



UNIVERSIDADE FEDERAL DO RIO DE JANEIRO  
CENTRO DE LETRAS E ARTES  
PÓS-GRADUAÇÃO EM LINGUÍSTICA  
CURSO DE DOUTORADO

Vânia Lisbôa da Silveira Guedes

Nominalizações deverbais em artigos científicos: uma contribuição para a análise e a indexação temática da informação.

Rio de Janeiro  
2010

Vânia Lisbôa da Silveira Guedes

Nominalizações deverbais em artigos científicos: uma contribuição para a análise e a indexação temática da informação.

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Linguística da Universidade Federal do Rio de Janeiro (UFRJ), como quesito necessário à obtenção do título de Doutor em Linguística.

Orientadora: Professora Doutora Maria Cecília de Magalhães Mollica e

Co-orientadora: Professora Doutora Marisa Leal.

Rio de Janeiro  
2010

## FICHA CATALOGRÁFICA

G 924 Guedes, Vânia Lisbôa da Silveira.

Nominalizações deverbais em artigos científicos: uma contribuição para a análise e a indexação temática da informação / Vânia Lisbôa da Silveira Guedes. -- 2010.

109 f.:il.

Tese (Doutorado em Linguística) - Programa de Pós-Graduação em Linguística, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2010.

Orientadores: Cecília de Magalhães Mollica e Marisa Leal.

1. Nominalização Deverbal. 2. Língua Portuguesa.. 3. Indexação da Informação 4. Ponto de Transição de Goffman. 5. Linguística-Teses. I. Mollica, Maria Cecília de Magalhães (Orient.). II. Leal, Marisa (Orient.). III. Universidade Federal do Rio de Janeiro. Programa de Pós-Graduação em Linguística. IV. Título.

CDD: 469.592

Vânia Lisbôa da Silveira Guedes

Nominalizações deverbais em artigos científicos: uma contribuição para a análise e a indexação temática da informação.

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Linguística da Universidade Federal do Rio de Janeiro (UFRJ), como quesito necessário à obtenção do título de Doutor em Linguística.

Orientadora: Professora Doutora Maria Cecília de Magalhães Mollica e

Co-orientadora: Professora Doutora Marisa Leal.

Aprovada em:

---

(Orientadora) Professora Dra. Maria Cecília de Magalhães Mollica – PPGL/UFRJ

---

(Co-orientadora) Professora Dra. Marisa Beatriz Bezerra Leal – IM/UFRJ

---

Professor Dr. Mário Eduardo Toscano Martelotta – PPGL/UFRJ

---

Professora Dra. Vera Lúcia Paredes Pereira da Silva – PPGL/UFRJ

---

Professora Dra. Rosali Fernandez de Souza - PPGCI/IBICT-FACC/UFRJ

---

Professora Dra. Suzana Borschiver – EQ/UFRJ

---

Professor Dr. Alessandro Boechat de Medeiros – PPGL/UFRJ (Suplente)

---

Professora Dra. Maria Cristina Guimarães de Goes Monteiro – LET/PUCRio (Suplente)

A Marcus Vinícius, Rickson e Andriws, a Julia e a meus pais, Walter (*In Memorium*)  
e Rita.

## AGRADECIMENTOS

A Deus, pelos sonhos que pude realizar.

Aos meus queridos pais, Walter da Silveira Guedes (*In Memorium*) e Rita Lisbôa da Silveira Guedes, exemplos de amor e dedicação, por terem propiciado as condições básicas indispensáveis à realização de projetos dessa natureza.

Aos meus queridos filhos, Marcus Vinícius, Rickson e Andriws, pelo carinho, incentivo e compreensão, especialmente nos momentos mais importantes desse percurso de doutoramento.

À Professora Maria Cecília de Magalhães Mollica, minha orientadora, e à Professora Marisa Leal, minha co-orientadora, pelo carinho, maestria e competência presentes em vários momentos do curso de doutorado, sobretudo na orientação da pesquisa de doutoramento, estabelecendo a verdadeira interface da Linguística com a Matemática.

À Professora Margarida Basílio, pelas aulas sobre Teoria Lexical, na PUCRio, permitindo a apropriação de relevantes conhecimentos sobre a teoria e a prática da nominalização e pela orientação na seleção da amostra.

Ao corpo docente e aos funcionários do Programa de Pós-graduação em Linguística da UFRJ, pelos conhecimentos adquiridos e pelas orientações.

À Fátima, Cíntia, Wellington, João e Fernanda, pelas importantes contribuições e parcerias.

Aos professores da Banca de Qualificação e da Banca de Doutorado, pela atenção e orientação.

À Professora Maria José Velloso, pela parceria no Curso de Biblioteconomia e Gestão de Unidades de Informação (CBG) da UFRJ, meu agradecimento eterno. Ao corpo docente, discente e técnico do CBG/UFRJ, pelas conquistas e pelo convívio tão enriquecedor.

Ao Cleber, pelo apoio operacional, desde as minhas primeiras aulas ministradas na Escola de Química até o meu ingresso como professora da FACC/UFRJ, agradeço a indispensável colaboração no processamento dos dados e formatação da tese.

A todos que, direta ou indiretamente, contribuíram para o desenvolvimento deste trabalho, agradeço sinceramente.

“A Ciência é o secular esforço para reunir, por meio do pensamento sistemático, os fenômenos perceptíveis deste mundo numa associação tão completa quanto possível.”

Albert Einstein

## LISTA DE FIGURAS

Figura 1 O artigo de periódico e outros gêneros de processo de pesquisa (SWALES, 1990)	43
Figura 2 Diagrama da inter-relação entre os quatro subcampos. (VANTI, 2002)	54



## LISTA DE FÓRMULAS

Fórmula 1 Primeira Lei de Zipf	55
Fórmula 2 Segunda Lei de Zipf	56
Fórmula 3 Segunda Lei de Zipf	56
Fórmula 4 Aplicação da Segunda lei de Zipf	56
Fórmula 5 Segunda Lei de Zipf	57
Fórmula 6 Segunda Lei de Zipf	57
Fórmula 7 Fórmula do Ponto T	57
Fórmula 8 Fórmula do Ponto T	57

## LISTA DE QUADROS

Quadro 1 Amostra.	49
Quadro 2 Leis e princípios bibliométricos, focos de estudo, principais aplicações e áreas de interesse (GUEDES; BORSHIVER, 2005).	53
Quadro 3 Relação das nominalizações deverbais investigadas no artigo 1.	61
Quadro 4 Representação da Região de Transição de Goffman no artigo 1.	64
Quadro 5 Relação das nominalizações deverbais investigadas no artigo 2.	65
Quadro 6 Representação da Região de Transição de Goffman no artigo 2.	68
Quadro 7 Relação das nominalizações deverbais investigadas no artigo 3.	70
Quadro 8 Representação da Região de Transição de Goffman no artigo 3.	72
Quadro 9 Relação das nominalizações deverbais investigadas no artigo 4.	74
Quadro 10 Representação da Região de Transição de Goffman no artigo 4.	77
Quadro 11 Relação das nominalizações deverbais investigadas no artigo 5.	78
Quadro 12 Representação da Região de Transição de Goffman no artigo 5.	81
Quadro 13 Relação das nominalizações deverbais investigadas no artigo 6.	83
Quadro 14 Representação da Região de Transição de Goffman no artigo 6.	86

## LISTA DE TABELAS

Tabela 1 Síntese da produtividade dos sufixos nominalizadores em análise no artigo 1 e suas frequências relativas de ocorrência.	61
Tabela 2 Síntese da produtividade dos sufixos nominalizadores em análise no artigo 2 e suas frequências relativas de ocorrência.	65
Tabela 3 Síntese da produtividade dos sufixos nominalizadores em análise no artigo 3 e suas frequências relativas de ocorrência.	70
Tabela 4 Síntese da produtividade dos sufixos nominalizadores em análise no artigo 4 e suas frequências relativas de ocorrência.	74
Tabela 5 Síntese da produtividade dos sufixos nominalizadores em análise no artigo 5 e suas frequências relativas de ocorrência.	78
Tabela 6 Síntese da produtividade dos sufixos nominalizadores em análise no artigo 6 e suas frequências relativas de ocorrência.	83
Tabela 7 Síntese da produtividade dos sufixos nominalizadores em análise nos artigos 1, 2, 3 e 4, bem como suas frequências relativas de ocorrência.	91
Tabela 8 Síntese da produtividade dos sufixos nominalizadores em análise nos artigos 5 e 6, bem como suas frequências relativas de ocorrência.	91
Tabela 9 Síntese dos resultados verificados na aplicação da Segunda Lei de Zipf aos artigos sobre a indústria de vinho	93
Tabela 10 Síntese dos resultados verificados na aplicação da Segunda Lei de Zipf aos artigos sobre a Economia	93

## LISTA DE GRÁFICOS

- Gráfico 1 Representação espacial dos resultados obtidos indicando os graus distintos de produtividades das cinco Regras de Formações de Palavras investigadas, nos artigos 1, 2, 3, 4, 5 e 6, com o número total de palavras de cada artigo. 92
- Gráfico 2 Representação espacial dos resultados obtidos indicando a frequência total de ocorrência de cada sufixo nominalizador deverbal investigado nos artigos 1, 2, 3, 4, 5 e 6, com o número total de palavras de cada artigo. 93

## LISTA DE SIGLAS, ABREVIATURAS E SÍMBOLOS

C	Constant
CBG	Curso de Biblioteconomia e Gestão de Unidades de Informação
cv	cultivar vinífera
F	frequency
I <sub>1</sub>	# de palavras que ocorrem uma vez
I <sub>2</sub>	# de palavras que ocorrem duas vezes
I <sub>3</sub>	# de palavras que ocorrem três vezes
I <sub>4</sub>	# de palavras que ocorrem quatro vezes
I <sub>5</sub>	# de palavras que ocorrem cinco vezes
I <sub>n</sub>	# de palavras que ocorrem <i>n</i> vezes
n	frequência
Ph	Potencial Hidrogeniônico
Ponto T	Ponto de Transição de Goffman
R	rank
RAE	Regra de Análise de Estruturas
RFP	Regras de Formação de Palavras
SciELO	Scientific Electronic Library Online
UFRJ	Universidade Federal do Rio de Janeiro
W3	Web Semântica (Web 3.0)
WWW	<i>World Wide Web</i> ou Internet
Web 2.0	conteúdo colaborativo entre usuários na Web
XML	eXtensible Markup Language
#	número

## SUMÁRIO

<b>1 INTRODUÇÃO</b>	17
<b>2 QUESTÕES, OBJETIVOS E HIPÓTESES</b>	19
<b>3 PRESSUPOSTOS TEÓRICOS</b>	25
3.1 BASES DA TEORIA LEXICAL	25
3.2 CONSIDERAÇÕES SOBRE GÊNEROS TEXTUAIS	35
3.3 BREVE HISTÓRICO DO ARTIGO DE PERIÓDICO	40
3.4 CIÊNCIA DA INFORMAÇÃO E ANÁLISE DE DOMÍNIO	44
<b>4 METODOLOGIA</b>	48
4.1 AMOSTRA	48
4.2 A PRODUÇÃO DO VINHO	49
4.3 PROCEDIMENTO DE ANÁLISE	51
4.4 BIBLIOMETRIA: LEIS DE ZIPF E PONTO DE TRANSIÇÃO DE GOFFMAN	52
<b>5 ANÁLISE E DISCUSSÃO DOS RESULTADOS</b>	60
<b>6 SÍNTESE DOS RESULTADOS</b>	90
<b>7 CONCLUSÕES E SUGESTÕES PARA PESQUISAS FUTURAS</b>	95
<b>REFERÊNCIAS</b>	98
<b>GLOSSÁRIO</b>	108

## RESUMO

GUEDES, Vânia Lisbôa da Silveira. *Nominalizações deverbais em artigos científicos: uma contribuição para a análise e a indexação temática da informação*, 2010. 109 f. Tese (Doutorado em Linguística) – Faculdade de Letras, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2010.

Este estudo analisa o discurso científico e tecnológico sobre a indústria de vinhos, em português do Brasil, investigando a recorrência das nominalizações deverbais no léxico do gênero artigo de periódico e suas funções de índices na indexação. Assim, o objetivo é, por um lado, desenvolver uma análise linguística léxico-morfológica do discurso sobre a indústria de vinhos e, por outro lado, contribuir para o desenvolvimento de modelos de sistemas semi-automáticos de indexação. Estabeleceu-se a hipótese de que as nominalizações deverbais formalmente descritas por  $[X] v \rightarrow [ [X] v -\text{ção} ] N$  são predominantes no campo semântico da indústria de vinhos e, portanto, estão presentes no léxico dos artigos de periódicos, nessa área do conhecimento, exercendo funções de índices mais relevantes, devido às suas frequências de ocorrências. Para tal, consideram-se como referencial teórico a análise de gêneros textuais, a teoria lexical e a indexação da informação, fundamentada em leis e princípios bibliométricos. A partir daí, selecionam-se quatro artigos sobre a indústria de vinhos, publicados na *Scientific Electronic Library Online*. Os artigos foram processados pelo *Software Rank Words 2.0*. Para cada artigo, foi produzida uma lista, em ordem decrescente, de frequência de ocorrência das palavras. Em cada lista, é verificada a aplicação das leis de Zipf e Ponto de Transição de Goffman, investigada a recorrência de nominalizações deverbais formuladas de acordo com  $[X] v \rightarrow [ [X] v -\text{ção} ] N$ ,  $[X] v \rightarrow [ [X] v -\text{mento} ] N$ ,  $[X] v \rightarrow [ [X] v -\text{ncia} ] N$ ,  $[X] v \rightarrow [ [X] v -\text{agem} ] N$ ,  $[X] v \rightarrow [ [X] v -\text{da} ] N$  e também verificadas suas frequências relativas e graus distintos de relevância de suas funções de índice, para a indexação da informação. Visando à confrontação dos dados, o método é aplicado a dois artigos sobre Economia. Finalmente, menciona-se que os dados obtidos corroboram a hipótese acima mencionada, revelam maior adequação do método aos artigos sobre Economia e evidenciam a importância das abordagens teóricas e descritivas da nominalização verbal para a indexação da informação científica e tecnológica, no âmbito da Ciência da Informação.

**PALAVRAS-CHAVE:** Análise de gêneros textuais. Leis de Zipf. Ponto de Transição de Goffman. Artigos científicos e tecnológicos. Nominalização verbal. Indexação.

## ABSTRACTS

GUEDES, Vânia Lisbôa da Silveira. *Nominalizações deverbais em artigos científicos: uma contribuição para a análise e a indexação temática da informação*, 2010. 109 f. Tese (Doutorado em Linguística) – Faculdade de Letras, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2010.

This study analyses the scientific and technology discourse about the wine industry in Brazilian Portuguese language, in order to investigate the recurrence of deverbal nominalization present in the lexicon of papers and their index role in the information indexing. The aim of this study is, on the one hand, the linguistic and statistical analysis of the discourse about the wine industry and on the other hand, the establishment of patterns to develop semiautomatic recognition systems to facilitate the information indexing. The hypothesis raised in this study is that deverbal nominalization formally described by  $[X] v \rightarrow [ [X] v -\text{ção} ] N$  is predominant in the wine industry semantic field and so it is present in the lexicon of scientific and technology papers in this area as their most important index functions based on their frequencies of occurrences. Therefore, the theoretical framework used was the textual genre analysis, the lexical theory and information indexing based on bibliometrics laws and principles. To identify cases of those lexico-morphological features in such terminology, four papers about the wine industry were selected which were published in the *Scientific Electronic Library Online*. These papers were processed by *Software Rank Words 2.0*. For each paper, a list of words frequencies of occurrences was obtained. Next, in each list it was investigated the Zipf's Laws and Goffman Transition Point adequacy, the recurrence of deverbal nominalization described by  $[X] v \rightarrow [ [X] v -\text{ção} ] N$ ,  $[X] v \rightarrow [ [X] v -\text{mento} ] N$ ,  $[X] v \rightarrow [ [X] v -\text{ncia} ] N$ ,  $[X] v \rightarrow [ [X] v -\text{agem} ] N$ ,  $[X] v \rightarrow [ [X] v -\text{da} ] N$  and also verified the relative frequencies of occurrence of these words and the different degrees of relevance of their functions as an index term for the information indexing. Aiming data confrontation, the method is applied into two text about Economics. Finally, given the data obtained the established hypothesis is confirmed, especially in Economics and the results also show the importance of studies developed on deverbal nominalization theory and practice for the scientific and technological information indexing, in the Information Science.

