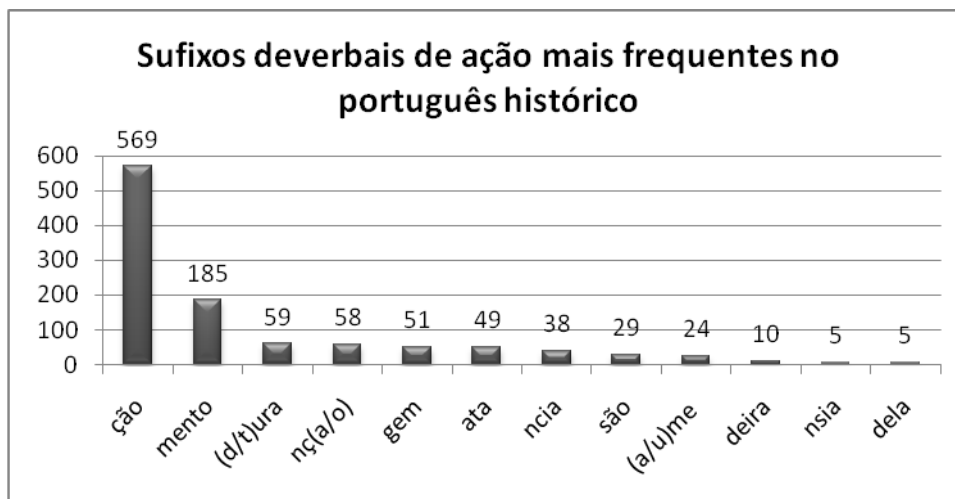


Figura 18: Gráfico dos sufixos deverbiais de ação mais frequentes no português.

O sufixo *-ção* é o mais frequente com 569 ocorrências no *corpus*, que correspondem a 35% do total de deverbiais encontrados. O segundo sufixo mais recorrente é *-mento* com 185. Ou seja, representativo de 11% do total de deverbiais do *corpus*. Ambos – *-mento* e *-ção* – são citados como sufixos formadores de deverbiais de ação por todos os autores da tabela 3 (Seção 1.2.1.1.) Esses são os sufixos mais frequentes no português atual e, segundo os dados coletados, os mais frequentes, também, no português histórico.

Observemos a Figura 20.



Figura 19: Gráfico das estruturas de deverbiais de ação mais frequentes construídas com o sufixo *-ção*

Para geração desse gráfico fizemos um recorte baseado na frequência 100 dos deverbais de ação terminado sem *-ção* (dada as 569 formas encontradas). Isto é, geramos um histograma de todas os deverbais do *corpus*, extraímos apenas as palavras que apresentavam frequência maior que 100 e, então, obtivemos uma lista de 46 deverbais distintos terminados em *-ção*. Dentre as possíveis estruturas morfológica encontradas, destacam-se as palavras construídas, revelando grande produtividade de deverbais de ação com sufixo *-ção* construídos no próprio português. Também é bastante representativa a quantidade de palavras de origem latina importadas para o PB. Para aquelas unidades lexicais que não foram encontradas nos dicionários supracitados acreditamos que sejam por razões de falta de dados, ou por não serem recorrentes na língua escrita.

Para analisarmos o sufixo *-mento*, seguimos a mesma metodologia utilizada para o sufixo *-ção*, pelo mesmo motivo – grande quantidade de dados. Acima da frequência 100, foram extraídos apenas 10 deverbais de ação sufixados em *-mento*, vejamos a Figura 21.

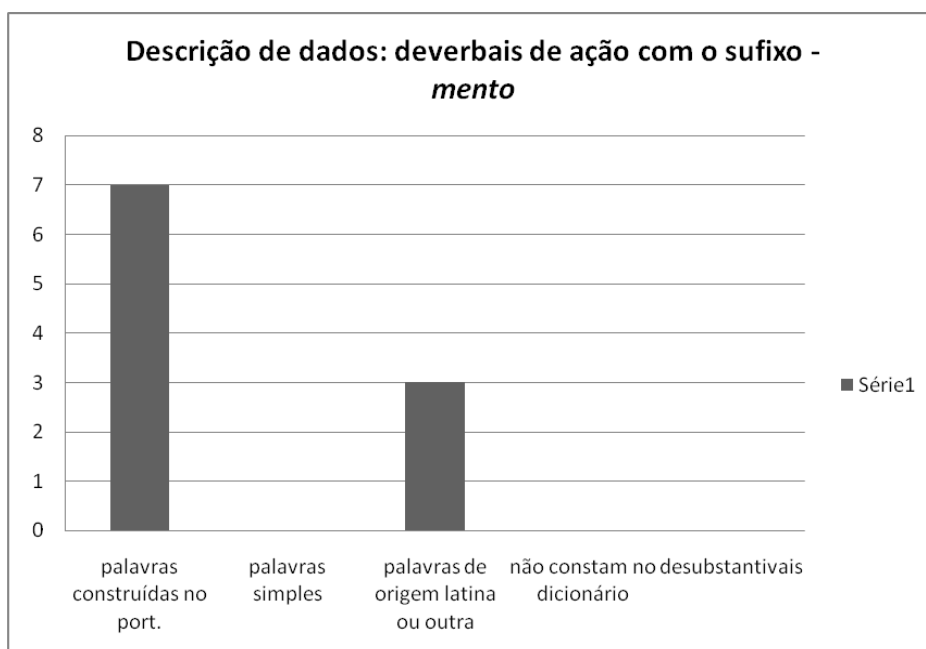


Figura 20: Gráfico das estruturas de deverbais de ação mais frequentes construídas com o sufixo -mente

Como podemos observar no gráfico, todos os deverbais mais frequentes constam nos dicionários e a maioria são palavras construídas no português.

Em terceiro lugar aparece o sufixo *-(t/d)ura* com 59 ocorrências, que se adjunge à temas do particípio passado e está presente como formador de verbal em todas as obras citadas na Tabela 3 (Seção 1.2.1.1) também. Em relação às formas existentes do português histórico, esse sufixo é bastante produtivo, combinando-se a 59 tipos diferentes de bases

verbais, porém, diferentemente, de *-ção* e *-mento*, que continuam muito produtivos no português atual, *-(t/d)ura* apresenta queda na produtividade ao longo dos anos. Atualmente, é raro encontrarmos nomes deverbais com esse sufixo em textos jornalísticos por exemplo ou mesmo na língua oral. Os deverbais construídos com o sufixo *(t/d)ura* estão listados na Tabela 35.

Tabela 33: Possíveis deverbais de ação terminados com o sufixo *(t/d)ura* encontrados no corpus histórico.

Estatura	Temperatura	Formatura	Criatura	Assinatura
Vocatura	Nomenclatura	Curvatura	Abreviatura	Quadratura
Prematura	Crispatura	Cravatura	Quebradura	Atadura
Semeadura	Embocadura	Fechadura	Cavalgadura	Rapadura
Dentadura	Ferradura	Queimadura	Arranhadura	Pregadura
Empunhadura	Cortadura	Bordadura	Andadura	Sangradura
Ordenhadura	Esfoladura	Coadura	Armadura	Trilhadura
Serradura	Torradura	Remadura	Roçadura	Rachadura
Pisadura	Olhadura	Molhadura	Meladura	Matadura
Limadura	Lavradura	Ligadura	embaraçadura	Desembocadura
crespadura	Cornadura	Codeadura	Chanfradura	Cercadura
Catadura	Caldadura	Borradura	Assadura	

Muitos dos deverbais acima são usados, hoje, como designadores de substantivos concretos que representam instrumentos da ação de X (verbo base) ou resultantes da ação de X (verbo base), como, por exemplo: *ligadura*, *armadura*, *empunhadura*, *atadura*, *assinatura*, *etc..*

Os substantivos deverbais de ação terminados em *-(d/t)ura* encontrados no PB podem ser palavras construídas em português e podem ser também palavras contruídas em outra língua e imputadas para o português, portanto palavras não construídas. Observemos o gráfico abaixo:

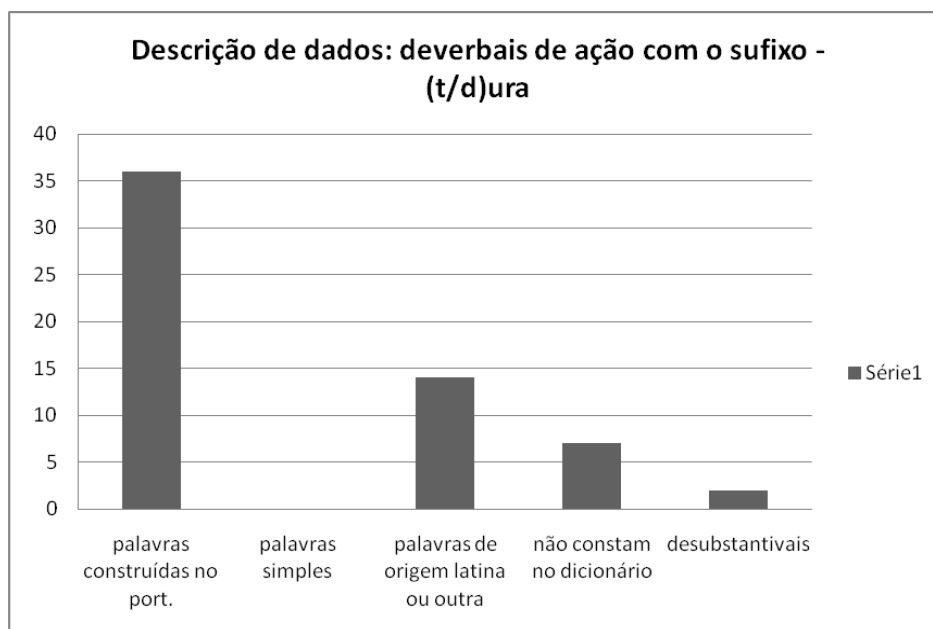


Figura 21: Gráfico das estruturas de deverbais de ação mais frequentes construídas com o sufixo **-mente**

Como podemos observar, a maioria das palavras extraídas do *corpus* com o sufixo *-(t/d)ura* são palavras construídas, ou seja, a base e o sufixo são reconhecidos no português; com um pouco menos da metade da quantidade das palavras construídas, aparecem os deverbais já construídos em outra língua como por exemplo *assadura*: do lat. *assatūra* e *sangradura*, que, segundo o dicionário Moraes, adaptação do castelhano. Algumas palavras não foram encontradas nos dicionários, ou pela baixa frequência na língua escrita ou pelo trabalho lexicográfico antigamente não se utilizar de grandes corpora e desse modo podendo perder algumas ocorrências típicas da língua. E com apenas duas ocorrências, palavras que não foram construídas sobre a base verbal e sim sobre a base nominal: é o caso de *detadura* (de *dente*) e *Cornadura* (de *cornu*). A coluna das palavras simples para este sufixo aparece vazia, porém, para outros sufixos ela não será nula, já que alguns sufixos na verdade não são sufixos, mas simples terminações.

Em seguida, com praticamente o mesmo número de ocorrências distintas (51), o sufixo *-nç(a/o)*. Esse sufixo aparece em três dos quatro trabalhos consultados e é o terceiro com mais ocorrências distintas no PB histórico. Observemos a Tabela 33.

Tabela 34: deverbais construídos com o sufixo –nça/o encontrados no corpus histórico.

esperança	Segurança	lembrança	confiança	mudança
vingança	Fiança	cobrança	vizinhança	desconfiança
Criança	Semelhança	Balança	governança	Ordenança
Vereança	Perseverança	Abastança	tardança	Matança
Aventurança	Intemperança	Aliança	temperança	Privança
pujança	Sustança	Usança	Picanço	esperanço
Afianço	Esquivança	livrança	sobrepujança	selança
provança	Presança	poupança	pitança	embalança
doença	Destemperança	dessemelhança	comilança	breança
abalança	Bastança	avezinhança	balanço	Remanço
Gravanço				

O gráfico abaixo exibe uma menor ocorrência de palavras construídas em outras línguas, menor ocorrência de desubstantival em relação ao sufixo *-(t/d)ura*. Por outro lado, aumentaram as palavras fora do dicionário e surgiram as primeiras unidades simples com a terminação equivalente a do sufixo.

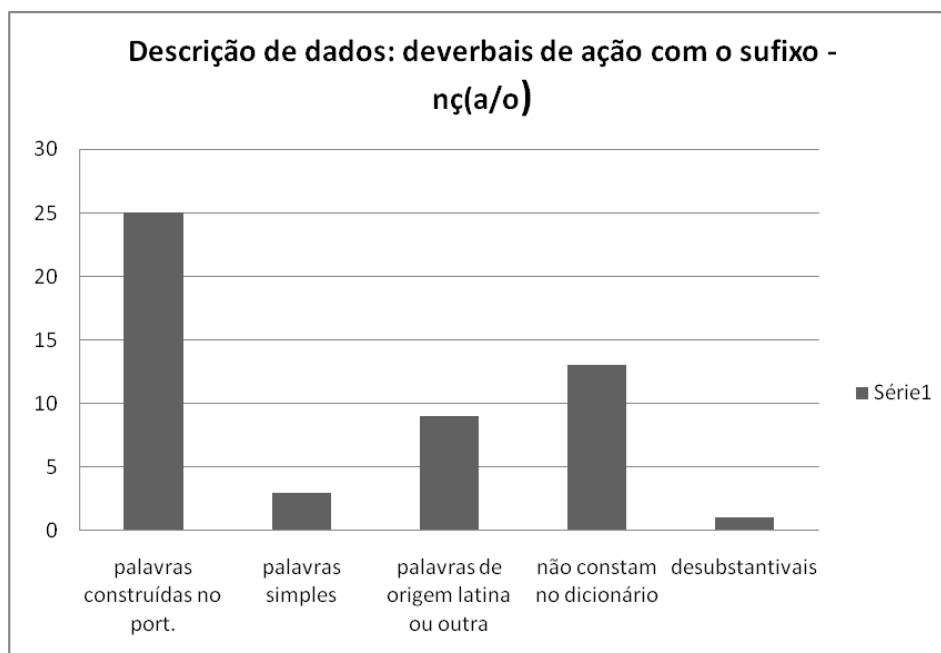


Figura 22: Gráfico das estruturas de deverbais de ação mais frequentes construídas com o sufixo -nç(o/a)

As palavras simples encontradas são: *balanço* (talvez do antigo italiano *balancio*), *balança* (do latim *bilancia*) e *gravação* (do castelhano *garbanzo*). Elas aparecem na lista de deverbais por apresentarem formas de grafia equivalente a deverbais construídos, nesses casos -nç(o/a) não é sufixo e sim uma simples terminação. O único desubstantival encontrado foi *vizinhança* (de *vizinho*).