**KEYWORDS:** Genre Analysis. Zipf's Laws. Goffman's Transition Point. Scientific and Technological Discourse. Deverbal Nominalization. Information Indexing.



## RÉSUMÉ

GUEDES, Vânia Lisbôa da Silveira. *Nominalizações deverbais em artigos científicos: uma contribuição para a análise e a indexação temática da informação*, 2010. 109 f. Tese (Doutorado em Linguística) – Faculdade de Letras, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2010.

Cette étude analyse le discours scientifique et technologique sur l'industrie du vin, en portugais, en faisant des investigations sur la récurrence des nominalisations des verbes dans le lexique du genre article de périodique et ses fonctions d'index dans l'indexation. Ainsi, l'objectif est, par un côté, développer une analyse linguistique lexicale-morphologique du discours sur l'industrie du vin et, par l'autre, contribuer pour le développement de modèles de systèmes semi-automatiques d'indexation. On a établi l'hypothèse que les nominalisations des verbes formellement décrites par  $[X] v \rightarrow [ [X] v -\text{ção} ] N$  sont prédominantes dans le champ sémantique de l'industrie du vin et donc, sont présentes dans le lexique des articles de périodique, dans ce domaine de connaissance, en exerçant des fonctions d'index plus importantes, en raison de leur fréquence d'occurrence. À cette finalité, on doit considérer comme rapport théorique l'analyse de genres textuels, la théorie lexicale et l'indexation de l'information, basée dans les lois et principes bibliométriques. De là, sont sélectionnés quatre articles sur la l'industrie du vins, publiés dans la *Scientific Electronic Library Online*. Les articles ont été processés par le *Software Rank Word 2.0*. Pour chaque article, une liste a été élaborée, dans l'ordre décroissant de fréquence d'occurrence des mots. Dans chaque liste, il est vérifiée l'application des lois de Zipf et Point de Transition de Goffman, en étant investiguée la récurrence des nominalisations des verbes formulées d'accord  $[X] v \rightarrow [ [X] v -\text{ção} ] N$ ,  $[X] v \rightarrow [ [X] v -\text{mento} ] N$ ,  $[X] v \rightarrow [ [X] v -\text{ncia} ] N$ ,  $[X] v \rightarrow [ [X] v -\text{agem} ] N$ ,  $[X] v \rightarrow [ [X] v -\text{da} ] N$  et aussi vérifiée leurs fréquences relatives et degrés différents d'importance de leur fonctions d'index, pour l'indexation de l'information. En objectivant à comparer les données, la méthode est appliquée à deux articles sur l'Économie. Enfin, il est mentionné que les données obtenus corroborent l'hypothèse mentionnée ci-dessus, montrent une plus grande adaptation de la méthode aux articles sur l'Économie et mettent en évidence l'importance de l'abordage théorique et descriptive de la nominalisation des verbes pour l'indexation de l'information scientifique et technologique au sein de la Science de l'Information.

**MOTS CLÉS:** Nominalisations des Verbes Analyse de Genres Textuels. Analyse Linguistique Lexicale-morphologique du Discours. Lois de Zipf. Point de Transition de Goffman. Articles Scientifiques et Technologique. Indexation de L'information.

## 1 INTRODUÇÃO

Este trabalho desenvolve uma análise morfológica de textos científicos, editados em artigos de periódicos, com base nos pressupostos da Teoria Lexical. O estudo focaliza os processos produtivos de nominalização deverbal regular e se vale ainda de recursos quantitativos utilizados na indexação temática para o exame dos dados, sob uma perspectiva bibliométrica, no âmbito da Ciência da Informação.

Em suas reflexões acerca das estruturas lexicais no português, Basílio (1999, p. 60) compreende como “construções morfológicas regulares [...] aquelas cujas propriedades semânticas e fonológicas correspondem exatamente ao previsto pela junção da base e do processo morfológico correspondente”. Albino (1993, p.23), ao analisar os sufixos nominalizadores *-ção* e *-mento*, considera que são “[...] regulares somente as formas cujas possíveis interpretações sejam previsíveis a partir da semântica dos verbos correspondentes”. Uma forma é classificada como regular se, no contexto de ocorrência, puder ser interpretada de acordo com sua função semântica, prevista no processo morfológico correspondente. Para Albino (1993, p.23) “uma forma nominalizada, dependendo do contexto em que se insere e da semântica do verbo, pode apresentar uma interpretação verbal - ato, processo ou fato - ou uma interpretação nominal - visão abstrata nominal, resultado concreto, estado.”

A motivação principal para a pesquisa prende-se fundamentalmente aos seguintes objetivos e metas: (a) verificar o grau alto de informatividade dos textos acadêmicos por meio da predominância de ocorrência de nominalizações deverbais; (b) determinar a frequência relativa de formação de nominalizações deverbais com os sufixos nominalizadores *-ção*, *-mento*, *-ncia*, *-agem* e *-da*; (c) fornecer algumas evidências da sistematicidade dos sufixos mais relevantes *-ção* e *-mento*, na escrita de periódicos científicos; (d) ratificar os indicadores já existentes acerca do gênero acadêmico por meio dos procedimentos sucintamente mencionados em (a) e (b); (e) investigar (a) também à luz de modelos quantitativos utilizados na indexação automática no âmbito da Bibliometria na Ciência da Informação; (f) com base na justificativa descrita em (d) e (e), contribuir para o enriquecimento do arcabouço teórico e prático da área de indexação da informação, no esforço de produzir índices temáticos importantes; (g) fortalecer a interface da Linguística com a Ciência da Informação.

As justificativas de (a) a (g) oferecem sustentação para as questões e hipóteses do trabalho que estão mais explicitadas no capítulo 2. As áreas da Ciência da Linguagem e da Ciência da Informação concorrem para o empreendimento desta pesquisa, na medida em que a pesquisa-problema que se impõe como desafio é o de demonstrar como uma análise morfológica sistemática de nominalizações deverbais no gênero artigo acadêmico pode contribuir para a indexação temática de modo a tornar os sistemas de busca mais inteligentes, econômicos e cientificamente constituídos.

A estrutura do texto desta tese se compõe dos seguintes capítulos. O primeiro introduz de forma panorâmica o trabalho. O segundo lança as questões e as hipóteses de pesquisa em consonância com o objeto de estudo. Os pressupostos teóricos acham-se resumidos no terceiro capítulo. O capítulo de metodologia descreve os critérios utilizados para a constituição da amostra e para os procedimentos de análise dos dados. Apresenta ainda os indicadores bibliométricos que alicerçam a análise quantitativa. No capítulo 5, apresentam-se a descrição, a interpretação e a discussão dos resultados, utilizando-se gráficos, tabelas e quadros, ligados a cada artigo estudado. No capítulo 6, são sintetizados os resultados confrontando-se os padrões importantes encontrados. As considerações finais e sugestões para pesquisas futuras encerram o texto, no capítulo 7, salientando alguns obstáculos encontrados durante o desenvolvimento do estudo e sugerindo novas direções de pesquisa.

## 2 QUESTÕES, OBJETIVOS E HIPÓTESES

O crescimento exponencial da informação científica e tecnológica e o conseqüente aparecimento de novos domínios especializados do conhecimento têm-se refletido, ao longo do tempo, em diversos contextos, como, na cultura, na documentação, na comunicação e, mais especialmente, na Linguística e na Ciência da Informação. Nesses domínios especializados do conhecimento, Méndez Rodríguez & Moreira González (1999) mencionam que a comunicação científica apresenta uma estrutura linguística e se estabelece em linguagem natural. Fundamentalmente, o discurso científico, na opinião desses autores, é caracterizado pela recepção e emissão de informação qualificada, pela organização estrutural útil à ciência, pelo modelo lógico-formal, pelo vocabulário especializado e pela determinação mais sistemática comparando-se ao nível baixo de especificidade da denominada linguagem comum.

Para Gil (2003, p.113), nos domínios de discursos técnico-científicos, “a renovação e o enriquecimento lexicais do sistema linguístico” se processam em ritmo mais intenso do que no sistema linguístico comum, acompanhando a dinâmica da Ciência e da Tecnologia. O autor esclarece que os sistemas linguísticos especializados compartilham de algumas características dos sistemas linguísticos gerais e apresentam outras características específicas. Assim, no sistema linguístico científico e tecnológico dos diferentes domínios do conhecimento, o léxico representa um subsistema do sistema linguístico pertencente a todos os falantes de uma dada língua. Cabré (1993) destaca ainda que as línguas de especialidade utilizam um fundo lexical comum e um vocabulário restrito, formado por unidades lexicais próprias de um domínio específico do conhecimento.

Bakhtin (2000) menciona que o gênero de discurso científico é um gênero de discurso secundário. Ao estabelecer a distinção entre gênero de discurso primário (simples) e gênero de discurso secundário (complexo), o autor destaca que os secundários surgem em situações de comunicação cultural mais complexa e relativamente mais culta, especialmente no que se refere à comunicação escrita. Segundo Bakhtin (2000, p. 281), durante a sua composição, os gêneros de discurso secundários se apropriam e transformam os gêneros primários, que se formaram “em circunstâncias de uma comunicação verbal espontânea”. Nesse sentido, Bakhtin chama a atenção para o fato de que o enunciado do gênero reflete as condições e as finalidades específicas de cada domínio de atividade humana através do conteúdo temático e

da seleção dos recursos linguísticos lexicais e gramaticais, isto é, do mecanismo estilístico adotado.

Este estudo tem por objetivo principal investigar as nominalizações deverbais regulares, recorrentes no gênero artigo de periódico, na área de indústria de vinhos, visando à demonstração da importância da análise do grau de produtividade de nominalizações deverbais no discurso acadêmico, na busca de evidências empíricas, sobretudo, de que o padrão de formação de nominalização deverbal, formalmente representada, de acordo com Basílio (1980) por  $[X] v \rightarrow [ [X] v -ção ] N$ , é predominante no campo semântico da indústria de vinhos, contribuindo para o conhecimento das estratégias léxico-morfológicas mais frequentes em textos científicos e para o refinamento da indexação da informação por categorias morfológicas. Assim, são investigadas as nominalizações deverbais regulares, empregadas no discurso escrito em artigos acadêmicos sobre a indústria de vinhos, em português, e suas funções de índices mais relevantes do ponto de vista da indexação. O objetivo não é dar conta de todo o processo de indexação, mas o de contribuir para o fortalecimento da interface Linguística e Indexação, sobretudo no que se refere à abordagem teórica e descritiva, como diria Basílio (1979a, p. 233), “do processo de identificação do conteúdo informativo” de artigos científicos e tecnológicos. Nesse contexto, Fujita (1992) salienta que a indexação, apoiada na Ciência da Linguagem, vem fortalecendo seu campo teórico e metodológico, com o intuito de atingir maior grau de precisão na recuperação da informação.

A indexação consiste na análise dos conceitos abordados em um determinado documento e sua representação, em linguagem natural (termos integrantes do documento), ou em linguagem controlada (termos integrantes de sistemas pré-estabelecidos). Assim, o conteúdo temático de um documento pode ser representado de maneira concisa, por meio de termos de indexação ou palavras-chave, visando à precisão na recuperação da informação.

Na Ciência da Informação, a indexação tem sido um dos temas mais relevantes de pesquisa, na medida em que o índice é o elo vital entre os dados bibliográficos armazenados e a recuperação da informação. Mendez Rodriguez & Moreira González (1999) destacam que a indexação é uma técnica de classificação do conteúdo tanto do documento quanto das consultas enunciadas pelos usuários com o objetivo de reter as idéias mais significativas representadas por

termos de indexação. Os autores esclarecem que os termos de indexação podem ser extraídos da linguagem natural, utilizada pelos autores, ou de um vocabulário controlado, selecionado *a priori*.

Na indexação, utilizando-se indicadores bibliométricos, a contagem das frequências de ocorrência das palavras refere-se à etapa inicial da análise bibliométrica do discurso escrito em documentos. Seu produto é uma lista em ordem decrescente de frequência de ocorrência das palavras que compõem o texto do documento. Esse procedimento constitui o ponto de partida para a aplicação das Leis de Zipf e do Ponto de Transição (T) de Goffman<sup>1</sup> que, por sua vez, associados, indicam a região de concentração de palavras de alto conteúdo temático, que, segundo Rouault (1987), são representantes das categorias morfológicas informativas: nomes, verbos e adjetivos. Parte-se do princípio de que as palavras mais frequentes, excluindo-se as não significativas, tais como artigos, preposições, conjunções e verbos auxiliares, são as mais representativas do conteúdo temático e, portanto, devem ser utilizadas como termos de indexação ou palavras-chave para representar a informação contida no documento.

Assim, a pesquisa baseia-se nas abordagens teóricas e descritivas referentes à nominalização deverbal, na Teoria Lexical, sobretudo nos estudos de Basílio (1979a; 1979b; 1980; 1993; 1999; 2004a; 2004b; 2007) sobre Morfologia Derivacional no português falado e de Albino (1993) sobre condições de produção de sufixos nominalizadores *-ção* e *-mento*, no português escrito formal, assim como nos estudos bibliométricos de Pao (1978, 1989) ligados às leis de Zipf e ao Ponto de Transição de Goffman, enquanto mecanismos semi-automáticos de indexação de artigos científicos e tecnológicos. Somam-se a esses enfoques, a análise de gêneros desenvolvida por Swales (1990; 2004), Bazerman (2006), Bakthin (2000), Bhatia (2004) e Hyland (2009), voltada especialmente para o discurso no gênero textual artigo acadêmico. Os autores apontam o uso de nominalizações como um dos aspectos que caracterizam os gêneros acadêmicos que, na visão bakthiniana, é um gênero secundário marcado por processos léxico-morfológicos.

A nominalização, segundo Basílio (2007, p.78), refere-se ao “conjunto de processos que formam substantivos a partir de adjetivos e, sobretudo, de verbos”. Considerando sua função

---

<sup>1</sup> O Ponto de Transição de Goffman determina graficamente a localização onde ocorre a transição das palavras de baixa frequência para as de alta frequência. Existe uma região, ao redor deste ponto, com probabilidade de concentrar as palavras de alto conteúdo temático e, portanto, aquelas que seriam utilizadas na indexação de um texto em análise (GUEDES & BORSCHIVER, 2005).

discursiva, a autora destaca (1) a função expressiva e (2) a função textual. A primeira volta-se para as funções envolvidas com a expressão de atitudes subjetivas em relação ao objeto do enunciado. A segunda refere-se à adequação da utilização de itens e/ou classes lexicais e/ou gramaticais ao gênero de discurso e à construção do enunciado. A autora esclarece que a nominalização contém um aspecto sintático e desempenha funções textuais e semânticas, sendo que a função semântica da nominalização é a de denominação, possibilitando a referência a um processo verbal, bem como a um evento, uma ação, um estado (BASÍLIO, 2007, p. 83).

Halliday (1994) afirma que, por meio da nominalização, processos normalmente representados por verbos e propriedades normalmente representadas por adjetivos surgem como substantivos. Para Huston (1994), os textos científicos precisam ser objetivos e neutros cientificamente, isto é, impessoais, e o emprego da nominalização possibilita atingir tais propriedades.

Basílio (1989) defende que os nomes abstratos, formados a partir de verbos, permitem atender a exigências sintático-semânticas do discurso, o que justifica o alto grau de produtividade dos processos de nominalizações deverbais. Em sua abordagem acerca da nominalização na estrutura textual, Basílio (2007) salienta que a composição de substantivos a partir de verbos é um dos processos mais produtivos de formação de palavras no português, devido à função que a palavra nominalizada apresenta na estrutura do texto escrito. Segundo Basílio, uma das funções da nominalização no texto é a de possibilitar a compactação textual sem perda de informações. Outra função, provavelmente a mais relevante, é a anafórica: a de recuperação das estruturas verbais pelo uso do verbo na forma nominal. Também identificada pela autora, é a função de caracterizar o tipo de discurso. Basílio (2007), assim como Bakhtin (2000), Bhatia (2004), Swales (1990), Hyland (2009) e outros autores dedicados à análise de gêneros, destacam que o uso acentuado de nominalizações marca fortemente o discurso formal escrito e, principalmente, o discurso científico.

Ao discutir a representação e o acesso lexical por radicais e regras de afixação ou por palavras previamente existentes, Basílio (1999) também ressalta que a nominalização deverbal pode ter função designadora, empregando o significado básico do verbo para denotar conceitos, processos etc. Salienta a autora que a nominalização denotativa é comum na linguagem científica. “Palavras como neutralização, derivação, referência e outras, da terminologia

linguística, são formadas com propósitos designadores” (BASÍLIO, 1999, p. 63). Há de se ressaltar igualmente a função da nominalização de compactação textual sem perda de informação, mencionada por Basílio (2007), por possibilitar atingir alto grau de precisão e de economia, que caracteriza o discurso científico.

A questão de pesquisa, que norteia este trabalho, refere-se à necessidade de maior compreensão desses processos morfológicos presentes no léxico dos artigos científicos e tecnológicos na área da indústria de vinhos e também à busca de maior precisão na indexação e na recuperação da informação nesse domínio de discurso. Atualmente, tanto na saúde quanto na Economia, como na gastronomia, na nutrição, na agronomia e em outros campos do conhecimento, a indústria de vinhos é uma importante área de investigação.

A escolha de textos sobre vinho se deve a grande demanda de estudos na área de indústria de alimentos, na qual a área de indústria de vinhos carece de trabalhos como o aqui desenvolvido. Dados estatísticos, apresentados por Mello (2010), mencionam que, em 2009, foram exportados 25,51 milhões de litros de vinhos de mesa, 146,5 % superior ao ano de 2008. Por outro lado, ocorreu o aumento de 2,79 na quantidade importada de vinhos e crescimento de 6,46 % no valor das importações.

A indexação e a recuperação da informação científica e tecnológica sobre a indústria de vinhos são fundamentais, uma vez que os pesquisadores necessitam acessar, sistematicamente, informações relevantes na literatura, no caso de vinhos, para adquirir conhecimentos sobre o cultivo de uvas e a produção de vinhos na área de vinicultura em geral.

Os métodos atuais de análise do discurso em artigos de periódicos, indexação temática e recuperação da informação não satisfazem à demanda e à rápida localização da informação. A necessidade de apropriação de conhecimentos se processa por meios linguísticos e bibliométricos que devem favorecer a elaboração de novos modelos semi-automáticos de indexação, no cenário dos sistemas de recuperação da informação.

Deve-se destacar que a virtualização de acervos bibliográficos, em bases eletrônicas de dados, vem revelando a necessidade crescente de refinamento do processo de indexação da informação. Existe grande flexibilidade para a inserção de informação, de um lado, e falta de



critérios científicos para a análise e indexação da informação, por outro. Analisar morfologicamente os artigos de periódicos científicos sobre a indústria de vinhos, à luz dos fundamentos teóricos da Teoria Lexical e sob a ótica da análise de gêneros textuais e da Bibliometria, oferece contribuição, seja à compreensão sobre nominalizações deverbais e acerca de gêneros acadêmicos, seja sobre fenômenos mais complexos de análise do conteúdo informativo e indexação da informação.

A produção de nominalizações deverbais, presentes no discurso científico e tecnológico do gênero textual artigo de periódico, pode ser mais bem entendida por meio da observação dos graus relativos de seus distintos padrões de formação e do desempenho de suas funções de índice mais relevante, no processo de indexação da informação. Conjugando a Linguística Textual, a Teoria Lexical e a Bibliometria parece uma orientação apropriada para a pesquisa em Indexação.

Postulam-se para o estudo os seguintes pressupostos:

- (1) O padrão de nominalização deverbal formalmente representado por  $[X] v \rightarrow [ [X] v -\text{ção} ]$   $N$ , é predominante em textos sobre a indústria de vinhos.
- (2) A frequência do item e processos léxico-morfológicos de nominalizações deverbais, das bases e dos sufixos, mantém relação com o gênero textual acadêmico.
- (3) Textos sobre indústria de vinhos apresentam vocabulários específicos deste domínio de discurso, marcados por graus relativos de produtividade de padrões de nominalização deverbal.
- (4) Do ponto de vista bibliométrico, as nominalizações deverbais de maior relevância ocorrem na Região de Transição de Goffmann<sup>2</sup>.
- (5) O estudo das nominalizações deverbais contribui para a adequação do método bibliométrico semi-automático de indexação em análise, sobretudo no que se refere à identificação de bases e sufixos de alto conteúdo informativo.

---

<sup>2</sup> Região na qual, de acordo com Goffman, deveria ocorrer a concentração de palavras com maior conteúdo temático, em relação a um texto analisado (PAO, 1978).

### 3 PRESSUPOSTOS TEÓRICOS

Neste item, são apresentados os pressupostos teóricos que alicerçam a pesquisa de doutoramento. Inicialmente, discorre-se sobre a Teoria Lexical. A seguir, abordam-se alguns postulados sobre gêneros textuais. Finalmente, versa-se sobre determinados aspectos da Ciência da Informação pertinentes ao estudo em tela.

#### 3.1 BASES DA TEORIA LEXICAL

A Morfologia é uma subárea em Linguística, que se ocupa da estrutura das palavras. Todas as línguas possuem palavras e, em todas as línguas, as palavras apresentam uma estrutura interna, formada por um ou mais morfemas. Assim, a forma *gatos* compreende o radical e a vogal temática nominal "gato", à qual se acrescenta o morfema "s" indicando plural. Em outras palavras, "s" é um sufixo e, ainda, um morfema vinculado ao morfema "gato" (SPENCER, 2003).

Spencer & Zwicky (2001) consideram a Morfologia como o estudo da estrutura das palavras as quais representam a interface Fonologia, Sintaxe e Semântica. As palavras têm propriedades fonológicas, se articulam junto a frases e sentenças; suas formas refletem geralmente suas funções sintáticas e são usualmente compostas de pequenas partes significativas.

Aronoff & Anshen (2001) defendem igualmente que a Morfologia trata inicialmente da estrutura interna das palavras potencialmente compostas de uma língua. Algumas dessas palavras podem não existir, mas todas se amoldam à estrutura morfológica da língua em questão. Aronoff (1994) salienta que a edição abreviada do *Oxford English Dictionary* conceitua Morfologia, na Linguística, como a área da gramática ligada à formação e flexão de palavras.

As palavras, segundo Matthews (1974), são reunidas para formar sentenças e se diferem uma das outras em som e significado, referindo-se a diferentes espécies de objetos. São distinguidas por várias consoantes no início e no final (e.g. *clock* e *gong*). As sentenças também diferenciam-se uma das outras em som e significado: *The clock has been sold* é diferente de *The gong has been sold*. Nos sistemas de línguas naturais, diferenciam-se

categorias de nomes, verbos, dentre outras. O autor acrescenta que são relevantes as questões morfológicas como objeto de estudo.

Câmara Jr. (1985) postula o conceito de palavra morfológica abrangendo nome, adjetivo e verbo, classes abertas, e palavras funcionais, como preposição, conjunção e determinativos, classes fechadas. Sendo a língua estruturada, convém investigar, do ponto de vista da origem, as palavras portuguesas que integram os diversos campos semânticos, no que se refere ao nível lexical.

O léxico, segundo Basílio (2004b, p. 50), pode ser conceituado como “um sistema dinâmico de produção e armazenagem de formas simbólicas, isto é, formas que evocam significados”. A autora acrescenta que o léxico não é apenas um conjunto de palavras, mas um sistema que também engloba estruturas de expansão, das quais as mais relevantes correspondem aos processos de formação de palavras. Em Basílio (2007, p. 7), destaca-se que “as palavras, ou itens lexicais, são elementos básicos que utilizamos para formar enunciado”. Basílio (1999) enfatiza que os estudos morfológicos no português do Brasil caracterizam-se pelo interesse nas questões lexicais. Uma segunda vertente desta preferência é a conexão semântica. Para a Morfologia Distribuída a noção de palavra toma outra acepção. Importante salientar que, no quadro teórico conceitual, não está contemplada a Morfologia Distribuída, ainda que seja relevante sua consideração nos estudos sobre a formação de estruturas nominalizadas, conforme verificado em Medeiros (2009).

De uma perspectiva estrutural, Cruse (2003) ressalta que uma língua se compõe, minimamente, de um conjunto de unidades básicas e de um conjunto de regras para a combinação dessas unidades básicas em frases e sentenças. Uma lista de unidades básicas constitui o léxico da língua, enquanto que uma especificação das regras combinatórias dessas unidades lexicais constitui a gramática. As unidades básicas devem ter forma e significado. As regras combinatórias indicam como as formas complexas são permitidas na língua e como podem ser computados seus significados.

Aronoff & Anshen (2001) defendem que o léxico de uma língua é uma lista de itens existentes nessa língua, considerados signos arbitrários. Segundo eles, a maior parte dos itens lexicais de uma língua é composta por palavras.

No que se refere à noção de palavra, Basílio (2007, p. 13) entende que, na língua escrita, o conceito de palavra refere-se a “qualquer sequência que ocorra entre espaços e/ou sinais de pontuação”. A autora chama a atenção para o fato de que, frequentemente, a palavra é uma unidade linguística básica, facilmente identificada pelo falante em sua língua nativa. As palavras, na opinião de Basílio (2007, p. 15-16), são compostas por “uma sequência de elementos constitutivos”. No caso de estrutura morfológica complexa, as palavras são compostas por mais de um elemento e são estruturadas pela combinação de uma base e um afixo. Segundo a autora, são palavras constituídas de vários elementos como, por exemplo, o adjetivo *infeliz*, composto pelo adjetivo *feliz* e o prefixo negativo *in-*; o substantivo *oralidade*, composto pelo adjetivo *oral* e pelo sufixo *-idade* e o substantivo *guarda-chuva*, formado pelos elementos *guarda*, tema do verbo guardar, e *chuva*, substantivo (BASÍLIO, 2007). Anteriormente, a autora assinala que as palavras têm a função de nomear o mundo. Se não ocorresse a formação de novas palavras a partir de determinadas bases, o sistema linguístico ficaria sobrecarregado e exigiria um esforço colossal de memória para que nos apropriássemos da língua como um todo.

O processo de formação de palavras por derivação se caracteriza pelo acréscimo de um afixo a uma base, obtendo-se a estrutura *base + afixo*, tanto base + sufixo (*indexar + -ção*), quanto prefixo + base (*re + classificar*). Uma das funções mais indicadas na literatura para a derivação é a de mudança de classe (exemplo: nominalização deverbal, transformação de um verbo em um substantivo - fermentar  $\Rightarrow$  fermentação). Entretanto, as classes das palavras podem ser definidas segundo critérios semânticos, sintáticos e morfológicos. As gramáticas normativas dedicam maior atenção ao critério semântico. O estruturalismo privilegia o morfológico e o funcional. O gerativismo, na sua versão padrão, volta-se para o critério sintático. No entanto, um item lexical contém propriedades morfológicas, sintáticas e semânticas, de modo que, “sua pertinência a classes deve ser estabelecida em termos morfológicos, semânticos e sintáticos” (BASÍLIO, 2007, p. 60).

Aronoff & Anshen (2001) postulam que a Morfologia e o léxico competem entre si e a competição ocupa papel central no amplo sistema da linguagem. Entretanto, a Morfologia, que forma palavras de palavras, encontra as bases sobre as quais ela opera no léxico. Logo, apesar de a Morfologia e de o léxico serem distintos e competirem entre si, a Morfologia depende do léxico, visto que as bases das palavras morfológicamente complexas são normalmente entradas lexicais. Assim, a Morfologia cria palavras regulares e o léxico

armazena palavras irregulares. Sob essa perspectiva, o léxico mental de um indivíduo é composto de uma lista de itens irregulares estocados na mente. Qualquer palavra armazenada no léxico mental de um indivíduo é considerada uma palavra existente. Por outro lado, uma palavra que atende a todos os critérios para ser considerada uma palavra em uma língua, mas que não está no léxico mental do falante, embora possa existir no léxico de outro falante, é considerada uma palavra potencial do ponto de vista morfológico. Sendo assim, em termos de léxico mental, pode-se estabelecer a distinção entre palavras existentes e palavras potenciais.

Na gramática tradicional, o léxico é definido como um conjunto de palavras que compõem a língua e a palavra é considerada a unidade mínima de análise linguística. Nesse contexto, como uma decorrência do modelo palavra-paradigma, os estudos se concentram na flexão, variação na forma da palavra, e na sintaxe, combinação das palavras na sentença. No século XIX, esse modelo vai sofrer um confronto maior, devido à proposição da linguística como ciência mais próxima do conceito de ciência atual, já adotado naquele momento. As palavras passam a ser analisadas em dois níveis: formas básicas e formas resultantes de uma série de processos que se aplicam às formas básicas. Ocorre o desmembramento da palavra em pequenas partes internas, obedecendo ao paradigma do método comparativo. A hipótese é a de que a análise comparativa das línguas possibilita verificar se elas pertencem, ou não, a um mesmo tronco, historicamente. Os critérios adotados eram a semelhança e a correspondência sistemática das raízes. Nessa etapa, já existia a preocupação com a divisão da palavra em constituintes. No Estruturalismo, que surge em oposição à Linguística Histórica, o efeito do método comparativo permanece. Não há mais como conceber a palavra como uma unidade mínima de análise.

No Estruturalismo, ocorrem então a ruptura com o diacronismo, que caracteriza os estudos linguísticos no século XIX, e a adoção do conceito de sincronismo. As pesquisas voltam-se para a determinação do *status* dos itens formativos das unidades lexicais. Como representante do estruturalismo europeu, Saussure defende a idéia de língua como sistema de valores, que se encadeiam em termos da noção de signo (signo = relação significado + significante). O Estruturalismo europeu modifica a idéia de léxico como conjunto de palavras, passando a concebê-lo como um sistema de signo na mente do falante. No entanto, esse paradigma não dá conta da distinção entre os tamanhos ou as extensões dessas unidades reconhecidas como signos. Quais são os limites entre os morfemas, as palavras, as frases e os textos? Preconiza-se que a frase e o texto representam os atos de fala comparativamente aos signos.

Nessa época, representando o Estruturalismo americano, Bloomfield (1926), ao analisar centenas de línguas indígenas, volta-se também para a análise morfológica. Assim, inaugura-se o Estruturalismo concentrado na estrutura do enunciado e não no processo que ocorre na mente do falante, defendida por Saussure. Bloomfield defende a importância fundamental de reconhecimento da palavra antes de se estabelecer o que é a palavra. Surge a palavra como forma mínima livre, resolvendo a questão da distinção entre palavra, morfema e sentença, não solucionada no modelo de Saussure.

A Hipótese Transformacionalista de Chomsky (1965) volta-se para a correspondência sintático-semântica existente entre estruturas nominais e estruturas verbais correspondentes. Ou seja, a idéia é a de que a estrutura verbal e a estrutura nominalizada têm o mesmo significado. Na versão do modelo em referência, os aspectos morfológicos não eram contemplados, de modo que as nominalizações deverbais são justificadas por meio de regras transformacionais. Há problemas na descrição de formas derivadas que não têm verbos correspondentes (*condição*) ou representam extensão de sentido (*redigir* → *redação*).

O léxico da Teoria Padrão fornece as bases para as formas nominalizadas. Assim, na Teoria Padrão, o léxico reduzido e simples apresenta generalizações, contendo menor número de símbolos e de formas lexicais. A partir dele, são geradas formas operadas pelas regras de transformações lexicais ou regras de transformações das bases. No léxico reduzido e simples, só vai aparecer o que não tem elementos adicionais, obedecendo ao princípio de generalização e eliminação de redundâncias.

Surge a Hipótese Lexicalista de Chomsky (1970) que propõe investigar as nominalizações como derivados nominais no próprio léxico e introduzir regras no componente de base para dar conta das estruturas nominais. Assim, os estudos de Morfologia Derivacional iniciam-se com a Hipótese Lexicalista, segundo a qual as relações entre formas, tais como verbos e suas nominalizações devem ser analisadas e descritas no contexto lexical.