O quarto sufixo mais produtivo é o -gem que apesar de participar da construção de deverbais de ação, também é produtivo em outras RCPs. Segundo os quatro trabalhos (entre gramáticas e obras acadêmicas) citados na Tabela 3 (Seção 1.2.1.1) apenas dois mencionam -gem como sufixo formador de deverbais de ação. Tal comportamento pode estar relacionado ao fato desse sufixo também participar da construção de substantivos coletivos a partir de bases nominais, daí a inconsistência entre os trabalhos visitados. Foram encontradas 51 ocorrências apresentadas na Tabela 34.

Tabela 35: Deverbais de ação construídos com o sufixo -gem .

Passagem	Paragem	Equipagem	ventagem	Tanchagem
Lavagem	Bagagem	Hospedagem	estalagem	Folhagem
Vassalagem	Tapagem	Coragem	marinhagem	Ferragem
Contagem	Carceragem	Talagem	carnagem	Aragem
Plumagem	Pilhagem	pastagem	remagem	Pilotagem
Menagem	Ancoragem	vendagem	aparagem	Salvagem
Limagem	Vragem	senhoreagem	portagem	Pesagem
Pardagem	Liagem	lavragem	lapidagem	Fradagem
Forragem	Esquipagem	Ervagem	costeagem	Cabotagem
Bobagem	Beberagem	Aviagem	aventagem	Abordagem

O sufixo *-gem* também apresenta frequência representativa no português histórico. Dessa maneira, a não referência a ele em alguns trabalhos visitados é um equívoco, sobretudo por ser este sufixo produtivo no português atual na formação de deverbais de ação.

Observemos a Figura 24.



Figura 23: Gráfico das estruturas de deverbais de ação mais frequentes construídas com o sufixo -gem

O gráfico revela uma grande quantidade de palavras não atestadas em dicionários, o que pode explicar a ausência nos trabalhos estudados deste sufixo como formador de deverbal de ação. Outro aspecto relevante é o fato de termos mais da metade de palavras estrangeiras em relação a quantidade de palavras construídas em português. De fato palavras como: lavagem, forragem, bagagem, ancoragem, etc. foram todas importadas do francês segundo dicionário de etimologia.

Em seguida, aparece o sufixo *-ata* (49). Observemos os deverbais na Tabela 35.

Tabela 36: Deverbais de ação construídos com o sufixo -ata

Remata	Relata	Aligata	concordata	Imbricata
apostata	Operata	Abata	mandata	Serrata
Cascata	bifurcata	Aromata	mediata	Confirmata
Anulata	Vastata	Revelata	gravata	Fraguata
exaltata	Elevata	Betata	vulnerata	Arreata
Tolerata	Sulcata	Revocata	rebata	Parata
Osculata	Notata	Nadata	moscata	Maculata
libata	Acurata	Inclinata	estriata	Disparata
desempata	denticulata	Curvata	corrugata	conturbata
citata	capitata	Aligata	amata	

Grande parte das palavras encontradas não aparecem em dicionários e, algumas aparecem, não são palavras construídas em língua portuguesa e sim importadas do latim é o caso de: *concordata* (do lat. medv. *concordata*, pl. de *concordatum.*); *apostata* (do lat. tard. *apostata*, deriv. do gr. *apostátes* 'desertor (da própria religião)').⁸⁸

Dos autores visitados apenas Bechara (1992) considera *-ata* como formador de deverbais. No dicionário Aulete esse sufixo é definido como designativo de nomes de coletividade, continuidade e extensão, no entanto, não podemos ignorá-lo, apesar de pouco frequente na língua em uso, há ocorrências de deverbais de ação construídos com *-ata*. Além disso, esse sufixo ainda ocorre no português atual: *passeata*, *carreata*, etc..

⁸⁸ Dicionário Aulete digital: <http://aulete.uol.com.br/>

O sufixo *-ncia* é o sétimo mais frequente no *corpus*. Citados por 3 autores dos quatro investigados. Vejamos as ocorrências dos deverbais terminados em *-ncia* no português histórico, na Tabela 36.

Tabela 37: Deverbais de ação construídos com o sufixo -ncia

distancia	ignorancia	observancia	abundancia	jactancia
instancia	constancia	importancia	repugnancia	melancia
vigilancia	sustancia	tolerancia	arrogancia	estancia
mercancia	ervancia	circunstancia	vacancia	consonancia
implicancia	extravagancia	dissonancia	discrepancia	superabundancia
relevancia	exuberancia	exorbitancia	concordancia	Dancia
assustancia	balancia	discordancia	entrancia	exabundancia
jatancia	comandancia	dependencia	traficancia	

Os deverbais terminados em *-ncia* apresentam algumas formas com alta frequência no português. Vejamos na Figura 25, as estruturas morfológicas encontradas para os deverbais sufixados em *-ncia*.

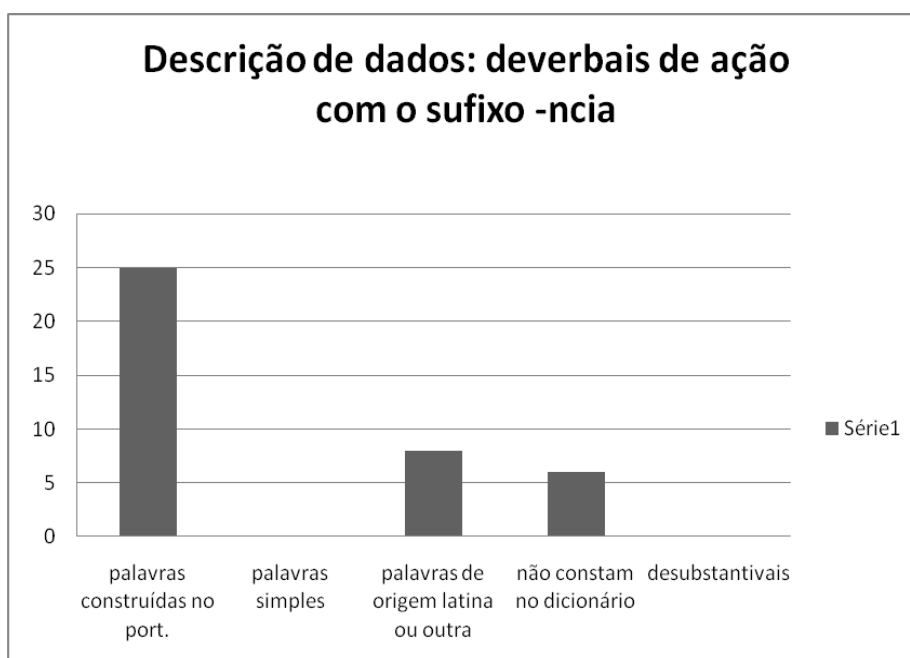


Figura 24: Gráfico das estruturas de deverbais de ação mais frequentes construídas com o sufixo -ncia

A maioria dos deverbais de ação sufixados em *-ncia* são palavras construídas do português. Uma pequena parcela são palavras importadas de outras línguas, especialmente do latim.

O oitavo sufixo com maior número de ocorrências de palavras distintas (29) é o sufixo *-são*, que como podemos observar na Tabela 37, são variações de grafia do sufixo *-ção*.

Tabela 38: Deverbais de ação construídos com o sufixo -são.

Avaliação	Quituação	Negociação	Ordenação	Carregação
embarcação	Povoação	Publicação	Administração	Cominação
declaração	Retificação	Purificação	Deixação	Enformação
obrigação	Separação	Conversação	Amasão	Entalção
procuração	Variação	Demarcação	Aumentação	
Doação	Salvação	Notificação	Nomeação	

O sufixo *-(u/a)me* aparece com 23 ocorrências no português histórico. Observemos na Tabela 38 as unidades lexicais extraídas do *corpus* para esse sufixo.

Tabela 39: Deverbais de ação construídos com o sufixo -(u/a)me .

vexame	Azedume	arume	Queixume	poleame
bicame	inflame	maçame	obstame	pastame
restame	segurame	animame	cavername	constame
mandame	velame	orname	certame	lembrame
ditame	gravame	arame		

Após a análise das ocorrências encontradas, verificamos que apenas *queixume* (de *queixar*) e *vexame* (de *vexar*) são palavras construídas em português. *Ditame* e *certame* foram importadas do latim; *gravame* também é importada do latim, porém segundo seu verbete, não é uma palavra construída, e sim uma palavra simples. As ocorrências restantes são palavras construídas em português, mas não a partir de bases verbais (*azedume* de *azedo*; *velame* de *vela*). E, por fim, a maioria das ocorrências, são das palavras que não existem no dicionário,

às quais podemos atribuir dois tipos de justificativa: homografia e equívoco dicionarístico. No caso desse sufixo, parece-nos que o pronome oblíquo uniu-se ao verbo como ocorre na língua espanhola. Esse fenômeno era bastante comum no português histórico.

Logo após, o sufixo *-deira* aparece com 10 ocorrências, mostradas na Tabela 39.

Tabela 40: Deverbais de ação construídos com o sufixo -deira

escumadeira	vedadeira	resfriadeira	espumadeira	chamadeira
choradeira	tiradeira	lavadeira	purgadeira	amadeira

Escumadeira (e *espumadeira*⁸⁹), *tiradeira*, assumem o sentido de instrumento da ação de *escumar* e de *tirar* respectivamente. *Resfriadeira* assume o significado de local da ação de *esfriar*; *choradeira* aparece no corpus como ação de *chorar* e também faz referência às mulheres que costumam chorar diante dos mortos. Semelhante é o comportamento do deverbais de ação que aparece como instrumento da ação de *lavar* (*lavadeira*) e também faz referência às mulheres que exercem a ação de *lavar*. *Purgadeira* aparece apenas como característica referente à escravas que trabalhavam no engenho de açúcar. A única ocorrência que não é um deverbais de ação é *amadeira*. Essa forma é um exemplo do resultado de fenômenos típicos de *corpus* histórico (RYDBERG-COX, 2003) como a junção de palavras (*amadeira* = a madeira).

Os dois últimos sufixos aparecem com a mesma ocorrência de formas distintas. Contudo, as palavras terminadas em *-nsia* também aparecem terminadas em *-ncia*, o que nos leva a pressupor uma coexistência de grafias para esses deverbais (Tabela 40).

Tabela 41: Deverbais de ação construídos com o sufixo -nsia.

mercansia	Estansia	instansia	sustansia	importansia
-----------	-----------------	-----------	-----------	-------------

Dos terminados em *-dela* sobre a base do participio, apenas *sacadela* tem entrada no dicionário, sendo esta, uma palavra construída no português. Podemos concluir que a combinação de *-dela* à bases verbais não era tão comum no português histórico dada a baixíssima ocorrência de formas distintas (Tabela 41).

Tabela 42: Deverbais de ação construídos com o sufixo -dela

sacadela	Bobadela	rodadela	aranhadela	arranhadela
----------	----------	----------	------------	-------------

⁸⁹ (**es.cu.mar**) v. 1. O mesmo que *espumar* [F.: *escuma* + *-ar*². Hom./Par.: *escuma* (fl.), *escuma* (sf.); *escumas* (fl.), *escumas* (pl. de *escuma*)(AULETE online: <http://aulete.uol.com.br>)

Partimos de um conjunto de sufixos “completo” a fim de evitar a perda de ocorrências dos possíveis deverbais de ação. E constatamos que todos pertencentes ao conjunto podem assumir o papel de um verbal de ação, ou seja, estão disponíveis para RCP de deverbais de ação.

O sufixo *-ria* foi descartado da análise por apresentar homografia com verbos no futuro do pretérito e por estes serem bastante frequentes no *corpus*. Observemos a tabela abaixo:

*Tambem me parece que Mestre João **aproveitaria** cá muito, porque a sua lingoa hé semelhante a esta, e mais aproveitar-nos-emos cá da sua theologia.*

Aqui, *aproveitaria* ocorre como verbo intransitivo no futuro do pretérito que concorda com o sujeito *Mestre João* e tem como complementos o advérbio *muito*, que intensifica o sentido do verbo, e o advérbio de lugar *cá*.

*(...) grande número de gente da fertilidade do dito rio pela abundancia dos seus peixes e da grande **mataria** para a cultura das sementeiras de milho, feijão e trigo.*

Nesta passagem *mataria* é um nome derivado de nome, portanto é um desubstantival de mata – *mata- -ria* – que significa uma grande quantidade de mata. Na gramática de Bechara esse sufixo aparece como formador de substantivos com o sentido de lugar e abundância, aglomeração e coleção. De fato, ele não aparece como um possível formador de deverbais de ação, porém como classificaríamos as seguintes ocorrências abaixo?

Outro verbal de ação encontrado nessa busca foi *roçaria*. Observemos o excerto abaixo:

*esta providência não podem bem governar-se os seus povoadores, porque lá mais do que em outra parte do mundo serem necessários os operários para os trabalhos, que temos dito das **roçarias** dos matos; e da navegação, preciso é dar-se algu? a providência (...)*

Roçaria pode ser parafraseado por ato ou efeito de roçar. Observemos a Tabela 42 com algumas das unidades lexicais terminadas por esse sufixo.