Nesse contexto, o modelo de representação lexical elaborado por Jackendoff (1975) defende a noção de relação lexical associada à facilidade de aquisição do léxico. De acordo com a Teoria da Entrada Plena em contraposição à entrada reduzida dos modelos anteriores, duas palavras relacionam-se lexicalmente na medida em que o conhecimento prévio de uma facilite a aquisição de outra ( *classificar* → *classificação*). Assim, o que é previsível por relações

lexicais não apresenta custo de aquisição. Além disso, ao observar que estruturas nominalizadas distintas podem ter o mesmo sentido e que uma mesma estrutura nominalizada pode apresentar diferentes interpretações, o autor propõe duas entradas lexicais, estabelecidas por regras de redundância morfológica e de redundância semântica. Apesar de sua importância para a descrição de relações lexicais, esse modelo não abarca o fenômeno da produtividade lexical.

Em Aronoff (1976, p. 36), a noção de produtividade lexical é considerada basilar no âmbito da Morfologia Derivacional. Apesar de o termo produtividade ser amplamente usado, discute-se uma característica específica para distinguir as Regras de Formação de Palavras (RFP) produtivas das não produtivas. Segundo o autor, existem restrições morfológicas ligadas aos tipos de palavras que se podem utilizar como base sobre a qual são operadas as regras de formação de palavras. Assim, *-ção* e *-mento* (fermentação e encaixotamento) por exemplo formam nomes a partir de verbos, apesar de somente *-ção* ser restrito a verbos latinos. Com o intuito de redução da classe potencial de palavras, Aronoff desenvolve um princípio segundo o qual uma nova palavra é gerada pela aplicação de uma RFP a um item lexical pré-existente no léxico. Aronoff também propõe as noções de graus de produtividade e de bloqueio.

Bybee (1988, p. 121) postula que as regras e as representações correspondem a pontos polares de um *continuum* e não a elementos inteiramente distintos e separados. O modelo também introduz os conceitos de conexão lexical e força lexical nos estudos morfológicos. Bybee (2007) destaca a frequência do item como um vetor facilitador para o seu acesso, diretamente relacionado ao grau de força lexical das relações morfológicas do item. Ela salienta que:

Palavras inseridas no léxico têm graus variados de força lexical. Principalmente devido a sua frequência de ocorrência, as palavras com força lexical intensa são fáceis de serem acessadas e servem como base de relações morfológicas, exibindo uma autonomia que as torna então resistentes à mudança e propensas à independência semântica (BYBEE, 2007, p.169, tradução nossa).

Basílio (1980, p.73-74) conceitua nominalização como processo de derivação de palavras determinado por “uma relação paradigmática geral entre verbos e nomes na língua”. A autora salienta que a nominalização consiste na associação lexical sistemática entre nomes e verbos correspondentes. Segundo Basílio, de acordo com o “significado extralingüístico”, no estruturalismo, “toda forma nominalizada é semanticamente relacionada ao verbo

correspondente, já que nome e verbo partilham, ao menos parcialmente, um significado lexical.”

Na opinião de Cunha & Cintra (1985, p.97), pode-se interpretar semanticamente as nominalizações deverbais com o sufixo *-ção* como *ação ou resultado da ação* (nomear → nomeação) expressa pela base verbal correspondente. Já as nominalizações deverbais com o sufixo *-mento* podem ser interpretadas semanticamente por meio de três formas distintas: (a) *ação ou resultado da ação* (ferir → ferimento), (b) *instrumento da ação* (ornar → ornamento) ou (c) *noção coletiva* (acolher → acolhimento).

Em Basílio (2004b), salienta-se que as nominalizações deverbais são classificadas em nomes de ação, cujo significado permanece como verbal, e nomes de agente e instrumento, que designam seres e objetos baseado no significado do verbo. Basílio (1980) afirma que a formação de nomes de ação deverbal indica particularmente função gramatical com a finalidade de uso da noção verbal em contextos sintáticos que demandam um substantivo. No entanto, os substantivos deverbais também podem revelar motivação denotativa, quando formados com a finalidade de designar seres, processos, eventos, situações, a partir do significado verbal (BASÍLIO, 2004b, p. 53-54).

Basílio (2004b) esclarece que, no português, quando ocorre a utilização do verbo em estruturas sintáticas verbais, o verbo é composto por categorias gramaticais de tempo, modo, aspecto, pessoa e número. Como essas exigências relacionadas ao uso do verbo não ocorrem no substantivo, a utilização do substantivo em substituição ao verbo correspondente não requer o uso de agentes, objetos ou beneficiários do processo verbal, assim como as indicações de tempo, modo e número/pessoa deixam de ser compulsórias (BASÍLIO, 2004b). A nominalização é construída “para fins de referência ao complexo semântico [...] de uma maneira nominal, ou seja, como a uma entidade em si, independente de instâncias particulares do evento, e suas associações de tempo, sujeito e objeto verbal etc” (BASÍLIO, 2007, p.79).

Em Basílio (2004a), explica-se que a formação de estruturas nominalizadas se deve a três principais motivações: (1) *denotativa*, que se refere ao emprego do significado do verbo correspondente para designar seres ou entidades, (2) *gramatical*, que se refere à adaptação do verbo a contextos sintáticos e (3) *textual*, que se refere a uso da nominalização, como referência a uma estrutura verbal utilizada anteriormente no texto. Abílio (1993), ao analisar o



discurso escrito em português formal, observa a predominância da função semântica denominadora das formas nominalizadas, estruturadas por meio de  $[X] v \rightarrow [ [X] v -\text{ção} ] N$ .

Nesta tese de doutoramento, são discutidos ainda os conceitos relevantes para o tema em foco de produtividade, bloqueio e frequência de bases e sufixos nominalizadores. Defende-se a importância da associação do estudo dos padrões de formação de nominalizações deverbais à análise de indicadores bibliométricos enunciados por Zipf e Goffman, na Ciência da Informação, enquanto mecanismos de identificação da faixa de ocorrência de palavras significativas em artigos científicos e tecnológicos e de seleção de termos de indexação. Nesse sentido, Basílio (1979a, p. 1) menciona que as investigações “[...] na área de estudos lexicais constituem um caminho promissor para o desenvolvimento de pesquisas interdisciplinares em Linguística e Ciência da Informação”.

Enfim, a Teoria Lexical é de grande relevância para a indexação. Ela permite o conhecimento de padrões lexicais produtivos de formação de nominalizações deverbais, assim como o reconhecimento da equivalência semântica das nominalizações deverbais com os verbos correspondentes. Contribui, assim, para a formulação de procedimentos semi-automáticos de análise e de identificação do conteúdo informativo de artigos científicos e tecnológicos.

Segue um exemplo de nominalização deverbal regular investigada no resumo do artigo de Rizzon; Zanuz; Miele (1998), revelando a função denominadora da nominalização deverbal *fermentação*, que corresponde ao processo de fermentar, e a equivalência semântica da nominalização com o verbo correspondente.

[...] Os vinhos foram elaborados por microvinificação na safra de 1995, sendo avaliadas a evolução do pH, acidez total, ácido tartárico e K em cinco fases da vinificação: 1. imediatamente após o esmagamento da uva; 2. na descuba; 3. após a *fermentação* alcoólica; 4. após a fermentação maloláctica; [...].

No trecho, observa-se a equivalência semântica da nominalização deverbal com o verbo correspondente em:

“[...] após a *fermentação* alcoólica [...],”

[...] após o álcool *ter fermentado* a uva.

A forma *fermentação* tem uma função denominadora uma vez que conota visão abstrata nominal do processo de fermentar a uva.

Vale ressaltar que, no trecho acima mencionado, foram utilizadas sete nominalizações deverbais distintas, mencionadas a seguir: *microvinificação*, *evolução*, *acidez*, *vinificação*, *esmagamento*, *descuba* e *fermentação*, sendo que a estrutura nominalizada *fermentação* foi usada na denominação dos processos de *fermentação alcoólica* e de *fermentação maloláctica*.

No que se refere à produtividade e ao bloqueio de estruturas morfológicas teoricamente possíveis, a produtividade morfológica pode ser informalmente definida como a extensão na qual um determinado afixo é provavelmente utilizado na produção de novas palavras na língua. Sob esse prisma, a produtividade é um *continuum* probabilístico que prediz o uso de palavras potenciais. De um lado do *continuum*, estão os afixos completamente não produtivos ou extintos, os quais provavelmente não serão utilizados na invenção de uma palavra. No outro lado do *continuum*, há os sufixos flexionais produtivos como *-ed*, *-ing* e *-s*, em inglês, que são adicionados às bases assim que as condições sintáticas se tornem apropriadas e não haja alguma forma irregular pré-existente que os bloqueie, como também os sufixos derivacionais altamente produtivos como *-ness* e *-action*. No meio, encontram-se os sufixos derivacionais tais como *-ity*. Ressalta-se que os afixos podem variar continuamente quanto à produtividade (ARONOFF & ANSHEN, 2001).

Assim, segundo Basílio (1993), o conceito de produtividade deve ser entendido como a medida do potencial de uma regra para a operação sobre determinadas bases, para produzir construções morfológicamente possíveis. Ao mencionar as restrições de operação sobre determinadas bases, ligadas a cada padrão de formação de palavras, Basílio (2004a) alerta para o fato de que, se os verbos corresponderem à estrutura morfológica *X -izar*, a sua nominalização terá que corresponder a  $[X] v \rightarrow [ [X] v -ção ] N$ .

Existem também fatores quantitativos relevantes para a questão da produtividade. Aronoff (1982 apud ARONOFF & ANSHEN, 2001, p. 245) menciona que a frequência e a produtividade de uma palavra também estão relacionadas entre si. A frequência é importante na seleção de bases e afixos. Em relação à produtividade de afixos, um afixo menos produtivo é encontrado geralmente ligado a bases de maior frequência. Em Aronoff (1976, p. 43), o

fenômeno da não ocorrência de uma nova forma devido à simples existência de outra é denominado bloqueio.

O conceito de bloqueio é também reconhecido como um fenômeno psicológico, baseado nos princípios gerais de economia de expressão. As abordagens teóricas e práticas ao fenômeno de bloqueio de formas morfológicamente regulares destacam a importância da frequência de ocorrência da forma para bloquear outras. Em geral, uma forma irregular usada mais frequentemente, tanto em termos absolutos como comparada à sua base, torna mais possível o bloqueio da forma regular correspondente. Quanto mais frequente uma forma irregular, mais resistente se tornará seu uso em relação à sua substituição por uma forma regular (ARONOFF & ANSHEN, 2001). Aronoff (1976) salienta que a noção de bloqueio é importante para a delimitação do léxico, na medida em que prevê o bloqueio de formas produtivas que, se fossem geradas, proveriam produtos não necessários ao léxico. Em síntese, o autor afirma que palavras novas em uma língua provêm apenas de palavras pré-existentes.

O bloqueio pode ser verificado na derivação quando a produção de uma palavra como *indexamento*, que corresponde à fórmula  $[X] v \rightarrow [ [X] v -mento ] N$ , teoricamente possível, é bloqueada pela existência prévia no léxico da palavra *indexação*, que corresponde à fórmula  $[X] v \rightarrow [ [X] v -ção ] N$  e que semanticamente exerce a mesma função que exerceria a outra forma bloqueada.

Para Basílio (1993) é relevante e necessário considerar-se a frequência e as condições de produção, pois a frequência de aplicação vai depender das condições externas de aplicação como, por exemplo, o gênero discursivo específico, ligado a uma produção concentrada de determinadas RFP. No discurso científico, verifica-se um alto grau de produção de nominalizações. Segundo Basílio, o estudo das condições de produção de construções morfológicas em tipos distintos de discursos é tão importante quanto o estudo das condições de produtividade lexical.

Do ponto de vista da Ciência da Informação, as abordagens teóricas e descritivas à nominalização no âmbito da Teoria Lexical são de grande importância para a indexação, especialmente, na Web Semântica (W3), idealizada por Tim Berners-Lee, inventor da *World Wide Web* (WWW) ou Internet. Berners-Lee; Hendler; Lassila (2001) afirmam que, na representação do conhecimento, os computadores devem ter acesso a coleções estruturadas de

informação e conjunto de regras de inferências que podem ser utilizadas para conduzir deduções automáticas. A base semântica da infra-estrutura da Web 3.0 vai ser constituída por páginas escritas em *eXtensible Markup Language* (XML), que utilizarão metadados e facilitarão o compartilhamento de informações através da Internet, privilegiando os vocabulários e as relações conceituais através de ontologias, por exemplo, para expressar a semântica das novas páginas na Web. O objetivo principal da Web 3.0 é o de melhorar os resultados diferenciados de busca, nos diversos domínios de discurso, destacando-se sobretudo as inferências conceituais em domínios específicos do conhecimento e a interpolaridade dos sistemas.

“Sobre essa estrutura de páginas ricas semanticamente, irão atuar programas especiais, os chamados agentes inteligentes, que terão acesso ao conhecimento embutido nas páginas da Web Semântica para realizar tarefas que demandem conhecimento, raciocínio e dedução.” (MARCONDES, 2001).

### 3.2 CONSIDERAÇÕES SOBRE GÊNEROS TEXTUAIS

A Linguística Textual surgiu na década de 60 do século XX, voltada para a descrição dos fenômenos sintáticos e semânticos dos enunciados e para a sequência de enunciados. Ao longo do tempo, o estudo dos textos, fundamentado por diferentes modelos teóricos, tem focado a atenção para questões referentes à Teoria do Texto (VILELA; KOCH, 2002).

Na década de 70, o texto é visto como uma unidade linguística hierarquicamente elevada, que constitui uma entidade do sistema linguístico, cujas estruturas possíveis, em cada língua, devem ser determinadas pelas regras de uma gramática textual. Chega-se assim à fase da Linguística Textual que se propõe a investigar a constituição, o funcionamento, a produção e a compreensão dos textos. A investigação se estende do texto ao contexto, sendo o contexto entendido como o conjunto de condições de produção, recepção e interpretação dos textos. A partir dos anos 80, os estudos destacam a questão da coesão e coerência textual, admitindo que a coerência se constroi na interação entre o texto e seus usuários, considerando fatores de ordem linguística, social, cognitiva e interacional (VILELA; KOCH, 2002).

Tendências recentes têm dedicado atenção à questão cada vez mais na área da cognição, com ênfase no processamento do texto. Nos últimos anos, também a tipologia textual volta à cena,

sob o enfoque dos gêneros textuais, demonstrando-se bastante relevante para as pesquisas no campo da Linguística Textual. Assim, o texto deixa de ser entendido como uma estrutura acabada, um produto, passando a ser abordado no seu próprio processo de planejamento, verbalização e construção. Nesse contexto, o conhecimento linguístico compreende o conhecimento gramatical e o lexical (VILELA; KOCH, 2002).

Bhatia (2004) menciona que, nos últimos anos, a teoria sobre o gênero do discurso tem contribuído imensamente para a compreensão de como o discurso é usado no ambiente acadêmico, profissional e em outros contextos institucionais. No que se refere à análise de gêneros, optou-se nesta pesquisa por ressaltar mais especificamente os estudos e os aspectos teóricos apresentados por Swales, pelo destaque que o autor dedica ao estudo dos gêneros acadêmicos, em especial, ao artigo de periódico e à interação com a Ciência da Informação, evidenciada em alguns de seus estudos sobre *abstracts* (SWALES & FEAK, 2009) e análise de citações (SWALES, 1986). Outros estudos relevantes sobre análise de gêneros devem ser lembrados como Bakhtin (2000), Bathia (2004), Bazerman (2006) e Hyland (2004; 2009).

Ao prefaciar o estudo de Swales (1990), Long & Richard salientam que a análise de gêneros enriquece a nossa compreensão de como a linguagem é usada no ambiente de um contexto discursivo. Nesse sentido, John Swales (1990) desenvolve uma abordagem para a compreensão do discurso acadêmico, fundamentando-se na Sociolinguística, com o fito de esclarecer a natureza do uso da linguagem em um cenário acadêmico e educacional.

Ao refletir sobre os gêneros discursivos, Bakhtin (2000) ressalta que o uso da língua se processa por enunciados orais e escritos, produzidos pelos integrantes de uma determinada esfera de atividade humana. O enunciado reflete suas condições específicas e seus objetivos, pelo conteúdo, pelo seu estilo verbal, pelos recursos lexicais, fraseológicos e gramaticais e, especialmente, pela construção composicional. “Cada esfera de utilização da língua elabora seus tipos relativamente estáveis de enunciados, sendo isso o que denominamos gêneros do discurso” (BAKHTIN, 2000, p. 279).

Bakhtin (1979, p. 14 e 147) ressalta que “a língua é o reflexo das [...] relações sociais estáveis dos falantes. [...] conforme o contexto presente tal ou qual objetivo específico, vê-se dominar ora uma forma ora outra, ora uma, ora outra variante de texto”. Para ele, se não houvessem os gêneros discursivos e “se não os dominássemos, se tivéssemos de criá-los pela primeira vez

no processo de fala, se tivéssemos de construir cada um de nossos enunciados, a comunicação verbal seria quase impossível”.

Ao discutir o conceito de comunidades de discurso, Swales (1990) reporta-se a redes sociais e retóricas que se formam a partir de um conjunto de objetivos comuns. O autor ressalta que uma das características que os membros das comunidades discursivas possuem é a familiaridade com os gêneros particulares de discurso (SWALES, 1990).

A noção de comunidade de discurso, portanto, fundamenta-se mais no que as pessoas fazem do que quem elas são em uma determinada comunidade de discurso. Ao destacar a importância do conceito de comunidade de discurso para a análise de gêneros, Swales (1990) defende que as comunidades de discurso detêm convenções próprias e tradições para as quais diversas atividades verbais convergem como encontros rápidos que produzem relatórios e publicação de atividades. As classes recorrentes de eventos comunicativos são os gêneros que regem a vida verbal de uma comunidade de discurso. Swales destaca alguns princípios indispensáveis à criação de uma comunidade de discurso:

- (1) delinear objetivos comuns;
- (2) desenvolver procedimentos de intercomunicação entre seus membros;
- (3) usar mecanismos participativos para a troca de informações;
- (4) recorrer a determinados gêneros para atender a objetivos próprios;
- (5) compartilhar um léxico específico;
- (6) ter constantemente membros com nível adequado de conhecimento relevante e *expertise* discursiva.

A noção de comunidade discursiva diz respeito àqueles que trabalham usual ou profissionalmente com determinado(s) gênero(s) e que, desse modo, possuem maior conhecimento de suas convenções. As comunidades de discurso, assim como grupos acadêmicos de vários tipos, são reconhecidas pelos gêneros específicos que empregam, que incluem tanto os tipos de eventos orais quanto os tipos de textos escritos. O trabalho em que os membros de uma comunidade de discurso estão engajados envolve o processamento de tarefas que refletem habilidades retóricas e discursivas específicas, utilizadas por determinada comunidade (SWALES, 1990).

Segundo Bakhtin (1979), o termo gênero, no âmbito da linguística textual, refere-se a tipos relativamente estáveis de enunciados, produzidos em cada esfera de comunicação (lugar social dos interlocutores) que são caracterizados pelo conteúdo temático, estilo e tipo de composição, adotados em função de uma situação determinada por alguns parâmetros (finalidade, destinatários, conteúdo). Logo, os gêneros textuais são tipos próprios de textos, usados para organizar a linguagem no processo de comunicação de uma determinada comunidade de discurso.

A noção de gênero textual apresentada por Swales (1990) é fortemente baseada nos conceitos de propósito comunicativo e comunidade de discurso. Para Swales, a expressão análise de gêneros, anteriormente associada a gênero literário, refere-se à análise de qualquer tipo particular de discurso, falado ou escrito, e a diferentes tipos de comunidade. Cada comunidade de discurso define os gêneros mais relevantes para atingir seus objetivos particulares. O autor defende que os gêneros são entidades dinâmicas, passíveis de transformações de acordo com os contextos sociais e históricos em que são produzidos. Swales (1990) ressalta que um gênero engloba uma classe de eventos comunicativos, cujos participantes compartilham algum conjunto de objetivos comuns. Esses objetivos são reconhecidos pelos membros especialistas da comunidade discursiva, constituindo-se em um conjunto de razões para o gênero, que moldam a estrutura esquemática do discurso e influenciam a escolha de conteúdo e de estilo. Na opinião do autor, o principal critério para a definição de gênero é o objetivo comunicativo compartilhado pelos membros da comunidade.

Bazerman (2006a) entende que os atos de fala são realizados através de formas textuais padronizadas, ou gêneros, que se relacionam com outros gêneros textuais, produzidos em circunstâncias relacionadas. Juntos, os diferentes gêneros textuais se acomodam em conjuntos de gêneros, que fazem parte dos sistemas de atividades humanas. O autor alerta que, normalmente, os gêneros têm características de fácil reconhecimento, que revelam o tipo de texto e frequentemente estão intimamente relacionados com as principais atividades realizadas pelo gênero. Ele acrescenta que a questão central para o estudo de gêneros requer o entendimento de como as práticas e conhecimentos surgem e são aprendidas por um determinado grupo social. Bazerman apresenta métodos analíticos que indicam como a produção, a circulação e o uso ordenados dos textos constituem, parcialmente, a própria atividade e organização de grupos sociais. No que se refere à análise de gêneros textuais acadêmicos, são destacados neste estudo sobretudo as considerações de Bazerman (2006), de

Swales (1990; 2004) , de Bakhtin (2000) e de Hyland (2004; 2009).

Segundo Swales (1990), os gêneros textuais acadêmicos apresentam alguns aspectos que caracterizam a sua escrita, que deve ser:

- explícita,
- lexicalmente densa,
- impessoal,
- nominalizada,
- cautelosa,
- intertextual,
- referenciada.

Hyland (2009, p.24), em sua abordagem com relação ao discurso acadêmico, afirma que o discurso formal escrito tende:

- (1) a ser mais denso lexicalmente, apresentando maior índice de palavras com conteúdo semântico, porém, simples gramaticalmente;
- (2) a apresentar um grande número de nominalizações, representando eventos preferencialmente por nomes em substituição a verbos;
- (3) a ser mais explícito, com clara sinalização de relações semânticas.

Eggins (2004, p. 93) postula que o discurso formal escrito apresenta formas lexicais de prestígio, alta densidade lexical e gramática simplificada. A autora ressalta que o uso da nominalização possui duas principais vantagens textuais: organização retórica e intensificação da densidade lexical. No que se refere à organização retórica, no discurso escrito, através do uso de nominalizações, é possível reorganizar frases, diferenciando-o da fala (EGGINS, 2004, p.96). A nominalização também possibilita o empacotamento ou a reunião de conteúdos lexicais por sentença; ou seja, pela transformação dos verbos em nomes, aumenta-se o conteúdo possível do texto e assim sua densidade lexical é intensificada.

A densidade lexical de um texto, observa Eggins (2004, p. 97), pode ser estimada através do número de palavras que expressam conteúdos em função do número total de palavras de um



texto/sentença. O termo palavras com conteúdo refere-se a nomes, parte principal do verbo, advérbios e adjetivos. Em contraposição, o termo palavras sem conteúdo refere-se a preposições, conjunções, verbos auxiliares e pronomes.

A seguir, apresentam-se a análise histórica do gênero artigo de periódico e uma análise sincrônica dos gêneros de processo de pesquisa, que tem no artigo de periódico seu foco principal. Para tal, são mencionados os principais aspectos teóricos apresentados por John M. Swales (1990) e Bazerman (2006a). Em meio a variadas correntes para a análise de gêneros, enfatiza-se a perspectiva anglo-americana dos autores em referência, reconhecidamente relevante para a análise do gênero artigo acadêmico.

### 3.3 BREVE HISTÓRICO DO ARTIGO DE PERIÓDICO

A história do artigo de periódico contribui para pensarmos no artigo de periódico como gênero textual. Através dela, entendemos a diferença entre os primeiros *Transactions* publicados em 1665 e os artigos publicados em periódicos científicos de hoje. Ela nos esclarece sobre a evolução pelas quais os gêneros passam, para se adaptar às necessidades de uma comunidade de discurso, no processo de comunicação entre os pares.

Swales (1990) ressalta que o artigo científico emergiu, ainda que embrionariamente, com o estabelecimento do primeiro periódico científico intitulado *The Philosophical Transactions of the Royal Society of London*, em 1665. De acordo com Ard, relata Swales (1990), o gênero textual artigo científico se desenvolveu a partir das cartas informativas que os cientistas escreviam para seus pares. Nessa etapa, entretanto, grande parte das primeiras contribuições dos autores para os *Transactions* consistia em narrativas na primeira pessoa do singular, associada ao tratamento de senhor (Sir), no início das cartas.

Bazerman (2006b) menciona que as cartas que circulavam entre filósofos naturais, com o propósito de trocar informações sobre suas pesquisas, concorreram para a criação de gêneros textuais distintos. Segundo ele, o primeiro artigo científico originou-se da correspondência trocada entre o alemão Henry Oldenburg e outros pesquisadores, em meados do século XVII. Como resultado dessa comunicação, em 1662, ele se tornou o secretário da recém-criada *Royal Society of London*. Nessa função, sua correspondência foi intensificada e utilizada para a edição de textos no *Philosophical Transactions of the Royal Society of London* que, junto

com o *Journal des Sçavants*, editado na França, são considerados os primeiros periódicos científicos que surgiram publicados no ano de 1665.

Os primeiros números do *Philosophical Transactions of the Royal Society of London* foram editados sob a forma de resumo dessa correspondência e das reuniões da *Royal Society of London*. Na medida em que os *Transactions* e os subsequentes periódicos científicos começaram a assumir o papel de prover uma arena regular para as discussões científicas, emergiu uma nova e recorrente situação retórica, que conduz à criação de um novo gênero textual, distinguindo-se gradualmente das cartas escritas inicialmente pelos pesquisadores. Ziman (1969) e Kuhn (2007) concordam que a colaboração entre cientistas de um mesmo domínio do conhecimento é uma marca única da Ciência. Price (1963) e Crane (1988) reconhecem o fenômeno como a formação de um *Colégio Invisível*, uma rede de comunicação entre cientistas que trabalham numa frente de pesquisa de um domínio específico do conhecimento.

Na opinião de Shapin (1984 apud SWALES, 1990), um outro fato marcante associado ao aparecimento dos primeiros artigos científicos, deve-se aos tratados científicos, publicados após as pesquisas realizadas por Roberto Boyle e seus pesquisadores bolsistas, na década de 50 do século XVII, que precedeu o aparecimento da primeira edição dos *Transactions*, com a finalidade de criar uma fundação própria para o conhecimento científico. Segundo Shapin (1984 apud SWALES, 1990), Boyle e seus colegas pretendiam, então, transformar suas pesquisas em conhecimentos aceitos universalmente. As contribuições de Boyle foram de grande relevância para a retórica científica.

Para Swales (1990), ao produzir relatórios de pesquisa, motivado pela preocupação com a qualidade da leitura por parte de um pesquisador, Boyle acrescentava informações sobre estudos prévios que não obtiveram resultados que atendessem ao objetivo delineado inicialmente. Boyle e seus pares preocupavam-se em relatar suas pesquisas retoricamente.

Podemos dizer que o que ocorreu com o artigo científico está também resumidamente descrito no estudo de Bazerman (2006b) sobre a carta e a base social de gêneros diferenciados, no qual é mencionado o desenvolvimento dos *Transactions*, desde sua criação até o surgimento das revistas populares do século XIX. Bazerman (1988) também relata as transformações que sofreram as representações de experimentos científicos, no período de 1665 a 1800. À medida

que os fenômenos começam a ser tratados como uma problemática maior, os artigos começam a adquirir uma organização distinta, iniciando-se com a introdução ao fenômeno problematizado. Com a questão estabelecida, o artigo passa a descrever uma série de experimentos sobre o assunto, cujo objetivo é desvendar o problema estabelecido referente a uma questão de pesquisa. Os *Transactions*, publicados entre cada dois experimentos, apresentavam conclusões para os experimentos leves e apontavam para uma resposta racional ou para uma pesquisa subsequente. Dentro de uma continuidade, desenvolvida intensivamente, verifica-se que o pesquisador obtinha gradualmente uma compreensão do fenômeno estudado.

Também Cintra *et al.* (2002) reafirmam que os primeiros periódicos científicos foram publicados em 1665: o *Philosophical Transactions of the Royal Society of London*, na Inglaterra, e o *Journal des Sçavants*, na França. Segundo eles, na segunda metade do século XVIII, 1760, as publicações dos *Transactions* praticamente duplicavam a cada quinze anos. Os autores acrescentam que não se sabe quantos desapareceram e, hoje, é praticamente impossível dizer o número de periódicos existentes ao todo.

Swales (1990) dedica a maior parte de sua análise ao artigo de periódico. Esta opção, segundo o autor, pode ser justificada pelas seguintes razões:

(1) em muitas comunidades de discurso dedicadas à pesquisa, o artigo de periódico é considerado o gênero principal, tanto quantitativa quanto qualitativamente;

(2) sem dúvida, sabemos muito mais a respeito do artigo de periódico do que sobre outro gênero textual para publicação de pesquisa científica e tecnológica.

Um dos objetivos de Swales (1990) é o de demonstrar “como” o processo atual de análise de gêneros deve ser conduzido, o que tem sido possível evidenciar pela seleção e análise de um pequeno número de gêneros textuais, dentre os quais se destaca o artigo de periódico.

Uma razão final para a evidência atribuída ao artigo de pesquisa deve-se à sua interação dinâmica com todos os gêneros textuais de publicações que descrevem processos de pesquisas. Com exceção das atividades comunicativas semiprivadas e privadas (e.g. cartas pessoais, relatórios críticos de avaliadores, questões demandadas para reedição de publicações

etc.), o artigo de periódico permanece no centro de uma “teia” formada por gêneros textuais de publicações de processos de pesquisa. As interrelações são bem conhecidas e reconhecidas no ambiente acadêmico, de acordo com Swales (1990). Segue a figura 1 com a representação gráfica das interrelações dos gêneros textuais de publicações de processos de pesquisa, estabelecidas por Swales (1990).

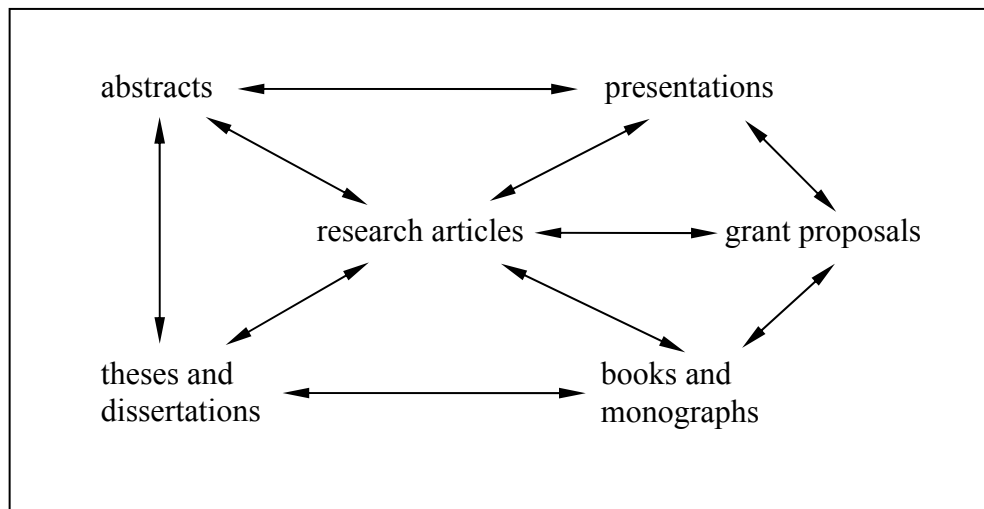


Figura 1 O artigo de periódico e outros gêneros de processo de pesquisa.  
Fonte: Swales (1990, p. 177)

O artigo científico, na opinião de Ziman (1969), é o mecanismo que tem como objetivo a publicação de fragmentos de trabalhos científicos. Em Swales (1990), infere-se que, geralmente, o artigo de pesquisa é prefaciado por um resumo desenvolvido em um só tópico, solicitado pelo corpo editorial do periódico. Após a publicação em artigos, um resumo pode aparecer em um periódico ou em uma base eletrônica de dados de resumos. Em resumos, evitam-se abreviações, jargões, símbolos e outros tipos de “linguagem abreviada”. O resumo é redigido em sentenças rapidamente enunciadas, evitando-se repetições, expressões insignificantes, superlativos, adjetivos, ilustrações, detalhes descritivos, exemplos, notas de rodapé. Enfim, eliminam-se redundâncias, o que pode ser ancorado na compactação da informação e no uso de anáforas (SWALES, 1990).