Tabela 43: Deverbais de ação construído com o sufixo *-ria*.

imaginaria	Levaria	Almotaçaria	evitaria	urinaria
Cantaria	Frontaria	Resultaria	roçaria	Cuidaria
Entraria	Lançaria	Chamaria	encontraria	Curaria
Pregaria	Obraria	Experimentaria	procuraria	duraria
Tornaria	Arbitraria	Gritaria	trataria	Estimaria
Acabaria	Tiraria	Importaria	vulneraria	facilitaria
Faltaria	Tomaria	Mostraria	Aconselharia	formaria
Zombaria	Vacaria	Ajudaria	entregaria	

Provavelmente a grande maioria das palavras terminadas em *-ria* encontradas no corpus são verbos. Apesar de já sabermos que *-ria* é muito produtivo na construção de desubstantivais locativos (*verduraria*, *bicicletaria*, etc.), não poderíamos deixar de considerá-lo um possível sufixo formador de deverbais de ação, já que ele também ocorre como tal. Assim, dada a baixíssima frequência de *-ria* na construção de deverbais de ação, podemos compreender o fato de não encontrarmos este sufixo no paradigma de deverbais de ação em nenhuma gramática ou dicionário.

Os sufixos *-ção* e *-mento*, assim como na atualidade, são os mais frequentes na língua e que se unem a uma maior diversidade de bases verbais. São também os únicos que não participam de outra RCP. Talvez a transparência resultante da união desses dois sufixos a suas bases, possa ser uma das razões principais para a preferência deles em detrimento de outros para expressar nominalmente a ação verbal.

Por serem altamente produtivos no português, muitos pesquisadores investigam *-mento* e *-ção*. Podemos citar o trabalho *Formação e uso da nominalização deverbais sufixal no português falado* de Margarida Basílio (1996) no qual a autora faz um levantamento da frequência de ambos os sufixos. O sufixo *-ção* representa 60,2% do *corpus* todo, o que constitui um teor de ocorrência três vezes maior que o segundo sufixo mais usado *-mento*, com 73 ocorrências (20,1%) :

Dentre os demais sufixos temos -nc(i)a, com 48 ocorrências (13.2%) -da com 18 (4.9%) e agem com 5 (1,4%). **Esses resultados indicam condições nitidamente superiores de produção para o sufixo -ção na língua falada. Entretanto, as razões para estas condições devem ser pesquisadas com maior cuidado, sobretudo no que tange a oposição -ção e -mento⁹⁰.**(p.26)

Se realizássemos uma comparação com os valores dos nossos dados, constataríamos que a ocorrência de -ção no corpus histórico é, praticamente, três vezes maior que o segundo sufixo -mento assim como no português contemporâneo. Esse é um dado interessantíssimo porque revela que a relação de frequência entre eles é a mesma desde o período de formação da língua portuguesa do Brasil.

Não sabemos o que determina essa regularidade e não é o foco deste trabalho tal investigação. Contudo, não é irrelevante fazermos uma reflexão sobre esse fenômeno, já que muitos linguístas vêm tentando compreender o que é significativo para escolha sufixal no momento da produção do discurso dado que ambos são funcionalmente iguais.

Basílio com o intuito de estabelecer algum fator determinante para escolha de um em detrimento do outro sufixo, iniciou uma investigação sob a perspectiva semântica e não obteve nenhum resultado convincente. E, então, ela partiu para a morfologia chegando a algum dado considerado interessante pela autora.

Acreditamos que respostas para questões como essas estejam mais próximas da morfologia e da prosódia (as nasais, por exemplo, são mais difíceis de serem articuladas do que as fricativas como ção – /s/). Com essa constatação podemos excluir fatores sociolinguísticos, por exemplo, pois ao longo de cinco séculos as proporções de uso não se alteraram. Todavia, essa é mais uma hipótese que poderia ser levada a cabo em estudos posteriores.

A maioria dos sufixos manteve suas proporções de frequência ao longo dos anos. Porém o sufixo -(t/d)ura foi o único que teve queda representativa no uso. No português histórico ele era bastante usado como deverbais de ação como ato ou efeito de X (verbo base), hoje em dia ele aparece mais como nomes de qualidade desubstantivais e quando aparece como deverbais de ação, geralmente, está relacionada com língua de especialidade e não com a língua geral (ex.: a varredura da superfície do metal foi realizada.).

⁹⁰ Grifo meu.

5.2.2. Conversão

O processo de conversão, por ter regras capazes de gerar muitas formas que são homógrafas na língua portuguesa, dificulta a contagem das palavras que são de fato deverbais de ação resultantes de conversão.

Foram encontradas 6.287 formas distintas que podem ser palavras construídas por conversão e 2.677 variantes. Dada a enorme quantidade de dados, realizamos um recorte dos deverbais de ação que apresentaram frequência igual ou maior que 100. Vejamos a Tabela 43.

Tabela 44: Deverbais de ação resultantes do processo de conversão

2734 falta	2647 caminho	2257 trabalho	1947 respeito	1804 come
1420 governo	1201 duvida	925 corre	890 rumo	874 prova
736 embargo	732 devo	697 sustento	686 uso	583 cura
559 perda	558 agrado	551 cria	542 caça	540 quero
525 levante	521 castigo	484 pratica	480 estima	475 mando
473 marcha	451 despacho	422 desejo	417 risco	397 conquista
394 venda	389 aviso	383 demora	379 demarca	372 trato
352 procura	339 empenho	336 curso	318 socorro	308 queixa
281 conversa	266 diferença	264 soma	262 gasto	259 lucro
259 custo	257 derrota	257 canto	253 erro	252 compra
250 cumprimento	240 encontra	308 pago	307 discurso	303 engano
292 interesse	292 batalha	288 espera	336 curso	335 vira
325 emprego	239 fortifica	238 confirma	238 busca	234 testemunho
230 voto	226 acerto	223 encontro	223 desgraça	219 movimento
218 salto	216 lida	216 estrago	209 visita	208 reparo
196 marco	195 cultivo	191 fala	190 esforço	190 corte
187 pergunta	184 demanda	178 acordo	174 pouso	169 renda
167 concerto	154 aumento	151 mistura	151 dote	147 seca
146 reforma	144 transporte	144 arremata	142 cerco	140 desprezo
140 cita	139 vende	139 receita	138 partilha	138 inveja
138 fico	134 magoa	131 auxilio	129 espanto	128 malicia
128 estudo	128 consumo	127 disciplina	126 passe	125 documento
125 descanso	125 documento	125 descanso	123 remo	121 desamparo
119 argumento	119 alento	115 marca	111 abrigo	110 volume
107 nado	105 registro	101 triunfo	100 suspeita	100 solto
100 suspeita	100 solto	100 escolha		

Foram encontradas 128 deverbais de ação distintos resultantes de conversão. Muitos dos deverbais encontrados podem ser parafraseados por ato ou efeito de X e resultado da ação de X. Dada a quantidade de palavras encontradas, podemos inferir que o processo de conversão é bastante produtivo no português histórico e se compararmos com as formas distintas construídas por sufixação acima de frequência 100, temos 87 substantivos sufixados contra 128 resultantes de conversão. O que pode indicar superioridade do processo de conversão na língua em uso em relação ao de sufixação.

5.2.3. Desflexionação

Foram encontrados 2.822 deverbais de ação resultantes da desflexionação do infinitivo, e 4.422 participípios. A desflexionação caracteriza-se por não apresentar variação na forma, por isso muitas das ocorrências podem ocorrer na forma verbal ao invés da nominal. Abaixo, apresentamos a Tabela 44 e 45 com os deverbais resultantes de desflexionação de infinitivo e de desflexionação de participípio respectivamente que ocorrem no *corpus* com frequência superior a 100.

Tabela 45: Deverbais resultantes do processo de desflexionação de infinitivo

791 conhecer	753 escrever	745 perder	724 tornar	696 matar
686 seguir	665 lanar	6622 dar	658 tratar	652 morrer
641 chamar	638 ouvir	619 continuar	618 andar	610 entregar
605 entender	580 mostrar	574 levantar	571 esperar	571 acabar
570 guardar	547 vender	545 querer	540 correr	5382 poder
537 conseguir	535 assucar	530 evitar	517 declarar	500 descobrir
499 viver	488 conservar	486 procurar	484 curar	483 recolher
465 trazer	458 falar	457 beber	457 abrir	451 faltar
448 acudir	442 executar	432 vencer	429 defender	426 cumprir
425 ajuntar	423 sair	423 formar	419 sustentar	408 constar
405 partir	400 trabalhar	400 povoar	385 cortar	382 examinar
377 obrar	376 ler	372 subir	369 dever	368 ordenar
364 tocar	363aproveitar	3560 haver	345 comprar	3453 dizer
335 prender	329 encontrar	328estabelecer	326 fugir	324 julgar
321 fundar	319 carregar	315 observar	3151 ver	312 voltar
312 colher	310 persuadir	310 conduzir	309 contar	303 navegar
298 reduzir	292 ajudar	291 impedir	288 descer	282requerer
280 retirar	280 conceder	279 assistir	278acompanhar	277 prover
274 livrar	271 governar	270 nomear	270 lavar	267praticar
266 usar	266 mudar	2668 mandar	265 embarcar	265cultivar
263 ensinar	257 dormir	252 informar	252 cobrar	250 obrigar
248responder	248 plantar	244 determinar	238 valer	233 crer
231 remediar	229 satisfazer	229 salvar	226 repartir	224 publicar
2247 ir	223 visitar	223 castigar	222 jantar	219concorrer
218confessar	215 perguntar	215 merecer	212 padecer	212 largar
206pertencer	205 sentir	205 regular	2044 vir	203 assinar
201 temer	2019 tomar	2013 ficar	199 render	198 remeter
198 gozar	197 empregar	196 reconhecer	195 parar	193 negar

192 mover	190 deitar	189 advertir	188 exemplar	187 encher
185 obedecer	185 nascer	1857 estar	183 purgar	182 segurar
181 ganhar	181 cobrir	1762 saber	175 produzir	175 cantar
174introduzir	1732 tirar	172 repetir	172 provar	171 durar
170 romper	170 chorar	168 considerar	166 concluir	165 cuidar
165 celebrar	163 gastar	1639 passar	162 marchar	161 fabricar
160 juntar	159 notar	158 exercitar	157 lembrar	157averiguar
156 lavar	156 erigir	156 consentir	155 eleger	154 resistir
1544 pagar	153 admirar	152 animar	151experimental	150 pescar
149 escapar	149 converter	148 penetrar	1483 achar	1473 parecer
1464 entrar	145 fiar	144 assentar	143 resolver	143facilitar
143 duvidar	143 botar	141 ervar	141 criar	141 apartar
140 crescer	138 resultar	138 reparar	138 diminuir	137 prazer
136 destruir	136 aceitar	13607 ser	135 desfazer	134principiar
134 minerar	133 pegar	132 aprender	130 cair	130 acontecer
1304 chegar	129 perceber	129 lograr	129 adquirir	128 quebrar
1284servir	127 queimar	126 referir	126 atalhar	125 perdoar
125 pelejar	125 demandar	124 bastar	123 moer	122 jurar
121conquistar	121 bater	119 deferir	118 pregar	118 dilatar
1184 levar	117 promover	117 enterrar	117 adiantar	1164 comer
1164 buscar	115 vagar	113representar	113 recorrer	113 explicar
112 restituir	111 causar	109 fechar	109 alimpar	107 montar
107 arrancar	1071 deixar	106	105 estender	105 descrever
104 vestir	104 escolher	10364 fazer	102 tentar	102 suceder
101 imitar	100transportar	100 conter	100 cessar	100 casar
100 aparecer				

Tabela 46: Deverbais resultantes do processo de desflexionação de participio

938 servido	922perguntado	909 entrada	880 lido	846 estrada
744 obrigado	715 armada	692 nomeado	691declarado	691 comprido
686 chegado	625 soldado	558 agrado	551 referida	545 recebido
513 ordenado	485 partido	439 chegada	438 tido	435 tomado
428 situada	4214 revisado	401 povoado	388 perdido	386 passada
374 ouvido	369 conhecido	340 tratado	333 ferida	328determinado
326 partida	315 segredo	312 tirado	310 provido	300conhecida
299experimental	299 acabado	298 treslado	286 medida	286 achado
277 acompanhado	273 madrugada	273 informado	268 devida	264 mostrado
263 encarregado	258 deixado	256 casado	254 comprida	247 fundada
244 sentido	231 dilatada	229 vestido	227 cedo	226 acabada
224 morada	219 devido	213 levado	212 retirada	211 formado
204 registada	203 havido	203 descuido	201 povoada	201 picada
199 tirada	197mencionado	196 pedido	191levantado	189 metido
187 formada	187 citado	186 sagrado	186 sagrada	184 datada
184 careado	1832 chamado	1812 estado	178entendido	178 agrada
173 tomada	173 bebida	172 dilatado	170 andado	167 criado
165 nascido	165 ficado	159 amado	157 arrecada	156 salgado
154 sepultado	154 privado	152 ferido	152declarada	151 honrado
150 pescado	149 fundado	149 contado	144 sobrado	143 vencido
143 obrado	142 lembrado	142denunciado	142 bocado	138 arado
137 entrado	137 concedida	135 fiado	134 machado	134destinado
132 mencionada	132 concedido	131 perdida	1318 dado	129carregado
129 avaliado	128 duvido	127 traslado	125estabelecido	124praticado
124 estimado	124 cunhado	124 cuidado	1242referido	123empregado
122 alada	120 ruido	120 bispado	1201 duvida	119 rematado
119 observado	119 jurada	119 cercado	119assentado	1191cuidado
118 reduzido	1182 mandado	117 seda	116 vendido	116requerido
116proporcionada	115 limitado	115 fugida	1154 passado	114retirado
113 sabido	113 acertado	112 situado	112 separado	111 estacada
110convencido	109 persuado	109continuado	108 trazido	107 nado
107 expedida	106persuadido	1061 chamada	105 salgada	1055 sido
104 armado	103 preparado	101 recolhido		

Temos muito mais ocorrências de formas resultantes de desflexionação do que Sufixação e Conversão. Entretanto, reforçamos o caráter ambíguo da função que tais palavras

exercem no discurso, dada a questão da homografia entre a forma verbal e nominal. Portanto, classificamos as palavras extraídas pela regra de desflexionação como possíveis deverbais.