A apresentação em conferências pode relatar uma pesquisa em desenvolvimento ou oferecer uma prova (*trial*) de idéias novas. Ela pode ser uma versão de um artigo de pesquisa não publicado ainda, ou pode ser baseada em um artigo previamente publicado. Uma apresentação pode também ser preparada visando a sua publicação em um *Proceedings*. Em relação a essa interação entre apresentações em eventos científicos e artigos de pesquisa, admite-se que o

esforço envolvido no desenvolvimento de uma participação em evento acadêmico aumenta a probabilidade do apresentador alcançar um *aceite* para um artigo de pesquisa, em periódicos científicos. Da mesma forma, a simbiose entre artigo de periódico e auxílio para pesquisa é um fenômeno reconhecido na literatura, mencionado como o Efeito Matthew na Ciência, que se refere aos processos psicossociais, que afetam o sistema de avaliação e distribuição de recompensas científicas. Enfim, artigos de pesquisa publicados aumentam as chances de serem seguidos de auxílio para pesquisa, e o auxílio para pesquisa, por sua vez, aumenta a chance de publicação de artigos de pesquisa. Na medida em que as propostas para auxílio de pesquisa objetivam a obtenção de financiamento e o desenvolvimento de projetos de pesquisa, reconhece-se a importância do gênero em termos tangíveis (SWALES, 1990).

As teses e dissertações são consideradas um rito de passagem ou uma qualificação necessária para capacitar o titular a galgar novos patamares. A redação de dissertações e teses utiliza um meta discurso ou uma redação mais envolvente em substituição a referências ao assunto. Autores de dissertações jamais aconselham seus leitores, visto que sua audiência é composta por um grupo de especialistas no assunto que atuam como conselheiros e julgam o produto finalizado em forma de tese ou dissertação. A solicitação de reimpressões consiste em uma requisição de uma cópia ou separata de um artigo de pesquisa, enviada ao autor da publicação (SWALES, 1990).

Dentre os gêneros textuais de publicações de processos de pesquisa, apresentado pelo autor, o artigo científico é o gênero escolhido para a pesquisa desta tese. Observa-se que o artigo de periódico ocupa posição central privilegiada, tanto quantitativamente quanto qualitativamente. Ou seja, o gênero textual artigo de periódicos situa-se na posição mais elevada do *ranking* dos gêneros textuais acadêmicos.

### 3.4 CIÊNCIA DA INFORMAÇÃO E ANÁLISE DE DOMÍNIO

Swales (1986), em seu artigo intitulado *Citation analysis and discourse analysis*, ressalta os aspectos interdisciplinares, cada vez mais frequentes nos estudos em Ciência da Informação, sobre a análise de citações, que utiliza indicadores bibliométricos e a análise de discurso. Ao discorrer sobre a Análise de Domínio, do ponto de vista da área de Organização do Conhecimento em Ciência da Informação, Hjørland (2004) menciona que a Análise de Domínio oferece uma perspectiva teórica e prática de grande relevância para o

desenvolvimento do arcabouço teórico e de subsídios práticos para a Ciência da Informação e salienta, nesse campo de pesquisa, a análise de discurso, a indexação, a análise de gêneros e a Bibliometria. Na opinião de Hjørland, os estudos teóricos e a prática da Análise de Domínio revelam uma visão coerente de todos os maiores conceitos da área de Ciência da Informação.

Nesse contexto, Hjørland (2004) menciona que a Análise de Domínio é capaz de reunir, em um só campo, diferentes disciplinas, tais como Organização do Conhecimento, Recuperação da Informação e Competência em Informação. Hjørland (1995) considera que o horizonte mais profícuo para a Ciência da Informação é o estudo dos domínios do conhecimento ou comunidades de discursos e pensamentos. Ele acrescenta que a análise de domínio permite mapear os diferentes atores, instituições e processos de comunicação de uma área do conhecimento, assim como possibilita a investigação de como os domínios do conhecimento se distinguem e se assemelham a partir dos diferentes pontos de vista. Em artigo anterior, Hjørland (2002), considerado o precursor da Análise do Domínio na Ciência da Informação, indica onze abordagens de análise dos domínios discursivos que, no conjunto, representam as competências específicas dos cientistas da informação:

- (1) produção e avaliação de guias da literatura e assuntos;
- (2) produção de classificações especiais e de *thesaurus*;
- (3) pesquisas sobre indexação e recuperação da informação em domínios especializados;
- (4) estudos empíricos de usuários em diferentes áreas de assunto;
- (5) produção e interpretação de estudos bibliométricos;
- (6) estudos dos gêneros de documentos em domínios do conhecimento;
- (7) estudos críticos e epistemológicos de diferentes paradigmas, hipóteses e interesses em domínios do conhecimento;
- (8) estudos terminológicos, linguagens para fins especiais e análise do discurso em áreas específicas do conhecimento;
- (9) estudo de estruturas e instituições na comunicação científica em um domínio do conhecimento;
- (10) análise de domínio, pela investigação do processo de cognição profissional e Inteligência Artificial;
- (11) estudos históricos de serviços e estruturas de informação nos domínios do conhecimento.

Hjørland (1995) compara a Análise de Domínio com outras abordagens importantes em Ciência da Informação, tais como a cognitiva. Ao mencionar questões relevantes de pesquisa, ele destaca que a diferença entre os domínios do conhecimento afeta o valor informativo dos pontos de acesso (palavras-chave) em bases eletrônica de dados. O autor ressalta ainda a relevância dos estudos sociolinguísticos para as pesquisas sobre o significado textual. Nesse sentido, destaca as contribuições de autores como Bakhtin e Halliday, entre outros.

Em síntese, os cientistas da informação podem investigar um domínio a partir de um conhecimento geral de base de dados e índices de citação e, por meio desses recursos, explorar as suas utilidades em um domínio específico. Em outras palavras, os especialistas em informação abordam um domínio a partir do modelo *top-down* (de cima para baixo), enquanto que os especialistas em análise de domínio abordam problemas de informação, pensando a organização do conhecimento a partir de modelos *botton-up* (de baixo para cima).

Na interface Linguística e Indexação, são analisados *corpora* textuais, utilizando-se conhecimentos linguísticos e estatísticos para a construção de referenciais, com o objetivo de contribuir para o desenvolvimento de sistemas de indexação automática. Como exemplos de trabalhos interdisciplinares dessa natureza, podemos citar Basílio (1979a ; 1979b), bem como Mauceri (1995) e Polanco et al.(1995), apresentados na *Journées Organisées par la Société Francaise de Bibliométrie Appliquée (SFBA)*, no período de 30 de maio a 02 de junho de 1995.

Nesta tese, foi aplicada: (1) uma análise estrutural do domínio e da adequação de indicadores bibliométricos, Leis de Zipf e Ponto de Transição (T) de Goffman, para a delimitação da região de concentração de palavras-chave; (2) um estudo sobre o fenômeno léxico-morfológico de nominalização deverbal presente produtivamente na terminologia utilizada em artigos científicos e tecnológicos na área de indústria de vinhos.

Na Bibliometria, especialmente no que se refere à aplicação das Leis de Zipf e Ponto de Transição de Goffman à indexação temática da informação, os pressupostos teóricos apontam para uma ligação entre a frequência de ocorrência de uma determinada palavra e o seu potencial de representação temática em determinado *corpus* em análise.

Hrebicek (2002), em seu artigo intitulado *Zipf's Law and Text*, alerta que, do ponto de vista estatístico, nenhuma objeção razoável, manifestando-se contra a redução de palavras e de *corpora*, é feita às Leis de Zipf. Do ponto de vista linguístico, entretanto, existe um fato marcante: o *gap* existente entre unidades lexicais e *corpora* indica uma ausência dificilmente admissível de fenômenos relevantes ligados às estruturas lexicais, atualmente ocorrendo em textos. Qualquer unidade lexical de uma linguagem natural tem uma abstração intuitiva de seu significado deduzido a partir de seu uso em textos. As propriedades semânticas reais das palavras podem ser mais adequadamente observadas quando são testadas como unidades lexicais em textos individuais. Nesse sentido, considerando-se as estruturas de textos individuais, Hrebicek propõe, entre outras, as seguintes questões:

(1) a abordagem mencionada está baseada em *corpora* capazes de expressar a real estrutura lexical de uma linguagem?

(2) existem abordagens teóricas, de alguma maneira associadas às Leis de Zipf, suficientemente fortes para suportar o impacto causado pela mudança de análise de *corpora* para textos?

Bybee & Hopper (2001, p.1) ao discutirem a relação da frequência com a emergência de estruturas linguística mencionam que o trabalho de George K. Zipf<sup>3</sup> foi o pioneiro nos anos 30. A obra *"The Psycho-biology of Language"* editada em 1935 se baseia no *Princípio do Menor Esforço*. Zipf antecipou muitos temas de investigação explorados recentemente sobre a relação entre a frequência e a estruturas linguísticas. Zipf postula que os itens mais frequentes são os itens mais curtos. O autor adota ainda o termo *"Dynamic Philology"* para nomear os estudos quantitativos da mudança linguística. Nos anos 80, outros linguistas procuraram respostas nos usos reais no âmbito do discurso. A propósito os tesouros são introduzidos nas linguagens documentárias valendo-se da noção de domínio e da frequência de uso dos itens lexicais.

---

<sup>3</sup> Chomsky & Miller (1971, p. 26), na construção do modelo gerativo, reportam-se à George K. Zipf observando:

Les problèmes relatifs à la définition du degré de grammaticalité, à l'interprétation des énoncés déviants, et à la construction de grammaires qui prévoient d'autres degrés de grammaticalité que 0 ou 1, sont tous intéressants et importants. Divers aspects de ces questions sont étudiés dans les travaux de Chomsky (1955, 1961b), Zipf (1960a, 1960b, 1961) et Katz (1963).



## 4 METODOLOGIA

A linha teórico-metodológica baseia-se na Teoria Lexical e na Bibliometria. Com essa orientação, são abordados conceitos linguísticos relevantes para o procedimento de pesquisa, assim como as leis e princípios bibliométricos utilizados na indexação da informação científica e tecnológica.

Em relação à amostra, inicialmente, são mencionados os critérios que nortearam a seleção dos seis artigos de periódicos. Em seguida, aborda-se brevemente a produção de vinhos e são referenciados os artigos selecionados. No item 4.2, descrevem-se passo a passo os procedimentos de análise dos artigos, do ponto de vista tanto da produtividade como da frequência de uso, bem como o mecanismo de delimitação da Região de Transição de Goffman e de reconhecimento dos processos léxico-morfológicos de nominalização deverbal. Considera-se ainda a regularidade de algumas nominalizações deverbais relevantes, selecionadas dos textos dos artigos em análise. O item 4.3 versa sobre a Bibliometria, destacando-se as Leis de Zipf e o Ponto de Transição de Goffman como mecanismos de delimitação da região de concentração dos termos de indexação.

### 4.1 AMOSTRA

Optou-se por um *corpus*, composto por quatro artigos de periódicos sobre a indústria de vinhos, em língua portuguesa, analisados à luz do referencial teórico e conceitual, sintetizado anteriormente. Visando à confrontação dos dados, foram também selecionados dois artigos sobre Economia. A seleção dos textos segue os seguintes critérios de (a) a (f) e constitui a amostra da pesquisa.

- (a) quatro artigos de periódicos científicos e tecnológicos sobre a indústria de vinhos e dois sobre Economia;
- (b) artigos de periódicos científicos e tecnológicos indexados na Base de Dados da *Scielo Brazil Library Online* ou em bases eletrônicas de dados de acesso livre na *web*;
- (c) artigos científicos e tecnológicos com textos integrais disponíveis na *web*;
- (d) artigos científicos e tecnológicos publicados em língua portuguesa;
- (e) artigos científicos e tecnológicos publicados no período de 1998 a 2008;
- (f) artigos científicos e tecnológicos com um total de 2.000 a 6.000 palavras.

As referências dos seis artigos de periódico selecionados para a análise dos dados estão elencadas no Quadro 1, a seguir.

RIZZON, Luiz Antenor; ZANUZ, Mauro C.; MIELE, Alberto. Evolução da acidez durante a vinificação de uvas tintas de três regiões vitícolas do Rio Grande do Sul. **Ciência e Tecnologia de Alimentos**, Campinas, v. 18, n. 2, p. 179-183, Maio/Jul.1998.

Disponível em: [http://www.scielo.br/scielo.php?pid=S0101-20611998000200007&script=sci\\_arttext](http://www.scielo.br/scielo.php?pid=S0101-20611998000200007&script=sci_arttext)  
Acesso em: 27/05/2009

RIZZON, Luiz Antenor; MIELE, Alberto. Avaliação da cv. Merlot para elaboração de vinho tinto. **Ciência e Tecnologia de Alimentos**, Campinas, v. 23, supl., p. 156-161, Dez. 2003.

Disponível em: <http://www.scielo.br/pdf/cta/v23s0/19489.pdf>  
Acesso em: 27/05/2009

RIZZON, Luiz Antenor; MIELE, Alberto. Avaliação da cv. Cabernet Sauvignon para elaboração de vinho tinto. **Ciência e Tecnologia de Alimentos**, Campinas, v. 22, n. 2, p. 192-198, maio/ago. 2002.

Disponível em: <http://www.scielo.br/pdf/cta/v22n2/a15v22n2.pdf>  
Acesso em: 27/05/2009

MAMEDE, Maria Eugênia de Oliveira; PASTORE, Gláucia Maria. Avaliação da produção dos compostos majoritários da fermentação de mosto da uva por leveduras isoladas da região da “Serra Gaúcha” (RS). **Ciência e Tecnologia de Alimentos**, Campinas, v. 24, n. 3, p. 453-458, jul./set. 2004.

Disponível em: <http://www.scielo.br/pdf/cta/v24n3/21942.pdf>  
Acesso em: 27/05/2009.

CRISOSTOMO, Vicente Lima. Dificuldades das empresas brasileiras para financiar seus investimentos em capital físico e em inovação. *Revista Economia Contemporânea*, Rio de Janeiro, v. 13, n. 2, Aug. 2009

Disponível em: <http://www.scielo.br/pdf/rec/v13n2/v13na04.pdf>  
Acesso em: 27/01/2010.

VASCONCELOS, Cláudio Roberto Fóffano; VASCONCELOS, Silvinha Pinto. Medidas "antidumping" e resultados colusivos: o caso do PEBDL na economia brasileira. **Nova Economia**, Belo Horizonte, v. 15, n. 3, p. 117-141, Set./Dec. 2005.

Disponível em: <http://www.scielo.br/pdf/neco/v15n3/v15n3a05.pdf>  
Acesso em: 27/01/2010

Quadro 1 Amostra

## 4.2 A PRODUÇÃO DO VINHO

As pesquisas arqueológicas, segundo Johnson (1999), consideram o acúmulo de caroços de uva como uma indicação de que se produziu vinho, em um passado longínquo. Os mais antigos caroços de uva cultivada foram achados na Geórgia soviética e referem-se ao período de 700 a 5000 a.C. Entretanto, os egípcios foram os primeiros a documentar, em pinturas, os detalhes da viticultura praticada por seu povo. As atividades de produção de vinho reproduzidas com clareza pelos egípcios revelam pleno domínio da tecnologia vinícola e ocorreram há cerca de 3000 a 5000 anos. Nessa época, os egípcios especialistas diferenciavam

qualidades de vinho com a precisão e o profissionalismo de exportadores ou corretores de vinhos do século XX.