Do ponto de vista da produtividade, ou seja, da disponibilidade de regras para construção de deverbais, o processo mais produtivo seria a sufixação, já que para toda base verbal existem pelo menos 12 sufixos diferentes e, por conseguinte, diversas possibilidades de combinação. A desflexionação viria em segundo lugar com cinco possibilidades e por último a conversão com três.

5.3. Deverbais de ação dos séculos XVI, XVII e XVIII à atualidade

Nesta Seção apresentamos alguns deverbais de ação que, a partir da observação da listas geradas a partir do *corpus*, apresentaram formas distintas das atuais⁹¹ e por isso despertaram a atenção. A distinção está relacionada a mudanças de sufixo ou em relação aos processos de construção. A Tabela 46 apresenta deverbais de ação com alteração de sufixo em relação ao português contemporâneo.

Tabela 47: Deverbais de ação que sofreram alteração sufixal -ção/ -mento.

<i>relaxação</i>	<i>congelação</i>	<i>juração</i>
<i>curação</i>	<i>nivelação</i>	<i>refinação</i>
<i>equipação</i>		

Os deverbais de ação supracitados podem parecer estranhos (embora passíveis de compreensão) ao falante do português contemporâneo, pois, atualmente, o sufixo eleito para essas palavras é o *-mento*. Observemos as formas históricas (1ª coluna) e as contemporâneas (2ª coluna):

<i>relaxação</i>	→	<i>relaxamento</i>
<i>curação</i>	→	<i>curamento</i>
<i>congelação</i>	→	<i>congelamento</i>
<i>nivelação</i>	→	<i>nivelamento</i>
<i>juração</i>	→	<i>juramento</i>
<i>equipação</i>	→	<i>equipamento</i>

No *corpus* foram encontradas somente ocorrências desses deverbais com o sufixo *-ção*. A partir disso podemos hipotetizar que a alteração na forma (em especial na mudança de sufixo), ocorreu posteriormente aos séculos aqui estudados, ou talvez a forma em *-mento* circulava na língua falada.

⁹¹ Para uma análise mais consistente sobre a frequência das formas selecionadas para análise em relação ao português contemporâneo, seria necessário uma busca em um corpus contemporâneo, como por exemplo o CentenFolha (www.linguateca.pt). As análises que seguem são baseadas na intuição e conhecimento linguístico.

Na Tabela 47 podemos encontrar deverbais construídos com o sufixo *-ção* (processo de sufixação) no português histórico e que hoje, são comumente encontrados construídos por outro processo – a conversão.

Tabela 48: Deverbais de ação históricos que atualmente são construídos por processo de conversão.

<i>exageração</i> 12	<i>protestação</i> 13	<i>enterramento</i> 12
<i>dipensação</i> 5	<i>cultivação</i> 2	<i>comandância</i> 2
<i>murmuração</i> 10	<i>reservação</i> 1	<i>comandamento</i> 2
<i>reformação</i> 25	<i>governação</i> 1	<i>sustentamento</i> 1
<i>trespassação</i> 2		
<i>vendagem</i> 3		

Na Tabela 47 podemos observar deverbais construídos por sufixação e que atualmente são mais (ou somente) recorrentes por conversão. Vejamos os deverbais de ação equivalentes construídos por conversão:

exageração → *exagero*
dispensação → *dispensa*
murmuração → *murmúrio*
reformação → *reforma*
protestação → *protesto*
cultivação → *cultivo*
reservação → *reserva*
governação → *governo*
trespassação → *trespasse*
encerramento → *enterro*
comandância → *comando*
sustentamento → *sustento*
vendagem → *venda*

Para buscarmos explicações para essa sobreposição de deverbais gerados por conversão sobre a sufixação procedemos a uma segunda etapa de análise: busca por deverbais

de ação construídos por conversão e suas respectivas frequências. A partir desses dados pudemos obter uma pista do porquê estes nomes deverbais por sufixação não são mais usados (ou são menos usados). Observemos os dados abaixo:

<i>exagero</i>	1	<	<i>exageração</i>	12
<i>dispensa</i>	41	>	<i>dispensação</i>	5
<i>múrmurio</i>	2	<	<i>murmuração</i>	10
<i>reforma</i>	0	<	<i>reformação</i>	25
<i>protesto</i>	79	>	<i>protestação</i>	13
<i>cultivo</i>	195	>	<i>cultivação</i>	2
<i>reserva</i>	53	>	<i>reservação</i>	1
<i>governo</i>	1420	>	<i>governação</i>	1
<i>trespasse</i>	1	>	<i>trespassação</i>	2
<i>enterro</i>	43	>	<i>enterramento</i>	12
<i>comando</i>	32	>	<i>comandância</i>	2
			<i>comandamento</i>	2
<i>sustento</i>	697	>	<i>sustentamento</i>	1

Como podemos observar apenas para a unidade lexical *reformação* não foi encontrado um deverbais de ação equivalente construído por conversão. Ambas as formas existiam no português histórico, já com uma notória soberania em relação à frequência dos deverbais de ação construídos por conversão. Somente os deverbais construídos por sufixação *exageração* e *murmuração* tiveram ocorrência superior.

A Tabela 48 apresenta formas pouco (ou não) utilizadas no português contemporâneo. Nos casos abaixo, o estranhamento não está relacionado somente aos sufixos eleitos mas também às bases verbais sobre as quais foram construídos os deverbais

Tabela 49: Deverbais em com baixa ou nenhuma frequência no português contemporâneo.

<i>pastoramento</i>	<i>varaço</i>	<i>roçaria</i>
<i>pastoração</i>	<i>detestação</i>	<i>filhamento</i>
<i>granjeamento</i>	<i>roçada</i>	

Os deverbais acima, atualmente e aparentemente, não são construídos por nenhum dos processos de construção de deverbais disponíveis . A explicação para esse fenômeno pode estar relacionada a questões extralinguísticas.

Deverbais como *pastoramento*, *pastoração*, *granjeamento*, *roçadura*, *roçaria* são termos relacionados à atividades agrárias e talvez ainda ocorram em determinados ambientes. São termos que são gerados devido à necessidade de nomear processos e seres que existem na língua num determinado período de tempo. Por exemplo, com certeza não encontraríamos a palavra *esmaltamento cerâmico* no *corpus* histórico, já que este verbal nomeia um processo recente característico de todo o desenvolvimento da área de materiais.

Observemos alguns excertos em que aparecem essas palavras:

*Assim passa pela maior parte, porque os naturais da terra se ocupam no **granjeamento** dos seus engenhos e no benefício de suas lavouras, sem quererem tratar de mercancias (...)*

*fizeram a parte separados melacias, e melões, como todos eles, e searas de legumes já aos 3 meses principam a desfrutar-se, pagando com boas colheitas a seus donos o grande trabalho da **roçaria**, e são estas frescuras tanto mais regaladas no Amazonas, quanto mais ardentes os seus calores, e todo o ano há estas frescuras havendo qualquer leve diligência de os semear, e conservar.*

*Vi eu algum novilho, a quem as cordas, que o seguravam pelas pontas, junto destas fizeram uma **roçadura**, em que por inadvertência saltou uma bicheira, que em poucos dias o comeo todo, e parecia o animal, que todo ele se convertera em bichos.*

Roçaria remete a ação de roçar, cortar, friccionar, já *roçadura* é o efeito da fricção, do roçar. Alguns outros deverbais suscitam nossa curiosidade, por exemplo *filhamento*, de *filhar* construído por sufixação, com apenas duas ocorrências no *corpus*, observemos:

*Esta nobre família de Almeidas Castanhos da capitania de S. Paulo traz a sua origem da villa de Monte-mor o Novo em Portugal, de onde veio para S. Vicente pelos annos de 1547 Antonio Rodrigues de Almeida, e tinha o foro de cavalleiro fidalgo da casa do Sr. rei D. João III, em cujo reinado foi este foro de cavalleiro fidalgo o mais superior que constituia grao de fidalguia, até que alterou a ordem dos **filhamentos** o Sr. rei D. Sebastião, de cujo tempo até o presente ficou este foro de cavalleiro fidalgo, sendo infimo; de sorte que o mordomo-mor do reino o confere ás pessoas mecanicas para passarem com elle ao primeiro grao de nobreza; e o foro de fidalgo cavalleiro ficou sendo **filhamento** superior com 1\$800 de moradia (...)*

O trecho nos dá pistas de ser esse um termo utilizado nas relações de vassalagem, específico do período feudal. Esse termo confirma a influência portuguesa no período de formação do PB.

Vendagem (3 ocorrências) que vem de *vender*, assim como *venda* (por conversão com 394), que atualmente é a forma disparada mais utilizada na língua, também era o deverbais mais recorrente para designar ação de vender. Esse é mais um caso de coexistência de duas formas para a mesma função e que aqueles formados por conversão prevaleceram. Vejamos a passagem em seguida:

*Quanto ao logar para **vendagem** do peixe á Preguiça, ainda lá está desafiando a critica que Vilhena já fazia ha 120 annos, aquella pessima feira, pois não ha mercado.*

*Tambem acompanha esta a conta de **venda** das coatro pipas dagoa ardente que foi **venda** como o tempo o premetia, e inda se me deve maior parte dellas;*

Pesquisando as ocorrências dos deverbais de *venda* encontramos muitas com o sentido de local onde se vende algo, diferentemente do que dizem muitas gramáticas que o sufixo formador de lugar é o *-ria*.

*Da mesma sorte, serão obrigadas a manifestarem-se todas as pessoas que tiverem *lojas, '**vendas**, // boticas, cortes de carne, pagando a capitação que lhes foi repartida, a saber, doze oitavas as lojas grandes, oito as medianas e as vendas, e quatro os mascates e lojas pequenas, de que se fará assento em livro separado e se darão bilhetes de paga aos que vierem manifestar-se*

Outra palavra interessante encontrada foi *detestação* do latim *detestation*. Não foram encontradas outras formas construídas dessa palavra no *corpus*. Encontramos quatro passagens com essa forma, das quais reportamos duas abaixo:

*Alli sendo nós mesmos os réos e os accusadores, confessamos espontaneamente todas nossas culpas: e se o fazemos com a verdadeira **detestação** e arrependimento que devemos a um Deus infinitamente bom, e infinitamente offendido (...)*

*E como a confissão verdadeira incluye essencialmente a detestação dos peccados commettidos, e resolução firme de nunca mais peccar; com a **detestação** abriu a porta ao perdão dos peccados passados, e com a resolução fechou a porta á continuação dos futuros.*

A Tabela 49 apresenta deverbais de ação históricos construídos por sufixação e que atualmente são muito mais frequentes por desflexionação de participio.

Tabela 50: Deverbais de ação construídos por sufixação no português histórico.

<i>paragem</i> 410	<i>paramento</i> 2	<i>saimento</i> 1
--------------------	--------------------	-------------------

Enquanto *paragem* ocorre 410 vezes, o deverbais *parada* que é o mais recorrente no português contemporâneo ocorre 38 vezes no *corpus* histórico. Observemos as passagens abaixo:

*(...) marchando de dia, & de noite, com pouco descanso, até que achem **paragem**, aonde possaõ parar.*

*Deo o mesmo Gentio Payagoa neste anno no sitio de Joaó de oliveira na **paragem** de Paragoai que era hum formozo Arrayal de muytos moradores ali estabelecidos (...)*

Paragem é usada com um dos sentidos que hoje também utilizamos: *parada* local onde se pára. O deverbais de ação construído por sufixação deu espaço a forma desflexionada do participio passado *parada*. Por existirem também ocorrências de *parada* no *corpus* analisamos alguns excertos abaixo:

*(...) porque neste genero de doença os póros estão fechados, e a circulação **parada**, ou retardada; e assim que abrindo os póros, e fazendo-se a circulação com os remedios assima (...).*

*(...) e o Cadette com as ordens p.[^]a fazer estabeleser as **paradas** da marinha da Villa de S.Luiz de Guaratuba até a da Conceição de Itanhaê.*

No primeiro excerto *parada* aparece como verbo na forma do participio passado, com o verbo auxiliar oculto *está parada*. Já na segunda ocorrência temos um deverbais de ação mas, diferentemente de *paragem* que pode também designar local onde se *para* (devido a polissemia deverbais que já foi mencionada), aqui *parada* significa ato ou efeito de *parar*.

*(...) que a não ter tantos desvios, seria hoje mais populosa, que a mesma cidade. O que nela mais se admira é o seu lindo templo edificado à moderna com belas vias sacras, espaçosa sacristia, e com **paramentos** muito ricos;*

Tem ordinariamente lindas igrejas nas suas missões, com bons ornamentos, e paramentos para os ofícios Divinos que não tem inveja aos templos das cidades com 3, ou mais sinos, feitas ordinariamente de madeira escolhida, retábulos bem feitos, e em algumas dourados,(...)

Encontramos ainda a forma *paramento* a qual ocorre duas vezes no *corpus* e significa um tipo de ornamento e não é uma palavra construída no português e sim originária do latim.

Enquanto *paragem* ocorre 410 vezes, *parada* ocorre 38. *Paragem* era muito mais frequente no português histórico que *parada*. Atualmente, parece-nos que o deverbais *parada* resultante da desflexionação do verbo no particípio incorporou também o sentido produzido por *paragem*. A semelhança dos significados pode ter causado um enxugamento lexical, desse modo, o processo de construção por sufixação parece ter perdido força. Uma investigação mais detalhada pode ser realizada em um trabalho futuro.

Outro fenômeno semelhante deu-se com as formas deverbais *saimento* e *saída*. Observemos as passagens abaixo:

o saimento das exequias que se celebraraõ na Sé desta Cidade (Pg. 144) o alivio que temos em ser governados por Vossa Magestade (...)

Expedição pertencente a Sua Magestade e não mostrando elle a saída de tudo Com as clarezas, que deve, o remetterá para a Fortaleza desta Villa, lhe mandará fazer sequestro em todos os bens athé se fizerem as contas, e ver ao que está responsavel á Fazenda Real.

Metteu-se no bosque, e tardando muito na saída d'elle, temeram os indios não tivesse acontecido algum repentino accidente ao seu superior.

Encontramos apenas uma ocorrência de *saimento* e 39 de *saída*. Com a análise dos excertos observamos que ambas eram usadas no sentido de ato ou efeito de sair, porém há de se analisar maior quantidade de excertos para uma constatação dos possíveis sentidos assumidos por esse deverbais. Já no português histórico a forma que utilizamos hoje – por desflexionação – era bem mais frequente.

Essa foi uma breve análise de alguns dos fenômenos envolvendo a categoria deverbais de ação no PB. Nosso intuito com esta pesquisa e essa análise foi suscitar, trazer a tona fenômenos interessantes de serem investigados mais profundamente por diferentes tipos de especialistas ou em nossos trabalhos futuros.

6. Conclusão

6.1. Considerações iniciais

A realização deste trabalho inicia o preenchimento de uma lacuna sobre o conhecimento de mecanismos de construção de deverbais no português do Brasil dos séculos XVI, XVII, XVIII e início do XIX. Esta pesquisa só foi possível devido à construção do *corpus* histórico do português do Brasil no projeto DHPB. A construção de um *corpus* com uma grande quantidade de dados, reunidos e em um formato adequado para investigação é uma tarefa trabalhosa e exige tempo. Ter à disposição o *corpus* DHPB foi imprescindível para idealização desta pesquisa de mestrado.

Com os dados disponíveis, o desafio passou a ser outro – o acesso a esses dados de maneira rápida. Seria inexequível procurar no *corpus* as ocorrências de deverbais manualmente por dois motivos: o primeiro, pela imensa quantidade de palavras no *corpus*; o segundo, por não partimos de um fenômeno exato e singular da língua, isto é, não estamos procurando determinada palavra, mas sim uma categoria de palavras construída por vários processos construcionais (o que acarreta diferentes estruturas morfológicas).

A primeira pergunta que surgiu foi: Como proceder diante dessas condições?

Chegamos à conclusão de que precisaríamos de alguma ferramenta que possibilitasse o acesso rápido aos deverbais de ação possíveis. Mas que tipo de ferramenta teria a habilidade de recuperar essa categoria com suas diversas formas? Para respondermos a essas perguntas, primeiramente, precisamos estudar o fenômeno linguístico – deverbais de ação. Somente após o conhecimento completo do objeto de estudo em termos de regularidades e idiossincrasias (objetivo deste trabalho), é que foi possível modelarmos uma ferramenta, ou mais especificamente, um sistema de recuperação semi-automática de deverbais de ação (segundo objetivo deste trabalho). E, por fim, analisar os deverbais de ação históricos do PB.

6.2. Sobre o modelo escolhido: SILEX

O caráter preditivo e hipergerativo foi fundamental para a adoção do modelo SILEX, já que a proposta justamente era revelar as estruturas possíveis e sua distribuição no português histórico. Somente a escolha de um modelo hipergerativo seria adequada para que não perdêssemos todas as possíveis ocorrências de deverbais de ação, já que essa é uma categoria

com alto teor de regularidade, porém (como vimos na Seção 4.5.1) dentro dessa regularidade há mais de 20 combinações possíveis, ou seja, para cada base verbal é possível termos pelo menos 20 estruturas morfológicas diferentes. Dessa maneira, passamos à investigação da categoria dos deverbais e à determinação dos constituintes participantes dessa categoria.

Outro aspecto importante do modelo para análise das estruturas lexicais, é que dado seu sistema estratificado é possível descrever minuciosamente as operações sofridas pelos constituintes lingüísticos até formarem uma palavra construída no português. Dessa maneira, pudemos conhecer quais dos deverbais encontrados, de fato, foram formados na língua portuguesa e quais foram importados para o português.

6.3. Sobre as regras linguísticas

Após uma vasta investigação da literatura sobre deverbais e processos de formação de palavras, encontramos algumas inconsistências, principalmente, em relação aos sufixos formadores de deverbais de ação. Diante disso e do fato de desconhecermos os deverbais que ocorriam no português histórico, consideramos todos os sufixos encontrados, já que por mais ou menos frequentes que fossem, ocorriam como deverbais de ação.

Eleitos os sufixos, demos continuidade às investigações dos processos de construção de deverbais de ação segundo o modelo de morfologia construcional e identificamos dois processos de construção – conversão e sufixação – e um de desflexionação. Salvo o processo de Sufixação, os outros dois trouxeram dificuldades em relação à contagem das ocorrências por apresentarem forma homógrafa em outras categorias.

6.4. Sobre o sistema de recuperação semi-automática – EXTRADEV

Com o fenômeno lingüístico conhecido e tendo sido definidas as suas irregularidades, passamos a pensar uma maneira dinâmica de recuperarmos tais formas no *corpus*. Agora, podemos retomar à pergunta do início desta Seção: como recuperar uma grande quantidade de deverbais de ação em um grande *corpus* de forma simples, precisa e rápida? Recorremos, então, aos recursos computacionais disponíveis para processamento de língua natural e propusemos uma arquitetura para um sistema de recuperação de deverbais de ação de maneira

semi-automática chamado EXTRADEV que inclusive foi implementado computacionalmente e avaliado.

A construção do sistema baseou-se na utilização de duas bases de conhecimento: o dicionário UNITEX-PB no formato DELAF e as formas variantes para um grande conjunto de palavras do próprio *corpus* do DHPB, extraídas pelo SIACONF. Ambas contribuíram para que o sistema ampliasse sua cobertura. O dicionário UNITEX-PB com mais de 14.000 verbos, permitindo a geração da grande maioria dos deverbais possíveis; e a saída do SIACONF que possibilitou que o EXTRADEV fosse também utilizado para um *corpus* contemporâneo, pois utilizando esse conhecimento não precisamos que as regras de variação de grafia fossem inseridas nas regras de deverbais de ação. Assim, vemos a modelagem e implementação do EXTRADEV como uma das contribuições deste trabalho de mestrado.

No estudo piloto, o objetivo era apenas gerar regras no formato de expressões regulares para cada um dos 50 verbos mais frequentes do *corpus* para serem usadas em qualquer sistema que suportasse a sintaxe das expressões regulares. Por isso, no Apêndice A, é possível verificar tais regras que também podem ser úteis dependendo do intuito do pesquisador. Caso o interesse seja apenas recuperar deverbais de ação para um determinado verbo ou para um pequeno conjunto, não é necessário utilizar todo o sistema e sim as regras através do comando ‘grep’ (ver Seção 4.3.2) ou mesmo via um *script* mais simplificado.

A utilização de expressões regulares foi adequada aos nossos objetivos. Como já foi dito, elas são bastante úteis para detecção de padrões de linguagem, entretanto, há de se ter cuidado com o uso de alguns metacaracteres, que muitas vezes, dependendo da linguagem do script, podem ser interpretados de maneiras diferentes ou não serem interpretados.

O *corpus* foi submetido ao primeiro módulo do sistema e obtivemos a primeira lista de deverbais históricos sem variação de grafia. Esta foi confrontada com a lista de variantes gerada pelo SIACONF e, então, a lista final foi gerada. O processamento do sistema durou mais de 2 horas devido à grande quantidade de dados envolvida e ao formato das próprias regras que foram construídas.

6.5. Sobre os dados gerados

Após todas essas investigações e processamentos, obtivemos enfim, os dados para análise. Como já foi dito, nosso objeto de pesquisa – os deverbais de ação – poderia ocorrer com múltiplos formatos. Nesse contexto, esperávamos que os dados pudessem nos revelar

algo em relação aos formatos encontrados (os mais frequentes, menos frequentes, por exemplo), ou seja, aspectos linguísticos interessantes para reportarmos e iniciarmos algumas reflexões.

Em uma breve observação da lista, muitas questões de pesquisa surgiram. Essa multiplicidade de fenômenos possíveis de serem investigados também dificulta a delimitação do tema a ser analisado. Desse modo, primeiramente, partimos de uma análise baseada na frequência: detecção dos processos construcionais e sufixos mais recorrentes no português histórico; análise do tipo de estrutura dos deverbais encontrados (palavras construídas ou palavras não construídas) e realização de uma análise comparativa com o trabalho de Basílio (1996) em relação à frequência de ocorrências no discurso dos sufixos *-ção* e *-mento*. A segunda análise baseou-se na observação das formas que despertaram a atenção por apresentarem construções diferentes das utilizadas hoje no português contemporâneo

Se partirmos do pressuposto que preservar uma determinada construção teoricamente é mais simples do que alterá-la, então, detectar uma mudança em uma determinada construção pode ser um indicativo de que diversos aspectos intervieram, e que alguns tenham tido mais peso, como, por exemplo, os extralinguísticos.

O trabalho de Basílio (1996) em comparação com este revelou-nos que não houve mudança, ou seja, não houve alteração na razão da frequência de sufixos em *-ção* e *-mento* ao longo de mais de cinco séculos.

Em relação à análise da estrutura, focamos os deverbais construídos por sufixação, devido à grande quantidade de dados e à menor obscuridade na detecção do termo. Podemos dizer que a maioria dos deverbais de ação do PB histórico são palavras construídas no próprio português e uma parcela menor, porém, significativa representa as palavras de origem latina importadas já construídas para o português.

6. 6. Considerações finais

Do ponto de vista teórico, este trabalho utilizou-se de enquadramentos teóricos opostos – a linguística de corpus e a morfologia construcional. A linguística de *corpus* pauta suas análises na língua em uso e concebe o *corpus* como um revelador dos fenômenos linguísticos. Por outro lado, a morfologia construcional inserida numa abordagem gerativa de viés semântico, trabalha com as possibilidades de construção lexical baseada na competência dos falantes.

Do ponto de vista da aplicação, as teorias são complementares. Ambas as abordagens foram requeridas em etapas distintas deste trabalho e, além disso, foram fundamentais para a execução completa desta pesquisa de mestrado.

Em relação aos aspectos computacionais e o processamento de língua natural, para podermos utilizá-los, devemos ter primeiramente uma modelagem linguística daquilo que queremos processar, seja para realizar buscas, ou para qualquer atividade que envolva a interpretação da língua pela máquina.

O computador entende e aprende a linguagem humana a partir de padrões, mesmo que estes sejam irregulares. Sempre existirão limitações por parte da máquina no tratamento da língua, porém, mesmo que os ganhos não sejam totais, acreditamos que a computação sempre trará contribuições e facilidades para os linguístas. Desse modo, podemos contar com a computação para dinamizar e facilitar as pesquisas acerca da linguagem humana, aumentando assim a produtividades das pesquisas linguísticas.

6.7. Trabalhos futuros

A realização deste trabalho suscitou diversas possíveis pesquisas futuras, algumas delas estão listadas abaixo:

- colaboração com regras novas para o SIACONF a partir de regras hipergerativas para deverbais de ação;
- construção de um ambiente web para recuperação de unidades lexicais construídas: deverbais de agente, desubstantivais, deadjetivais; estudo comparativo da função textual do verbal de ação em *corpus* histórico;
- análises comparativas do que foi revelado no *corpus* com o que dizem os manuais de gramática, tanto as históricas quanto as contemporâneas;
- aprofundamento nas análises sobre sufixos, processos mais frequentes, coexistência de sufixos, mudança linguística, influência do PE, etc.

Referências

- ALUISIO, S.M., PINHEIRO, G., FINGER, M., NUNES, M.G.V., TAGNIN, S.E. The Lácio-Web Project: overview and issues in Brazilian Portuguese corpora creation. In: CORPUS LINGUISTICS 2003, Lancaster, UK. **Proceedings** of Corpus Linguistics 2003 (Also as UCREL Technical Report, Vol 16 Part). Lancaster: 2003. v. 16, p. 14-21.
- ALUISIO, S., PINHEIRO, G.M., MANFRIM, A.M.P, OLIVEIRA, L. H. M. de, L. C. GENOVES Jr., TAGNIN, S. E. O. The Lácio-Web: Corpora and Tools to advance Brazilian Portuguese Language Investigations and Computational Linguistic Tools. In LREC 2004. **Proceedings** of LREC, 2004, Lisboa, Portugal, p. 1779-1782.
- AULETE, F.J.C., GARCIA, H. - **Dicionário contemporâneo da língua portuguesa**, 3.ed. Rio de Janeiro, Ed. Delta, 1980.
- BASILIO, M. M. P. **Estruturas lexicais do português: uma abordagem gerativa**. Petrópolis: Vozes, 1980. p.128.
- BASILIO, M. M. P. A morfologia no Brasil: indicadores e questões. D.E.L.T.A., Vol. 15, N.º ESPECIAL, 1999 (53-70).
- BASILIO, M. M. P. **Polissemia Sistemática em Substantivos Deverbais**. In: Roberta Pires de Oliveira; Apóstolo T. Nicolacópulos. (Org.). Semantics: Lexicon, Grammar and Use. Florianópolis: Editora da UFSC, 2004, v. 47, p. 49-71.
- BASILIO, M. M. P. **Formação e classes de palavras no português do Brasil**. 2. ed., 1ª reimpressão. – São Paulo: Contexto, 2006.
- BECHARA, E. **Moderna Gramática Portuguesa**. 37. ed. Rio de Janeiro: Lucerna, 1992.
- BAEZA-YATES, R.; RIBEIRO-NETO, B. (1999). **Modern Information Retrieval**. New York: ACM Press, Addison-Wesley.
- BYBEE, Joan. *Morphology. A study of the relation between meaning and form*. Amsterdam: John Benjamins, 1985.
- BIDERMAN, M. T. C. *Teoria lingüística: lingüística quantitativa e computacional*. Rio de Janeiro: Livros Técnicos e Científicos, 1978.
- BIDERMAN, Maria Tereza Camargo. **Teoria linguística**. Teoria lexical e linguística computacional. 2.ª ed. São Paulo, Martins Fontes, 2001.
- BORBA, F. S. **Dicionários de Uso do Português do Brasil**. São Paulo, Ed. Ática, 2002.
- CAETANO MOCHO, Maria do Céu. **Formação de palavras em Gramáticas Históricas do Português**. Análise de algumas correlações sufixais. Dissertação de Doutorado em Linguística, especialidade Morfologia. Lisboa, Universidade Nova de Lisboa, Faculdade de Ciências Sociais e Humanas, 2003.
- CAMACHO, Roberto Gomes. **A função textual dos nomes deverbais**. In Estudos Lingüísticos XXXIV, p. 183-188, 2005 Disponível em: <http://www.gel.org.br/4publica-estudos-2005/4publica-estudos-2005-pdfs/a-funcao-textual-dos-nomes-692.pdf?SQMSESSID=a38ffc79c82bcbe561e1c641326fd16c> Acesso em : jan/2009
- CÂMARA Jr., J. Mattoso. **Dicionário de Filologia e Gramática - referente à Língua Portuguesa**. 4ª ed. Rio de Janeiro. J.Ozon, 1970.
- CÂMARA Jr., J. M. **Princípios de lingüística geral**. 4. ed., Rio de Janeiro, Acadêmica, 1964.
- CÂMARA Jr., J. M. **História e Estrutura da Língua Portuguesa**. 2ª ed. Rio de Janeiro. Ed. Padrão, 1976.
- CÂMARA Jr., J. M. **Estrutura da Língua Portuguesa**. 7ª ed. Rio de Janeiro. Ed. Vozes, 1976.
- CANDIDO, Jr, A., **Criação de um ambiente para o processamento de corpus de Português Histórico**. 2008. Dissertação (Mestrado em Ciência da Computação e