Macnell (2001, p. 30) também infere que o vinho existe entre nós há mais de 5000 anos. No entanto, o processo natural e complexo, pelo qual é produzido atualmente - fermentação - só foi compreendido há pouco mais de 150 anos. Somente em meados do século XIX, quando a pesquisa de Louis Pasteur em Microbiologia ligou a conversão do açúcar em álcool (fermentação) à existência de organismos vivos chamados de levedura, é que a produção de vinho migrou do domínio do ocultismo para o domínio científico. Mais de um século se passou para que o próximo avanço significativo ocorresse na área de produção de vinho. Até a Segunda Guerra Mundial, com exceção dos vinhos fortificados (exemplo: Vinho do Porto) e dos espumantes (exemplo: *Champagne*), a maior parte dos vinhos era produzida de acordo com dois processos clássicos: (1) produção de vinho branco e (2) produção de vinho tinto. Os dois processos de produção de vinho eram especiais, complexos e particulares. Por volta de 1960, em nível internacional, os avanços na produção de vinho e o surgimento de tecnologias e equipamentos mais sofisticados, especialmente tanques em aço inoxidável com controle de temperatura, permitiram aos produtores maior capacidade para produzir aromas, sabores, texturas e durabilidade do vinho. Nascia um poderoso mundo ligado à produção de vinho cujo objetivo continuava imutável: proteger e estimular as características do vinho, provenientes de vinhedos. A cadeia de produção de vinho sempre inicia no vinhedo, com a escolha do local, a seleção das variedades de uvas e dos clones a plantar, o planejamento para a distribuição dos vinhedos e a maneira pela qual as videiras serão arrumadas em grades ou entrelaçadas e cuidadas.

Johnson (1999) esclarece que a videira pertence a uma espécie de trepadeira, que possui parentes por todo o hemisfério norte. Cerca de quarenta espécies são incluídas no gênero botânico *vitis*. Seu nome específico é *Vinífera*, “que produz vinho”. A *vitis Vinífera* possui um teor de açúcar que corresponde a cerca de um terço do seu volume e, ao mesmo tempo, contém elementos ácidos que permitem transformar seu suco em uma bebida límpida e revigorante: o vinho.

Em Macnell (2001), ressalta-se que, com raras exceções, o sumo de todas as uvas tinto e branca são incolores. A diferença entre vinho tinto e vinho branco é a seguinte: no vinho tinto, o sumo é fermentado com a casca de uvas tintas. As cascas acabam agindo como um pacote

de corantes. Durante a fermentação gera-se calor, assim como álcool. Ambos ajudam a extrair os pigmentos vermelho-púrpura das cascas, tingindo o vinho. No caso do vinho branco, as cascas são rapidamente separadas do sumo antes que este fermente. É o que também ocorre com vinhos brancos feitos de uvas tintas. Por exemplo, no caso do *Champagne*, o sumo é rapidamente separado das cascas tintas antes que ocorra qualquer coloração.

### 4.3 PROCEDIMENTO DE ANÁLISE

Neste item, são listados os procedimentos de análise e os critérios que nortearam o processamento dos artigos selecionados, que podem ser resumidos nas seguintes etapas:

(a) delimitação da amostra através da seleção de quatro artigos científicos e tecnológicos sobre a indústria de vinhos e dois sobre Economia;

(b) contagem das palavras por meio da utilização do *software* contador de palavras *Rank Words 2.0*, disponível em [http://download.cnet.com/Rank-Words3000-2279\\_4-10909564.html](http://download.cnet.com/Rank-Words3000-2279_4-10909564.html), considerando-se os seguintes critérios: (1) excluir títulos dos artigos, nomes, afiliações e títulos dos autores, resumos, palavras-chave, sumários, referências, notas de rodapé, bibliografias, agradecimentos, tabelas, gráficos, ilustrações, diagramas, equações e pontuações; (2) considerar citações como partes dos textos e formas flexionadas das palavras como palavras distintas;

(c) listagem e ordenação das palavras operando o *software* mencionado em (b) que produziu uma listagem em 3 colunas assim distribuídas: as palavras, as frequências em ordem decrescente de ocorrência e o *rank* das palavras;

(d) resolução de cálculos matemáticos e estatísticos a partir do transporte dos resultados em planilhas utilizando-se o *Software Microsoft Office Excel 2003*, para a projeção dos seguintes produtos:  $r \times f = C$ , total de palavras, total de palavras distintas e *ranking* das nominalizações deverbais;

(e) identificação dos seguintes padrões de nominalizações deverbais regulares: (1)  $[X] v \rightarrow [ [X] v -\text{ção} ] N$ , (2)  $[X] v \rightarrow [ [X] v -\text{mento} ] N$ , (3)  $[X] v \rightarrow [ [X] v -\text{ncia} ] N$ , (4)  $[X] v \rightarrow [ [X] v -\text{agem} ] N$ , (5)  $[X] v \rightarrow [ [X] v -\text{da} ] N$ ;

(f) estimativa dos graus relativos de produtividade das nominalizações deverbais, pela consideração da frequência relativa a cada nominalização e sua relação com o somatório das frequências das bases e dos sufixos de formação em análise;

(g) aplicação da Primeira e da Segunda Leis de Zipf ;

(h) identificação do Ponto de Transição (T) de Goffman, com o objetivo de delimitar a Região de Transição que, segundo Goffman, reúne os itens lexicais de alto conteúdo semântico, com probabilidade de ser usados na indexação;

(i) investigação das nominalizações deverbais que se situam na região de concentração de palavras de alto conteúdo semântico;

(j) confrontação dos dados observados nos textos sobre a indústria de vinhos com os observados nos textos sobre Economia.

#### 4.4 BIBLIOMETRIA: LEIS DE ZIPF E PONTO DE TRANSIÇÃO DE GOFFMAN

A área de Bibliometria, na Ciência da Informação, é constituída por um conjunto de leis e princípios empíricos, que contribuem para o estabelecimento da fundamentação teórica da Ciência da Informação. Miranda Pao (1989) refere-se à Bibliometria como um termo introduzido por Allan Pritchard, em seu artigo *Statistical Bibliography or Bibliometrics*, publicado em 1969, para denotar a área de assunto, que utiliza métodos matemáticos e estatísticos, com o objetivo de analisar os processos de comunicação escrita. Pritchard, segundo Pao (1989), percebeu que a literatura é o “ingrediente chave” no processo de comunicação científica e que os atributos de uma unidade de literatura podem ser analisados estatisticamente. Pao (1989) acrescenta que publicações, autores, palavras-chave, usuários, citações e periódicos são alguns dos parâmetros pesquisados em estudos bibliométricos da literatura.

A seguir, apresenta-se o Quadro 2, com as leis e princípios bibliométricos, os focos de estudo, principais aplicações e áreas de interesse, desenvolvido por Guedes & Borschiver (2005).

<b>Ciência da Informação</b>		
<b>Bibliometria (aplicada à Literatura e a Textos Científicos e Tecnológicos)</b>		
<b>Leis e Princípios</b>	<b>Focos de Estudo</b>	<b>Principais Aplicações</b>
Lei de Bradford	Periódicos	Estimativa do grau de relevância de periódicos em dada área do conhecimento
Lei de Lotka	Autores	Estimativa do grau de relevância de autores em dada área do conhecimento
Leis de Zipf	Palavras	Indexação automática de artigos científicos e tecnológicos
Ponto de Transição (T) de Goffman	Palavras	Indexação automática de artigos científicos e tecnológicos
Colégios Invisíveis	Citações	Identificação da elite de pesquisadores, em dada área do conhecimento
Fator de Imediatismo ou de Impacto	Citações	Estimativa do grau de relevância de artigos, cientistas e periódicos científicos, em determinada área do conhecimento
Acoplamento Bibliográfico	Citações	Estimativa do grau de ligação de dois ou mais artigos
Co-citação	Citações	Estimativa do grau de ligação de dois ou mais artigos
Obsolescência da Literatura	Citações	Estimativa do declínio da literatura de determinada área do conhecimento
Vida média	Citações	Estimativa da vida média de uma unidade da literatura de dada área do conhecimento
Teoria Epidêmica de Goffman	Citações	Estimativa da razão de crescimento e declínio de determinada área do conhecimento
Lei do Elitismo	Citações	Estimativa do tamanho da elite de determinada população de autores
Frente de Pesquisa	Citações	Identificação de um padrão de relação múltipla entre autores que se citam
Lei dos 80/20	Demanda de Informação	Composição, ampliação e redução de acervos

Quadro 2 Leis e princípios bibliométricos, focos de estudo, principais aplicações e áreas de interesse (GUEDES; BORSHIVER, 2005).

Segundo Macias-Chapula (1998), Tague-Sutcliffe (1992) estabelece a distinção entre a Bibliometria, a Cienciometria e a Informetria e as conceitua da seguinte forma:

Bibliometria é o estudo dos aspectos quantitativos da produção, disseminação e uso da informação registrada. Usada pela primeira vez por Pritchard em 1969, a Bibliometria desenvolve padrões e modelos matemáticos para medir esses processos, usando seus resultados para elaborar previsões e apoiar tomadas de decisão.

Cienciometria é o estudo dos aspectos quantitativos da ciência enquanto uma disciplina ou atividade econômica. A cienciometria é um segmento da sociologia da ciência, sendo aplicada no desenvolvimento de políticas científicas. Envolve estudos quantitativos das atividades científicas, incluindo a publicação e, portanto, sobrepondo-se à bibliometria.

Informetria é o estudo dos aspectos quantitativos da informação em qualquer formato, e não apenas registros catalográficos ou bibliografias, referente a qualquer grupo social, e não apenas aos cientistas. A informetria pode incorporar, utilizar e ampliar os muitos estudos de avaliação da informação que estão fora dos limites tanto da bibliometria como da cienciometria.

Braun e Schubert (2003) ressaltam que a Cientometria investiga os aspectos quantitativos ligados à produção, disseminação e uso de informações científicas, com a finalidade de contribuir para uma melhor compreensão do processo de pesquisa científica, enquanto atividade social.

Segue o diagrama representando graficamente a superposição da Bibliometria, Cientometria, Informetria e ainda a Webometria, “área emergente na Ciência da Informação que aplica técnicas bibliométricas e cienciométricas para medir a informação disponível na Web” (VANTI, 2002).

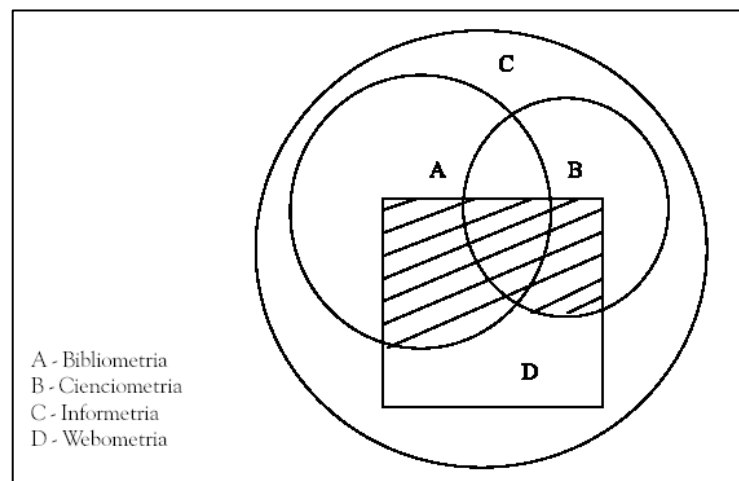


Figura 2 Diagrama da inter-relação entre os quatro subcampos.  
 Fonte: Vanti (2002, p. 161)

No estudo, utiliza-se a Bibliometria, no âmbito da Cientometria, com o objetivo de analisar os aspectos estatísticos da nominalização deverbal na linguagem científica e tecnológica, nos artigos que integrarão o *corpora* em análise.

Dentre as leis e princípios bibliométricos, são de interesse deste trabalho as Leis de Zipf e o Ponto de Transição de Goffman, fundamentados na frequência de ocorrência de palavras. Os estudos estatísticos, baseados na frequência de ocorrência de palavras de artigos científicos e tecnológicos, têm revelado grande potencial para automatização do processo de indexação temática. Eles partem da premissa básica de que, ao produzir um artigo científico e tecnológico, um autor tende a decidir por um determinado nível de especificidade de linguagem e comunicação de palavras, para transmitir sua mensagem (LUHN, 1957). Existe também, segundo Luhn (1957), a probabilidade de quanto mais frequentemente uma noção e combinação de noções ocorrerem, maior importância o autor atribuirá para expressar a essência de sua idéia no todo. Ele observa que a probabilidade de um autor empregar palavras diferentes, para expressar a mesma idéia, é pequena. Observa, ainda, que as palavras de frequência média são as de maior carga semântica, conseqüentemente são as mais adequadas para a indexação temática de um texto.

Hans Peter Luhn (1957) propôs que a indexação deveria ser baseada na própria literatura. Tópicos relevantes, termos e a relação entre eles, sobre um assunto dado, deveriam ser derivados da análise de uma amostra de documentos sobre aquele assunto. Vários estudos têm se baseado nos pressupostos lançados por Luhn (1957) e usado a frequência de ocorrência das palavras, como indicador de seu grau de relevância em relação a um determinado texto.

Na Bibliometria, as Leis de Zipf associadas à frequência de ocorrência das palavras e o Ponto de Transição de Goffman estão diretamente relacionados a tentativas de representação temática da informação, objetivando maior grau de precisão na indexação de artigos científicos e tecnológicos. Zipf (1949) observou que existe uma lei que comanda toda a atividade humana: a “Lei do Menor Esforço”. Ele pressupõe que o comportamento humano obedece ao Princípio do Menor Esforço. Como linguista, ele observa que, ao analisarmos um texto suficientemente longo e ordenarmos as palavras de acordo com sua frequência de ocorrência decrescente, o produto da ordem de série pela frequência é uma constante ( $C$ ) para cada texto analisado. A partir dessa observação, Zipf enuncia sua primeira lei, representada por  $r \times f = C$ . A Primeira Lei de Zipf estabelece que o produto da ordem de série ( $r$ ) de uma palavra, pela sua frequência ( $f$ ) é aproximadamente constante ( $C$ ). Enunciou assim que

$$(1) \quad r \times f = C$$



Estudos sobre a aplicabilidade dessa lei indicam que ela somente se aplica às palavras de alta frequência de ocorrência em um texto. Para palavras de baixa frequência de ocorrência (alta ordem de série), Zipf propôs uma segunda lei, revisada e modificada por Booth.

A Segunda Lei de Zipf enuncia que, em um texto, várias palavras de baixa frequência de ocorrência (alta ordem de série) têm a mesma frequência. Essa lei é também conhecida como Lei de Zipf-Booth, associada às palavras de baixa frequência de ocorrência. Booth (1967), ao modificá-la, a representa matematicamente pela seguinte fórmula:

$$(2) \quad \frac{I_1}{I_n} = \frac{n(n+1)}{2} \quad (3) \quad I_n = \frac{2I_1}{n(n+1)}$$

Onde  $I_1$  é o número total de palavras que têm frequência 1 e  $I_n$  é o número total de palavras que têm frequência  $n$ , 2 sendo a constante válida para a língua inglesa. Assim,  $I_n$  corresponde ao número total de palavras de qualquer frequência ( $n$ ) observada. Por exemplo, para se calcular o total de palavras que ocorrem três vezes ( $n = 3$ ), tendo por hipótese que o total de palavras com frequência 1 no *ranking* ( $I_1$ ) é 250, o cálculo é:

$$(4) \quad I_3 = \frac{2 \times 250}{3(3+1)} = \frac{500}{12}$$

Logo, segundo Zipf, o total de palavras que ocorrem 3 (três) vezes é igual a 41,666..., ou seja, aproximadamente 42.

Esses dois comportamentos distintos definem as duas extremidades da lista de distribuição de palavras de um texto. Assim, é razoável esperar uma região crítica, na qual ocorre a transição do comportamento das palavras de baixa frequência para as palavras de alta frequência. Goffman, segundo Pao (1978), admitiu como hipótese que, nessa região de transição, estariam as palavras de maior conteúdo semântico de um texto.