- Matemática Computacional) - Icmc Usp.
- CANDIDO, Jr, A., Aluísio, S. M.. **Procorph: um Sistema de Apoio à Criação de Dicionários Históricos**. In: VI Workshop em Tecnologias da Informação e da Linguagem Humana (TIL 2008), Vilha Velha, v. 1. p. 1-6, 2008.
- COELHO, Juliana Soledade Barbosa Coelho. **A sufixação nominal no português arcaico em sua segunda fase (sécs. XV e XVI)**. Dissertação de Doutorado em Letras, área de Linguística Histórica. Salvador, Universidade Federal da Bahia (inédita), 2005.
- CORBIN, Danielle. **Les bases non autonomes en français ou comment intégrer l'exception dans un modèle lexical**. In *Langue Française*, nº 66, 1985, p. 54-76.
- CORBIN, Danielle. **Morphologie dérivationnelle et structuration du lexique**. 2 vols. Tubinga: Max Niemeyer Verlag, 1987.
- CORBIN, Danielle. **La place de l'histoire dans une morphologie synchronique**. In *Acta Universitatis Wratislaviensis*, nº 1064, 1989, p. 51-67.
- CORBIN, Danielle. **Introduction - La formation des mots: structures et interprétations**. In *Lexique 10*. Villeneuve d'Ascq: Presses Universitaires de Lille, p. 7-30, 1991.
- CORBIN, D. **La représentation d'une "famille" de mots dans le Dictionnaire dérivationnel du français et ses corrélats théoriques, méthodologiques et descriptifs**. In *Recherches linguistiques de Vincennes*, 1997, p. 5-37 + errata.
- CORREIA, M., L. L. S. P. **Inovacao Lexical em Portugues**. Lisboa, Edições Colibri, 2005.
- CORREIA, M. **A denominação das qualidades em português – contributos para a compreensão da estrutura do léxico português**. Dissertação de doutoramento. Lisboa, Universidade de Lisboa, 1999.92
- CORREIA, M. **Homonímia e polissemia contributos para a delimitação dos conceitos. Palavras** (Lisboa), v. 19, p. 57-75, 2001.
- CORREIA, M. **Denominação e construção de palavras**. 1. ed. Lisboa: Colibri, 2004. v. 1. 364 p.
- CRYSTAL, David. **Dicionário de Linguística e Fonética**. Tradução e adaptação [da 2ª ed. Inglesa rev. e ampliada, publicada em 1985], Maria Carmelita Pádua Dias. Rio de Janeiro: Jorge Zahar Ed., 2000.
- CASTILHO, A. ; BASÍLIO, M. (org.). **Gramática do Português falado**. Vol. IV. Campinas, editora da Unicamp, 1996.
- CUNHA, Antonio Geraldo da. **Dicionário Etimológico Nova Fronteira da Língua Portuguesa**. 1ª ed. – Rio de Janeiro: Nova Fronteira, 1982.
- CUNHA, Celso., L. C. **Nova Gramática do Português Contemporâneo**. 5. ed., Lisboa, Edições João As da Costa, 1989.
- CUNHA, Celso., L. C. **Nova Gramática do Português Contemporâneo**. 3ª ed. – Rio de Janeiro: Nova Fronteira, 2001.
- DUBOIS, et. al.. **Dicionário de Linguística**. 6-7-8-9-10 ed. – São Paulo: Cultrix Ltda., 97-98.
- DUNCAN, John. **A Frequency Dictionary of Portuguese Words**, Stanford: Ph.D. Dissertation, 1971
- FLEXOR, Maria Helena Ochi. **Abreviaturas: Manuscritos dos séculos XVI ao XIX**. 2. ed. São Paulo: Editura Unesp, 1991. 468 p.
- FINGER, M. . **Técnicas de Otimização da Precisão Empregadas no Etiquetador Tycho Brahe**. In: V Encontro para o Processamento Computacional da Língua Portuguesa Escrita e Falada (PROPOR,2000), 2000, Atibaia, SP. **Anais do V Encontro para o Processamento Computacional da Língua Portuguesa Escrita e Falada (PROPOR2000)**, 2000.

⁹²Parte desta dissertação foi publicada em Lisboa, pelas Edições Colibri, em 2004, com o título *Denominação e construção de palavras*. O primeiro Capítulo da dissertação, que apresenta o modelo SILEX tal como ele se encontrava definido em 1999, não foi publicado. Daí a referência à versão original da tese.

- FINGER, M. 1998: Tagging a morphologically rich language; in: P. Sojka, V. Matousek, K. Pala, and I. Kopecek (ed) **Proceedings** of the 1st Workshop on Text, Speech and Dialogue TDS'98. Brno: Masaryk University, 39-44.
- FINGER, M. ; BRITTO, H. ; GALVES, C. . Computational and linguistic aspects of the construction of the tycho brahe parsed corpus of historical portuguese. In: 1st Freiburg Workshop on Romance Corpus Linguistics, 2000, Freiburg, Germany. **Proceedings** of the 1st Freiburg Workshop on Romance Corpus Linguistics, 2000.
- GALVES,C.;BRITTO,H.**A construção do corpus anotado do português histórico tycho brahe**. In: IV Encontro para o Processamento Computacional da Língua Portuguesa Escrita e Falada (PROPOR99). Évora, Portugal: Universidade de Évora, 1999. p.81–92.
- GIUSTI, R., Candido Jr, A. MUNIZ, M., CUCATTO, L. A. ALUÍSIO, S. Automatic detection of spelling variation in historical corpus: An application to build a Brazilian Portuguese spelling variants dictionary. In: **Proceedings** of the Corpus Linguistics 2007 (to appear).
- GOMES, L. ; MACHADO FILHO, Américo V. L. . Considerações sobre participios e nomes deverbais no Foro Real. In: 58ª Reunião Anual da SBPC, 2006, Florianópolis. **Anais e Resumos**. Florianópolis : SBPC, 2006.
- HIROHASHI, A. S. **Aprendizado de regras de substituição para normatização de textos históricos**. Dissertação de Mestrado – Instituto de Matemática e Estatística, USP, São Paulo, Brazil, 2004.
- HOUAISS, A. Dicionário da Língua Portuguesa. 1ª ed. Editora Objetiva, Rio de Janeiro, 2001.
- HOPCROFT, MOTWANI & ULLMAN. **Introduction to Automata Theory, Languages, and Computation**. 2nd edition, Addison Wesley, 2001.
- INSTITUTO DE MATEMÁTICA E ESTATÍSTICA - USP (São Paulo). **Corpus Histórico do Português Tycho Brahe**. Disponível em: <<http://www.ime.usp.br/~tycho/corpus/>>. Acesso em: 29 set. 2006.
- JANSSEN, Maarten; José Pedro FERREIRA. **Nomes deverbais eventivos: contributos para o seu estudo**. Comunicação a apresentar ao XXIII Encontro Nacional da Associação Portuguesa de Linguística, Évora, Outubro de 2007 (em preparação).
- JARGAS, A. M.. Expressões Regulares: uma abordagem divertida. 1 ed. São Paulo: Editora Novatec, 2006.
- KEHDI, Valter. **Formação de palavras em português**. 2ª ed. São Paulo: Editora Ática, 1997.
- KEHDI, V. **Morfemas do Português**. 4ª ed. São Paulo: Editora Ática, 1998.
- KEHDI,V. **A derivação regressiva em português**. In Filologia e Lingüística Portuguesa, nº 2, 1998, p. 205-213.
- LIMA, Rocha. **Gramática Normativa da Língua Portuguesa**. 1ª ed. Rio de Janeiro: vis mea in labore, 1957.
- LYONS, J. **Introdução à lingüística teórica**. Trad. de Rosa V. Mattos e Silva & Hélio Pimentel. São Paulo: Nacional, 1979.
- MACHADO FILHO, Américo Venâncio Lopes. **A pontuação em manuscritos medievais portugueses**. 2000. 215 f. Dissertação (Mestrado) - Departamento de Instituto de Letras, UFBA, Salvador, 2000.
- MACHADO, José Pedro. **Dicionário Etmológico da Língua Portuguesa**. 2ª ed. volume II e III. Editora livros horizonte ltd, Lisboa e São Paulo, 1967.
- MARONEZE, B. O. . **Um modelo teórico de formação de palavras e sua aplicação aos deverbais do português**. In: VIII Enapol, 2007, São Paulo. Interface da Ciência Lingüística com as demais áreas do conhecimento: domínios e fronteiras. São Paulo : Paulistana, 2005.
- MATEUS, Maria Helena Mira. **Gramática da Língua Portuguesa**. 1983.

- MEILLET, A. **L'évolution des formes gramaticales**. In: A. MEILLET. Linguistique historique et linguistique générale. Paris: Champion, [1912] 1948.
- MENEGATTI, T. A. **Regras Lingüísticas para Tratamento Computacional da Variação de Grafia e Abreviaturas do Corpus Tycho Brahe**. Campinas: Universidade de Campinas, 2002. Relatório técnico.
- MUNIZ, M. C. M. **A construção de recursos lingüístico-computacionais para o português do Brasil: o projeto de Unitex-PB**. Dissertação de Mestrado. Instituto de Ciências Matemáticas de São Carlos, USP. 72 p. 2004.
- NASCENTES, Antenor. **Dicionário Etmológico Resumido**. Instituto nacional do livro, 1966.
- NUNES, José Joaquim. **Compêndio de gramática histórica portuguesa (Fonética e Morfologia)**. 9ª edição. Lisboa: Clássica Editora, 1989.
- OLIVEIRA, A. S. . O estudo de deverbais no Vocabulário da Carta de Pero Vaz de Caminha. In: IV Congresso Internacional da ABRALIN, 2005, Brasília. **Anais**, 2005.
- PAIXÃO, M. C S.. and T. TRIPPEL. Building a historical corpus for Classical Portuguese: some technological aspects, In P. Baroni et al. (eds.) **Proceedings** of V International Conference on Language Resources and Evaluation (LREC 2006), pp. 1831-1836. Genova: ELRA, 2006.
- PAUMIER, Sébastien. **Unitex: Manual de utilização**. Si: Université Marne-la-valée, 2002. Disponível em: <<http://ladl.univ-mlv.fr/brasil/Ferramentas/ManualUnitex-0-1-2-3-4.zip>>. Acesso em: 24 abr. 2006.
- PAUMIER, S. **Manuel d'utilisation du logiciel UNITEX**. IGM, Université Marne-la-Vallée. Available on-line <http://www-igm.univ-mlv.fr/~unitex/ManuelUnitex.pdf>, 2006 (accessed 22 June 2007).
- PIEL, Joseph. **A formação dos substantivos abstractos em português**. In Biblos, vol. XVI, 1940, p. 209-237.
- PINHEIRO, Gisele Montilha e Sandra Maria ALUÍSIO. "Córpus Nilc: descrição e análise crítica com vistas ao projeto Lacio-Web." NILC-TR-03-03, fevereiro 2003.
- PINTO, R. M. História da Língua Portuguesa IV. Século XVIII. São Paulo: Editora Ática, 1988.
- RANCHHOD, E. M. **O uso de dicionários e de autômatos finitos na representação lexical das línguas naturais**, In E. Ranchhod (ed.) Tratamento das Línguas por Computador: uma Introdução à Linguística Computacional e suas Aplicações, Lisboa: Caminho, 2001, p. 13-47.
- REIS, N. D.. **Deverbais em um texto português do século XIV: Considerações sobre o étimo**, 2005. Disponível em: www.prohpor.ufba.br/deverbetimo.pdf>. Acesso em: 8 ago. 2008.
- RIO-TORTO, Graça Maria. Do ser à acção: “O facto de ser X”, “atitude de (quem é) x” e “condição (estatuto) de x”. Separata da *Revista da Universidade de Coimbra*, vol. de homenagem à Luis de Albuquerque, pp. 427-456.
- RIO-TORTO, Graça Maria. **Sincronia, diacronia e análise genolexical**. In *Diacrítica*, nº 11, pp. 227-244, 1996.
- RIO-TORTO, Graça Maria (org.). **Verbos e nomes em português**. Coimbra, Almedina, 2004.
- RODRIGUES, Alexandra Filipa Soares. **Formação de substantivos deverbais sufixados em português**. Dissertação de Doutoramento em Linguística Portuguesa. Coimbra, Faculdade de Letras da Universidade de Coimbra (inérita).
- RYDBERG-COX, J. A. **Automatic disambiguation of Latin abbreviations in early modern texts for humanities digital libraries**. In: Joint Conference on Digital Libraries. Houston, USA: IEEE Press, 2003. v. 3, p. 372-373.
- SAID ALI, Manuel. **Gramática Histórica da Língua Portuguesa**. 3ª edição melhorada e

- aumentada de Lexicologia e Formação de Palavras e Sintaxe do Português Histórico [Estabelecimento do texto, revisão, notas e índices pelo Prof. Maximiano de Carvalho e Silva]. São Paulo, Edições Melhoramentos, 1964.
- SANTOS, Carla Maria Bastos dos. **Os sufixos -ção e -mento na construção de nomes de ação e de processo: contribuições à prática lexicográfica**. 2006. Dissertação (Mestrado em Letras) - Universidade Federal do Rio Grande do Sul.
- SANDBMANN, Antônio José. **Formação de palavras no Português Brasileiro Contemporâneo**. Curitiba: Scientia et Labor / Ícone, 1989.
- SANDBMANN, Antônio José. **Morfologia Geral**. São Paulo: Editora Contexto, 1991.
- SANDBMANN, Antônio José. **Competência Lexical: Produtividade, Restrições e Bloqueio**. Curitiba: Editora da UFPR, 1991.
- SANDBMANN, Antônio José. **Morfologia lexical: Formação de palavras, ampliação do léxico, produtividade lexical**. São Paulo: Editora Contexto, 1992.
- SARDINHA, Tony Berber. **Linguística de Corpus**. Barueri, SP: Manole, 2004.
- SARDINHA, Tony Berber. **A língua portuguesa no computador**. Campinas: Mercado de Letras, 2005. 296 p.
- SILBERZTEIN, Max. **Text Indexing with INTEX**. In Computer and the Humanities. Amsterdam: Academic Publishers, v.33, 1999.
- SILVA, Antonio de Moraes. **Dicionário da Língua Portuguesa**. 2ª ed. Lisboa typ. Lacerdina, 1813.
- TOLEDO, M. A. R. A.. **Representação Sintático-semântica do deverbais em Português**. 1976. Tese de doutoramento (Doutorado em Letras) - Universidade de São Paulo.
- UNIVERSIDADE FEDERAL DO RIO DE JANEIRO. PHPB - **Para uma História do Português do Brasil**. Disponível em: <<http://www.letas.ufrj.br/phpb-rj/>>. Acesso em: 29 set. 2006.
- VALE, O. A., et al., **Building a large dictionary of abbreviations for named entity recognition in Portuguese historical corpora**. In: LATECH 2008. Paris: ELRA, v. 1. p. 1-10, 2008.

Apêndice A: regras no formato de expressões regulares para deverbais de ação com variação de grafia

<p>Estar</p> <p>est-</p>	<p>Por sufixação:</p> <p>Radical 1: estt? + vt1 + sufixo</p> <hr/> <p>Por conversão:</p> <p>Radical1: estt? + dg ou 0</p> <hr/> <p>Por desflexionação:</p> <p>Infinitivo – estt? + conj1</p> <p>Particípio1 – estt?[ao]d[ao] ou (estado esttada estado estada)</p>
<p>Haver</p> <p>hav-</p>	<p>Por sufixação:</p> <p>Radical 1: ha[vu] + vt2 + sufixo</p> <hr/> <p>Por conversão:</p> <p>Radical1: ha[vu] + dg ou 0</p> <hr/> <p>Por desflexionação:</p> <p>Infinitivo – ha[vu] + conj2</p> <p>Particípio1 – ha[vu][ieyj]d[ao] ou (havido havida havydo havedo havjdo)</p>
<p>Poder</p> <p>part-</p>	<p>Por sufixação:</p> <p>Radical 1: podd? + vt2 + sufixo</p> <hr/> <p>Por conversão:</p> <p>Radical1: podd? + dg</p>

	<p>Por desflexionação:</p> <p>Infinitivo – podd? + conj2</p> <p>Particípio1 – podd?[ieyj]d[ao]</p>
--	--

Passar pas-	<p>Por sufixação:</p> <p>Radical 1: (pass pas paç) + vt2 + sufixo</p>
	<p>Por conversão:</p> <p>Radical1: (pass pas paç) + dg</p>
	<p>Por desflexionação:</p> <p>Infinitivo – (pass pas paç) + conj3</p> <p>Particípio1 – (pass pas paç)[a]d[ao] ou (passar pasar paçar passado pasado paçado passada pasada paçada)</p>

Achar Ach-	<p>Por sufixação:</p> <p>Radical 1: (ach ax) + vt1 + sufixo</p>
	<p>Por conversão:</p> <p>Radical1: (ach ax) + dg</p>
	<p>Por desflexionação:</p> <p>Infinitivo – (ach ax) + conj1</p> <p>Particípio1 – (ach ax)[a]d[ao] ou (achar axar achado achada axado axada)</p>

Ficar	Por sufixação:
fic-	Radical 1: f[iyj]c + vt1 + sufixo
	Por conversão:
	Radical1: f[iyj] + dg
	Por desflexionação:
	Infinitivo – f[iyj] + conj1
	Particípio1 – f[iyj]c + [a]d[ao] ou (ficar ficada ficado)

Chegar	Por sufixação:
cheg-	Radical 1: (cheg xeg) + vt + sufixo
	Por conversão:
	Radical1: (cheg xeg) + dg
	Por desflexionação:
	Infinitivo – (cheg xeg) + conj1
	Particípio1 – (cheg xeg) [a]d[ao]

Querer	Por sufixação:
quer-	Radical 1: (quer) + vt + sufixo
	Por conversão:
	Radical1: (quer) + dg
	Por desflexionação:
	Infinitivo – (quer) + conj2
	Particípio1 – quer[i]d[ao]

Saber	Por sufixação:
sab-	Radical 1: (sabb sab) + vt2 + sufixo
	Por conversão:
	Radical1: (sabb sab) + dg
	Por desflexionação:
	Infinitivo – (sabb sab) + conj1
	Particípio1 – (sabb sab)[a]d[ao]

Tomar	Por sufixação:
tom-	Radical 1: (ttom tomm tom) + vt + sufixo
	Por conversão:
	Radical1: (tom tomm ttom) + dg
	Por desflexionação:
	Infinitivo – (tom tomm ttom) + conj1
	Particípio1 – (tom tomm ttom)[a]d[ao]

Parecer	Por sufixação:
parec-	Radical 1: pare[çsc] + vt2 + sufixo
	Por conversão:
	Radical1: pare[çsc] + dg
	Por desflexionação:
	Infinitivo – pare[çsc] + conj1
	Particípio1 – pare[çsc] [iyje]d[ao] ou (parecer pareser parecer paresido parecido pareçido)

Matar	Por sufixação:
mat- mort-	Radical 1: matt? + vt1 + sufixo
	Radical 2: mortt? + vt1 + sufixo
	Por conversão:
	Radical1: matt?? + dg
	Por desflexionação:
	Infinitivo – matt? + conj1
	Particípio1 – matt?[a]d[ao]
	Particípio2 – mortt?[oa]

Faltar	Por sufixação:
falt-	Radical 1: fa[lu]tt? + vt1 + sufixo
	Por conversão:
	Radical1: fa[lu] + dg
	Por desflexionação:
	Infinitivo – fa[lu] + conj1
	Particípio1 – fa[lu]ad[ao]

Deixar	Por sufixação:
deix-	Radical 1: de[iyj]x + vt1 + sufixo
	Por conversão:
	Radical1: de[iyj]x + dg

	<p>Por desflexionação:</p> <p>Infinitivo – de[iyj]x + conj1</p> <p>Particípio1 – de[iyj]x [a]d[ao]</p>
--	--

Servir	<p>Por sufixação:</p> <p>Radical 1: [sc]er[vu] + vt2 + sufixo</p>
serv-	<p>Por conversão:</p> <p>Radical1: [sc]er[vu] + dg</p>
	<p>Por desflexionação:</p> <p>Infinitivo – [sc]er[vu] + conj1</p> <p>Particípio1 – [sc]er[vu][iyje]d[ao]</p>

Tirar	<p>Por sufixação:</p> <p>Radical 1: t[iyj]r + vt1 + sufixo</p>
tir-	<p>Por conversão:</p> <p>Radical1: t[iyj] + dg</p>
	<p>Por desflexionação:</p> <p>Infinitivo – t[iyj] + conj1</p> <p>Particípio1 – t[iyj]rad[ao]</p>

Levar	<p>Por sufixação:</p> <p>Radical 1: ll?e[vu] + vt1 + sufixo</p>
lev-	<p>Por conversão:</p> <p>Radical1: ll?e[vu] + dg</p>

	<p>Por desflexionação:</p> <p>Infinitivo – ll?e[vu] + conj1</p> <p>Particípio1 – ll?e[vu][a]d[ao]</p>
--	---

Pedir ped-	<p>Por sufixação:</p> <p>Radical 1: p[eij]dd? + vt3 + sufixo</p>
	<p>Por conversão:</p> <p>Radical1: p[eij]dd? + dg</p>
	<p>Por desflexionação:</p> <p>Infinitivo – p[eij]dd? + conj1</p> <p>Particípio1 – p[eij]dd?[ij]d[ao]</p>

Pagar peg-	<p>Por sufixação:</p> <p>Radical 1: peg + vt1 + sufixo</p>
	<p>Por conversão:</p> <p>Radical1: peg + dg</p>
	<p>Por desflexionação:</p> <p>Infinitivo – peg + conj1</p> <p>Particípio1 – peg[a]d[ao]</p>

Seguir segu- sig-	<p>Por sufixação:</p> <p>Radical 1: [sc][eij]gu + vt2 + sufixo</p>
	<p>Por conversão:</p> <p>Radical = sig-</p> <p>Radical1: [sc][ij]gu + dg</p>

	<p>Por desflexionação:</p> <p>Infinitivo – [sc][eij]gu + conj1</p> <p>Particípio1 – [sc][eij]gu[ij]d[ao]</p>
--	--

Escrever	<p>Por sufixação:</p> <p>Radical 1: escre[vu] + vt2 + sufixo</p> <p>Radical 2: escr[ij]tt? + vt1 + sufixo</p>
escrev-	
escrit-	<p>Por conversão:</p> <p>Radical1: escre[vu] + dg</p>
	<p>Por desflexionação:</p> <p>Infinitivo – escre[vu] + conj2</p> <p>Particípio1 – escr[vu][ij]d[ao]</p> <p>Participio2 – escr[ij]tt?[ao]</p>

Viver	<p>Por sufixação:</p> <p>Radical 1: [vu][ij][vu] + vt2 + sufixo</p>
viv-	
	<p>Por conversão:</p> <p>Radical1: [vu][ij][vu] + dg</p>
	<p>Por desflexionação:</p> <p>Infinitivo – [vu][ij][vu] + conj2</p> <p>Particípio1 – [vu][ij][vu] [ij]d[ao]</p>

Conhecer	<p>Por sufixação:</p> <p>Radical 1: co[nñm]he[cs] + vt2 + sufixo*</p>
conhec-	

	<p>Por conversão:</p> <p>Radical1: co[nãm]he[csç] + dg</p>
	<p>Por desflexionação:</p> <p>Infinitivo – : co[nãm]he[csç] + conj2</p> <p>Particípio1 – : co[nãm]he[csç][iyj]d[ao]</p>

Mostrar	<p>Por sufixação:</p> <p>Radical 1: mostt?r + vt1 + sufixo</p>
mostr-	<p>Por conversão:</p> <p>Radical1: mostt?r + dg</p>
	<p>Por desflexionação:</p> <p>Infinitivo – mostt?r + conj1</p> <p>Particípio1 – mostt?r [a]d[ao]</p>

Trabalhar	<p>Por sufixação:</p> <p>Radical 1: trabalh + vt1 + sufixo</p>
trabalh-	<p>Por conversão:</p> <p>Radical1: trabalh + dg</p>
	<p>Por desflexionação:</p> <p>Infinitivo – trabalh + conj1</p> <p>Particípio1 – trabalh [a]d[ao]</p>

perguntar	<p>Por sufixação:</p> <p>Radical 1: pergu[nm]t + vt1 + sufixo</p>
pergunt-	

	<p>Por conversão:</p> <p>Radical1: pergu[nm]t + dg</p>
	<p>Por desflexionação:</p> <p>Infinitivo – pergu[nm]t + conj1</p> <p>Particípio1 – pergu[nm]t [a]d[ao]</p>

referir	<p>Por sufixação:</p> <p>Radical 1: reff?er + vt2 + sufixo</p>
refer-	<p>Por conversão:</p> <p>Radical1: reff?er + dg</p>
	<p>Por desflexionação:</p> <p>Infinitivo – reff?er + conj3</p> <p>Particípio1 – reff?er [iyj]d[ao]</p>

correr	<p>Por sufixação:</p> <p>Radical 1: corr + vt2 + sufixo</p>
corr-	<p>Por conversão:</p> <p>Radical1: corr + dg</p>
	<p>Por desflexionação:</p> <p>Infinitivo – corr + conj2</p> <p>Particípio1 – corr [iyj]d[ao]</p>

Obrigar	<p>Por sufixação:</p> <p>Radical 1: obr[iyj]gg? + vt1 + sufixo</p>
obrig-	

	<p>Por conversão:</p> <p>Radical1: obr[iyj]gg? + dg</p>
	<p>Por desflexionação:</p> <p>Infinitivo – obr[iyj]gg? + conj1</p> <p>Particípio1 – obr[iyj]gg?[a]d[ao]</p>

perder	<p>Por sufixação:</p> <p>Radical 1: perdd? + vt2 + sufixo</p>
perd-	<p>Por conversão:</p> <p>Radical1: perdd? + dg</p>
	<p>Por desflexionação:</p> <p>Infinitivo – perdd? + conj2</p> <p>Particípio1 – per [iyj]d[ao]</p>

guardar	<p>Por sufixação:</p> <p>Radical 1: guard? + vt1 + sufixo</p>
guard-	<p>Por conversão:</p> <p>Radical1: guardd? + dg</p>
	<p>Por desflexionação:</p> <p>Infinitivo – guardd? + conj1</p> <p>Particípio1 – guardd?[a]d[ao]</p>

esperar	<p>Por sufixação:</p> <p>Radical 1: esp?er + vt1 + sufixo</p>
esper-	

	<p>Por conversão:</p> <p>Radical1: espp?er + dg</p>
	<p>Por desflexionação:</p> <p>Infinitivo – espp?er + conj1</p> <p>Particípio1 – espp?er [a]d[ao]</p>

constar	<p>Por sufixação:</p> <p>Radical 1: co[nm]stt? + vt2 + sufixo</p>
const-	<p>Por conversão:</p> <p>Radical1: co[nm]stt ? + dg</p>
	<p>Por desflexionação:</p> <p>Infinitivo – co[mn]stt? + conj1</p> <p>Particípio1 – co[mn]stt? [a]d[ao]</p>

Vir	<p>Por sufixação:</p> <p>Radical 1: v + vt2 + sufixo</p>
vir-	<p>Por conversão:</p> <p>Radical1: v + dg ????</p>
	<p>Por desflexionação:</p> <p>Infinitivo – v + conj3</p> <p>Particípio1 – v[iyj]nd[ao] ou</p>

Dizer	Por sufixação:
diz- / dit-	Radical 1: di[sz] + vt2 + sufixo
	Radical 2: d[ijy]tt? + vt1 + sufixo
	Por conversão:
	Radical1: di[sz]+ dg
	Por desflexionação:
	Infinitivo – d[ieyj][sz]er
	Particípio1 – d[ieyj]tt?[ao]
	Dizido?
	ou
	(dizer diser dyzer djzer dezer dyser djser deser ditto dito)

Script - Estudo Piloto

```
#!/usr/bin/python
# -*- coding: iso-8859-1 -*-

import os
#import locale
import re

if __name__ == "__main__":
    show_line_numbers = False

    corpus_file = open('corpus_100.iso-8859-1.txt', 'r')
    corpus = corpus_file.readlines()
    corpus_file.close()

    # Sufixação
    sufixacao = []
    sufixacao.append({'base':
'(cheg|xeg|lav|llav|estt|est|cham|xam|chamm|paç|pass|pas|trabalh|pag|mandd|mamd|mand|ditt|di
t|dytt|dyt|djtt|dj|feit|feitt|feyt|feytt|fejt|fejtt)',
        'vt': '[a]',
        'suffix': '(çã|sã|çaõ|cão|mento|ência|ancia|dela|ura)'})
    sufixacao.append({'base':
'(escriv|screu|escriv|escriu|scryv|escryu|escrjv|escrju|dev|deu|podd|pod|diz|dis|comm|com|faz|f
as|parec|pares)',
        'vt': '[eiyj]',
        'suffix': '(çã|sã|çaõ|cão|mento|ência|ancia|dela|ura)'})
    sufixacao.append({'base': '(partt|partv)',
        'vt': '[eiyj]',
        'suffix': '(çã|sã|çaõ|cão|mento|ência|ancia|dela|ura)'})

    # Conversão
    conversao = []
    conversao.append({'base':
'(cheg|xeg|dev|deu|lav|llav|estt|est|cham|xam|chamm|paç|pass|pas|trabalh|pag|mandd|mamd|ma
nd|parec|pareç)',
        'dg': '[aeo]'})

    # Desflexionação
    desflexionacao = []
    # Infinitivo
    desflexionacao.append({'base':
'(cheg|xeg|lav|llav|estt|est|cham|xam|chamm|paç|pass|pas|trabalh|pag|mandd|mamd|mand)',
        'conj': 'ar'})
```

```

    desflexionacao.append({'base':
'(escrev|screu|escriv|escriu|scriv|scriyu|escrijv|escriju|dev|deu|podd|pod|diz|dis|comm|com|faz|f
as|parec|pares|pareç)',
        'conj': 'er'})
    )
    desflexionacao.append({'base': '(partt|part|v|abr|p)',
        'conj': '[iyjo]r'})
    )
    # Participio
    desflexionacao.append({'base':
'(escrev|screu|escriv|escriu|scriv|scriyu|escrijv|escriju|dev|deu|podd|pod|lav|llav|estt|est|cham|x
am|chamm|paç|pass|pas|trabalh|pag|mandd|mamd|mand|partt|part)',
        'conj': '[iyja]d[ao]'})
    )
    # Participio Irregular
    desflexionacao.append({'base': '(ditt|dit|dytt|dyt|djtt|djt|feit|feitt|feyt|feytt|fejt|fejtt)',
        'conj': '[ao]'})
    )

line_no = 1
for line in corpus:
    for deverbale in sufixacao:
        compiled_re = re.compile(r'\b(' + deverbale['base'] + deverbale['vt'] + deverbale['sufix'] +
r'\b)')
        matchs = compiled_re.findall(line)
        if len(matchs) != 0:
            if show_line_numbers:
                print str(line_no) + ': ',
                print ''.join([x[0] for x in matchs])

    for deverbale in conversao:
        compiled_re = re.compile(r'\b(' + deverbale['base'] + deverbale['dg'] + r'\b)')
        matchs = compiled_re.findall(line)
        if len(matchs) != 0:
            if show_line_numbers:
                print str(line_no) + ': ',
                print ''.join([x[0] for x in matchs])

    for deverbale in desflexionacao:
        compiled_re = re.compile(r'\b(' + deverbale['base'] + deverbale['conj'] + r'\b)')
        matchs = compiled_re.findall(line)
        if len(matchs) != 0:
            if show_line_numbers:
                print str(line_no) + ': ',
                print ''.join([x[0] for x in matchs])

    line_no = line_no + 1

```

Apêndice B: lista de deverbais de ação sufixados em - mento e -ção

Sufixo -mento

destacamento	casamento	adiantamento	acompanhamento
medicamento	lançamento	temperamento	salvamento
melhoramento	livramento	abarracamento	aproveitamento
comportamento	acampamento	regulamento	arranchamento
acatamento	arrendamento	aviamento	arrojamento
aforamento	mandamento	armamento	ornamento
campamento	apartamento	encerramento	firmamento
arrançamento	predicamento	entabulamento	ajustamento
entrincheiramento	enfadamento	apontamento	parlamento
madeiramento	enterramento	arruamento	amancebamento
Alinhamento	resfriamento	requerimento	fardamento
Andamento	acrescentamento	retardamento	encantamento
descontentamento	municiamto	arrebatamento	acomodamento
Quentamento	orçamento	aquartelamento	mantimento
emadearamento	paramento	juntamento	emprazamento
Avançamento	encaixamento	descaroçamento	derramamento
Arbitramento	alevntamento	seguimento	prolongamento
Procedimento	movimento	incitamento	guisamento
Encadeamento	costeamento	conhecimento	chamamento
Agastamento	achamento	acabamento	vimento
Rompimento	rodamento	quebrantamento	quebramento
Passamento	louvamento	granjeamento	filhamento
Filamento	estabelecimento	esfalfamento	engradamento
Douramento	descobrimento	descaramento	desavergonhamento
Cortamento	comandamento	cerceamento	açodamento
Alistamento	aldeamento	sustentamento	serramento
Rodeamento	retelhamento	requerimento	represamento
Rendimento	relacionamento	regimento	recrutamento
Recreamento	povoamento	pastoreamento	nivelamento
Nascimento	maltratamento	ligamento	laçamento
Laboramento	julgamento	fretamento	fechamento
Falecimento	fabricamento	estagnamento	espancamento
Esfriamento	esfregamento	escoamento	escalvamento
Envigamento	entrevamento	entaipamento	enojamento
encorajamento	encanamento	dicamento	desovamento
desmoronamento	desmembramento	desfilamento	desenfreamento
desencantamento	embaixamento	consentimento	congraçamento
comprimimento	completamento	cerramento	brotamento
Atrevimento	atrasamento	assentamento	arrombamento
Arripiamento	arredamento	arrecadamento	arrancamento
Arejamento	aposentamento	alçamento	alvidramento
Alteamento	alargamento	agrupamento	agravamento
Afirmamento	acolhimento	acantonamento	acanhamento
Abrasamento	custeamento	descobrimento	Procedimento
Conhecimento	resfriamento	seguimento	Requeirimento
movimento			

Sufixo –ção

povoação	navegação	obrigação	coração
administração	relação	salvação	estimação
declaração	conservação	informação	demarcação
consideração	embarcação	admiração	oração
fundação	doação	situação	conjuração
consolação	comparação	duração	continuação
geração	imitação	capitação	criação
determinação	arrecadação	conversaço	representação
restauração	quietação	fortificação	confirmação
invocação	operação	reputação	sustentação
sublevação	arrematação	veneração	habitação
demonstração	plantação	edificação	alteração
nomeação	inclinação	comunicação	observação
mineração	averiguação	participação	estação
exportação	procuração	justificação	careação
imaginação	formação	ponderação	narração
avaliação	contemplação	moderação	separação
confederação	quitação	pregação	congregação
deliberação	publicação	consternação	graduação
multiplicação	elevação	perturbação	ocupação
educação	desesperação	propagação	negociação
mortificação	explicação	aplicação	tentação
indagação	carregação	emulação	asseveração
delatação	aceitação	conspiração	condenação
ordenação	acareação	preparação	denominação
aprovação	individação	desconsolação	remuneração
recomendação	vexação	significação	respiração
fermentação	viração	obstinação	meditação
inundação	limitação	consignação	configuração
combinação	apelação	privação	ostentação
equivocação	dissimulação	inquietação	vegetação
subordinação	reformação	circulação	especulação
revelação	ratificação	peregrinação	declinação
agitação	adoração	reedificação	denunciação
modificação	interpretação	exploração	alagação
recreação	perpetração	indignação	confiscação
acomodação	inchação	imputação	aclamação
tripulação	transmigração	predestinação	insinuação
importação	confrontação	vereação	penetração
lotação	civilização	variação	notificação
expectação	exaltação	evaporação	esquipação
enformação	dilatação	celebração	capinação
relaxação	protestação	precipitação	citação
tribulação	resignação	gratificação	filiação
exageração	evacuação	arrumação	armação
apresentação	transpiração	renovação	reconciliação
negação	dominação	computação	comiseração
renúnciação	recuperação	recopilação	purificação
promulgação	murmuração	laboração	despovoação
deixação	aliciação	abonação	usurpação
retratação	reparação	realização	provação
preocupação	suplicação	regulação	recordação
purgação	prestação	inflamação	exortação

examinação	dissertação	depravação	contestação
comutação	cominação	calcinação	aluação
abominação	simulação	manifestação	fabricação
emigração	cooperação	convocação	cogitação
atestação	arribação	acusação	visitação
trasladação	transformação	saudação	pronúnciação
permutação	numeração	mediação	maturação
legislação	indenização	importunação	ervação
dissipação	desolação	dedicação	consagração
cerração	assolação	antecipação	adulação
sonegação	regeneração	organização	meação
marcação	investigação	inspiração	impugnação
habilitação	fecundação	extirpação	estagnação
especificação	enumeração	dispensação	devastação
deslocação	desinquietação	degeneração	decoração
circunvalação	apuração	aproximação	aceleração
violação	vadiação	trituração	transplantação
revogação	preservação	pacificação	manipulação
insubordinação	inovação	indicação	hesitação
expugnação	exclamação	esfregação	encarnação
detestação	custeação	constipação	congelação
complicação	compensação	colação	capação
articulação	amarração	abjuração	varação
transmutação	transação	sazonação	santificação
proclamação	piração	morigeração	mastigação
intimação	implicação	explanação	estipulação
escavação	distilação	deterioração	designação
derrogação	deprecação	delegação	coroação
copelação	consumação	conglutinação	conciliação
coloração	coagulação	certificação	capitulação
avaloação	associação	animação	ampliação
altercação	admoestação	vibração	versificação
turbação	trespassação	terminação	revalidação
refinação	recitação	prevaricação	peroração
ocultação	mareação	maduração	ligação
juração	interpolação	infestação	humilhação
germinação	fiscalização	ficação	expulsão
expedição	exalação	espectação	equipação
emborcação	destinação	destilação	desobrigação
desmembração	desertação	desembarcação	derivação
deputação	degradação	cultivação	costeação
constelação	comprovação	comemoração	colocação
coação	classificação	cavação	canonização
anotação	amoestação	ameação	amalgamação
afetação	vulgarização	voação	verificação
ventilação	vaporação	vadeação	turvação
tributação	triangulação	travação	transitação
traição	testação	taxação	solação
socavação	sideração	serração	secularização
satisdação	sagração	safação	rotação
roração	revocação	retificação	retardação
resfriação	reservação	reiteração	refutação
recusação	reclamação	reciprocção	rebelação
ralação	quartação	prostração	prorrogação
prolongação	profanação	postulação	perpetuação

pastoreação	paração	palpitação	paleação
paginação	pactuação	ovação	opilação
ondeação	niveação	mudação	mitigação
mineralização	mineiração	metrificação	meneação
memoração	maneação	maceração	logração
liberação	liação	legação	laxação
lambicação	labutação	isolação	irritação
irradiação	interrogação	intermediação	instigação
inspissação	inoculação	inflação	individualização
incorporação	inauguração	ilustração	identificação
hebetação	governação	glorificação	frutificação
fornicação	formalização	filtração	falsificação
facetação	exterminação	extenuação	expição
expatriação	evocação	evitação	estração
estilação	espiculação	espevitación	espetação
espedição	escorificação	escavacação	entaipação
encorporação	encampação	encadernação	encadeação
embriagação	emancipação	emanação	eliminação
elaboração	douração	desorganização	desoneração
desfiguração	desentoação	depuração	denotação
denegação	demostração	delineação	deleitação
defraudação	decapitação	debilitação	danificação
danação	curação	cristalização	contratação
contiguação	contemporização	confutação	conformação
condensação	condecoração	concentração	compilação
clarificação	cintilação	cementação	cavilação
catequização	carreação	cariação	caiação
boleação	beatificação	avexação	avaliação
autuação	autorização	atuação	atracação
aspiração	arpoação	apreciação	aperreação
anunciação	anulação	aniquilação	amotinação
amofinação	alvoração	alucinação	alouração
alimentação	alienação	alegação	agregação
afundação	afogueação	aduação	abnegação
abitação			