De acordo com Goffman, para chegar a essa região de transição, onde estariam as palavras de alto conteúdo semântico (termos de indexação ou palavras-chave), a expressão da Segunda Lei de Zipf teria que fornecer o comportamento típico das palavras de alta frequência, isto é, o número de palavras que têm frequência  $n$  tenderia a 1 (unidade). Substituindo-se, na expressão da Segunda Lei de Zipf-Booth,  $I_n$  por 1, obtém-se:

$$(5) \quad \frac{I_1}{1} = \frac{n(n+1)}{2}$$

Ou ainda, rearranjando

$$(6) \quad n^2 + n - 2I_1 = 0$$

cujas raízes são

$$(7) \quad n = \frac{-1 \pm \sqrt{1 + 8I_1}}{2}$$

Da expressão acima, interessa somente determinar a raiz positiva, assim

$$(8) \quad n = \frac{-1 + \sqrt{1 + 8I_1}}{2}$$

Ao valor de  $n$  assim determinado corresponde o Ponto de Transição (T) de Goffman.

Se considerarmos como exemplo  $I_1 = 250$ , teremos  $n$  igual a 21,866; isto é, o valor de  $n$  (frequência que corresponde ao Ponto T) é aproximadamente 22. O Ponto T de Goffman determina graficamente a localização onde ocorre a transição das palavras de baixa frequência para as de alta frequência. Existe uma determinada região, ao redor desse ponto, com probabilidade de concentrar as palavras de alto conteúdo semântico, aquelas que seriam usadas para indexação de um texto em questão. Goffman apresenta, com o Ponto de T, a primeira oportunidade de se decompor um texto sintaticamente, objetivando a sua indexação temática.

Pao (1978), ao testar a hipótese de que a Região de Transição de Goffman produziria as palavras de maior conteúdo semântico de um determinado texto, delimitou a região projetando para baixo o mesmo número de palavras que ocorreu acima da frequência que corresponde ao Ponto T. Nessa região, foi calculado o total de palavras e, desse total, excluídas as palavras sem conteúdo semântico, tais como: preposições, artigos, conjunções, pronomes e verbos auxiliares. Essas palavras foram excluídas dos dados na medida em que o objetivo do estudo era a investigação temática dos termos na Região de Transição de Goffman. Ao analisar os resultados alcançados, a autora conclui que eles foram satisfatórios, para os textos analisados, e parecem indicar que a seleção automática de termos de indexação,

a partir de uma lista de frequência de ocorrência de palavras, é uma perspectiva promissora para a indexação automática.

No caso específico de textos “suficientemente longos”, a aplicação das leis de Zipf, Zipf/Booth, Ponto T de Goffman e similares produz uma listagem de palavras acompanhadas de respectivas frequências de aparecimento no texto. Segundo Braga (1996):

Há uma total ‘descontração’ do contexto das palavras, transformando-as em seqüências de caracteres entre espaços e pontuações. Apesar disso, os poucos estudos existentes indicam que é possível determinar o conteúdo semântico dos textos, através de tais procedimentos. (BRAGA, 1996)

Rouault (1987), em sua discussão sobre métodos estatísticos de indexação temática automática, apesar de questionar a validade das leis de Zipf, destaca que podemos delimitar aproximadamente três zonas em qualquer índice e verifica que as fronteiras entre elas não são claras. Acrescenta, entretanto, que essas observações são válidas para todo o *corpus* analisado.

Segundo Rouault (1987), a primeira zona é constituída de formas com um número elevado de ocorrências. Ela contém, acima de tudo, palavras cuja presença deve-se a razões de sintaxe (estrutura linguística do idioma, no qual é escrito o *corpus*). No idioma francês, encontramos, essencialmente, determinantes, tais como: “le” (o), “la” (a); preposições: “des” (dos, das), representando cerca de 10% do *corpus*; conjunções, verbos auxiliares etc. Essa zona contém poucos representantes das categorias de substantivos, adjetivos, verbos. Já a segunda zona, na opinião de Rouault, caracteriza-se, sobretudo, pelo fato de conter uma quantidade maior que a primeira zona de representantes de categorias morfológicas “informativas”, tais como: substantivo, adjetivo, verbo, entretanto, as formas gramaticais que caracterizam a primeira zona ocorrem, ainda, sobretudo no topo dessa segunda zona. A terceira zona contém as formas que jamais serão utilizadas como descritores, como palavras-chave. A parte mais importante da terceira zona é composta por formas que ocorrem uma única vez no *corpus* e que representam frequentemente 50% das formas distintas de um *corpus* (ROUAULT, 1987). O autor volta a enfatizar que o limite entre a primeira, a segunda e a terceira zonas é muito tênue. Segundo ele, a fronteira deve ser resultado de uma decisão arbitrária de fixação de um limite entre as efetivas formas de cada zona e depende do *corpus* utilizado. Finalmente, ao abordar a identificação de três zonas de ocorrência de palavras, ele ressalta que a segunda e o

final da primeira caracterizam-se pela presença de palavras com potencial para representar tematicamente os textos analisados.

Os estudos de indexação, fundamentados na frequência de ocorrência de palavras, têm evoluído para tentativas de desenvolvimento de algoritmos. O objetivo principal é o de contribuir para a automatização, em parte ou no todo, da indexação temática da informação.

Neste estudo, é analisado o potencial de índice de cada base e de cada sufixo nominalizador, investigando-se os padrões mais produtivos de formação de nominalizações deverbais e calculando-se o grau de relevância relativo a cada padrão identificado e a sua relação com o processo de indexação.

## 5 ANÁLISE E DISCUSSÃO DOS RESULTADOS

Nessa etapa, descreve-se, morfológica e estatisticamente, a importância relativa dos padrões distintos de nominalização deverbal em foco. Em seguida, aplica-se um método de análise bibliométrico aos artigos referenciados no item 4.1.2. Na análise morfológica, à luz das teorias abordadas no item 3, são investigados os graus relativos de produtividade de padrões recorrentes das nominalizações deverbais e dos diferentes sufixos nominalizadores que as compõem. Na análise bibliométrica, foram aplicadas as fórmulas das leis de Zipf e do Ponto de Transição de Goffman, assim como delimitada a região de concentração de palavras de alto conteúdo semântico, de acordo com Miranda Pao (1978). A partir daí, foram verificados os padrões de nominalizações deverbais e os sufixos nominalizadores que possivelmente ocorrem nessa região e seus potenciais de termos índice relevantes, na indexação. Finalmente, considera-se a regularidade de algumas nominalizações deverbais relevantes, do ponto de vista da indexação, apontando-se a correspondência entre as estruturas nominais e verbais.

### Artigo 1

RIZZON, Luiz Antenor; ZANUZ, Mauro C.; MIELE, Alberto. Evolução da acidez durante a vinificação de uvas tintas de três regiões vitícolas do Rio Grande do Sul. **Ciência e Tecnologia de Alimentos**, Campinas, v. 18, n. 2, p. 179-183, Maio/Jul. 1998.  
Disponível em: [http://www.scielo.br/scielo.php?pid=S0101-20611998000200007&script=sci\\_arttext](http://www.scielo.br/scielo.php?pid=S0101-20611998000200007&script=sci_arttext)  
Acesso em: 27/05/2009

O artigo 1 reúne um total de 2126 palavras, contendo 604 palavras distintas que equivalem a aproximadamente 28 (28,41)% do total de palavras. A frequência média por palavra, isto é, o índice médio de repetição de cada palavra no *corpus* analisado é de aproximadamente 4 (3,51).

A análise das estruturas léxico-morfológicas do artigo 1 revelou a superioridade no padrão de formação de nominalizações deverbais com o sufixo *-ção* sobre o número de ocorrências de cada padrão investigado. Com o sufixo *-ção*, foram identificadas 36 bases distintas que ocorreram 98 vezes, com a frequência média aproximada igual a 3 (2,72), correspondendo a cerca de 94 (94,23)% do total de ocorrências (104) dos diferentes sufixos e a cerca de 88 (87,80)% do total de bases distintas (41).

O sufixo *-mento* foi ligado a 2 bases distintas que ocorreram 3 vezes, com a frequência média aproximada igual a 1,5, correspondendo a cerca de 3 (2,88)% do total de ocorrências (104) e a

cerca de 5 (4,87)% do total de bases distintas (41). Com o sufixo *-agem* não foram identificadas nominalizações deverbais. O sufixo *-ncia*, com 2 bases distintas, foi empregado 2 vezes, com frequência média 1, correspondendo a cerca de 2 (1,92)% do total de ocorrências (104) e a cerca de 5 (4,87)% do total de bases distintas (41). O sufixo *-da*, com 1 base que ocorreu 1 vez, apresentou frequência média 1, correspondendo a cerca de 1 (0,96)% do total de ocorrências (104) e a cerca de 2 (2,43)% do total de bases distintas (41).

Segue a tabela 1 que sintetiza os sufixos nominalizadores no artigo 1 e suas frequências relativas de ocorrência, assim como o grau de produtividade dos padrões de formação de palavras, estimados pela observação da ocorrência de bases/produtos distintos de nominalização. A tabela também demonstra a força lexical relativa dos distintos padrões de produção de nominalizações.

Tabela 1 Síntese da produtividade dos sufixos nominalizadores do artigo 1 e da frequência relativa de ocorrência.

Rank	Sufixos	Frequência Total por Sufixo	Frequência Relativa do Sufixo/ Frequência Total de Sufixos	Bases Distintas	Bases Distintas/ Total de Bases Distintas	Frequência Média das Bases Distintas
1	-ção	98	94,23%	36	87,8%	2,72
2	-mento	3	2,88%	2	4,87%	1,5
3	-agem	0	0	0	0%	0
4	-ncia	2	1,92%	2	4,87%	1
5	-da	1	0,96%	1	2,43%	1

A seguir apresenta-se o Quadro 3 com a relação das nominalizações deverbais investigadas no artigo 1 e suas frequências relativas.

-ÇÃO			-ÇÕES		
1	fermentação	15	1	concentrações	3
2	vinificação	12	2	alterações	2
3	precipitação	9	3	determinações	1
4	concentração	6	4	variações	1
5	relação	4	<b>Totais</b>	<b>4</b>	<b>7</b>
6	maceração	4			
7	redução	4			
8	diminuição	3			
9	estabilização	3			
10	absorção	2			
11	dissolução	2			
12	extração	2			

13	formação	2
14	liberação	2
15	maturação	2
16	adubação	1
17	composição	1
18	constituição	1
19	correlação	1
20	degradação	1
21	determinação	1
22	elaboração	1
23	elevação	1
24	eluição	1
25	evolução	1
26	insolubilização	1
27	introdução	1
28	operação	1
29	oxidação	1
30	produção	1
31	proporção	1
32	separação	1
33	titulação	1
34	transformação	1
<b>Totais</b>	<b>34</b>	<b>91</b>
<b>-MENTO</b>		
1	esmagamento	2
2	envelhecimento	1
<b>Totais</b>	<b>2</b>	<b>3</b>
<b>-NCIA</b>		
1	interferência	1
2	variância	1
<b>Totais</b>	<b>2</b>	<b>2</b>
<b>-AGEM</b>		
<b>Totais</b>	<b>0</b>	<b>0</b>
<b>-DA</b>		
1	retirada	1
<b>Totais</b>	<b>1</b>	<b>1</b>
<b>-MENTOS</b>		
<b>Totais</b>	<b>0</b>	<b>0</b>
<b>-NCIAS</b>		
<b>Totais</b>	<b>0</b>	<b>0</b>
<b>-AGENS</b>		
<b>Totais</b>	<b>0</b>	<b>0</b>
<b>-DAS</b>		
<b>Totais</b>	<b>0</b>	<b>0</b>

Quadro 3 Relação das nominalizações deverbais investigadas no artigo 1.

A análise bibliométrica do artigo indicou que, no *rank* 1, uma única palavra (de) foi empregada 142 vezes, equivalendo aproximadamente a 7 (6,68)% do total de palavras. Em contrapartida, 380 itens ocorreram uma única vez ( $I_1 = 380$ ), equivalendo aproximadamente a 18 (17,87)% do total de palavras do artigo e a cerca de 63 (62,91)% do total de palavras distintas (604).

A aplicação da Primeira Lei de Zipf, considerando como palavras de alta frequência as que se situam na ordem de serie (  $r$  ) igual ou menor que 5 ( $r \leq 5$ ), apresentou desvios da média de

216,4, para mais, de 10,9% (  $r=3$ ), de 46,02% (  $r = 4$ ), de 1,66% (  $r = 5$ ) e, para menos, de 34,38% (  $r=1$  ), de 24,21%(  $r = 2$  ). A aplicação da segunda Lei de Zipf (  $n \leq 5$ ), onde  $n = f$ , apresentou desvios, para menos, de cerca de 29,21% (  $n = 2$ ), 52,62% (  $n = 3$ ), 60,52% (  $n = 4$ ) e 32,88% (  $n = 5$ ) (Tabela 9).

O Ponto T de Goffman localiza-se aproximadamente na frequência 27 (27,07). Se projetarmos para baixo o mesmo número de palavras que ocorre acima da frequência 27, um total de 21 palavras é identificado. Nessa região, ocorrem 15 operadores sintáticos, tais como: preposições, artigos, conjunções e pronomes. Esses itens foram excluídos não só porque não têm potencial de derivação para nominalizações, mas também por não poderem ser utilizados na representação temática da informação. Assim, na Região de Transição de Goffman, foram identificadas as seguintes palavras: *vinhos*, *ácido*, *Sant*, *Ana*, *Livramento*, a sigla *pH* (Potencial Hidrogeniônico) e o símbolo *k*, que integram o resumo e são de alto teor semântico em relação ao artigo em análise.

Logo após, identificam-se as palavras *tartárico* (frequência 18), *acidez* (frequência 17), *fermentação* (frequência 16), *vinho* (frequência 13), *vinificação* (frequência 12), entre outras, que pertencem à terminologia da área da indústria de vinhos, sendo que *acidez*, *fermentação* e *vinificação* constam do título e/ou do resumo do artigo 1. Se considerarmos a soma das frequências de *ácidos* (frequência 8) e *ácido* (frequência 23), teríamos a frequência 31, onde também ocorre *vinhos*. Foram identificadas as seguintes nominalizações deverbais relevantes do ponto de vista da indexação: *Livramento* (frequência 20), *acidez* (frequência 17), *fermentação* (frequência 15), *vinificação* (frequência 12), *maceração* (frequência 4), *maturação* (frequência 4), *esmagamento* (frequência 2), entre outras, da terminologia da área de indústria de vinhos.

Há de se ressaltar que o artigo tem como objetivo avaliar a evolução da acidez durante o processo de vinificação da uva e esclarece que a *acidez* e o *pH* dos vinhos podem ser alterados durante a vinificação de acordo com o teor de *K* da uva.

Segue o quadro 4 com o *rank*, a palavra, a frequência e o produto da ordem de série multiplicada pela frequência, verificados na Região de Transição de Goffman.



Rank	Word	Frequency	C		
1	De	142	142		
2	A	82	164		
3	Do	80	240		
4	E	79	316		
5	Da	44	220		
6	O	42	252		
7	Os	32	224		
8	Vinhos	31	248		
9	pH	30	270		
10	Na	29	290		
11	Em	27	297		
12	Com	25	300		
13	Ácido	23	299		PONTO T DE GOFFMAN: 27 (27,07)
14	K	23	322		
15	Ana	20	300	TOTAL DE PALAVRAS:	2126
16	Livramento	20	320	I <sub>1</sub>	380
17	Sant	20	340	I <sub>2</sub>	89
18	Se	20	360	I <sub>3</sub>	30
19	Que	19	361	I <sub>4</sub>	15
20	As	18	360	I <sub>5</sub>	17
21	Dos	18	378		

Quadro 4 Representação da Região de Transição de Goffman no artigo 1.

## Artigo 2

RIZZON, Luiz Antenor; MIELE, Alberto. Avaliação da cv. Merlot para elaboração de vinho tinto. **Ciência e Tecnologia de Alimentos**, Campinas, v. 23, suppl., p. 156-161, dez. 2003.  
Disponível em: [http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S0101-20612003000400029&lng=en&nrm=iso](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0101-20612003000400029&lng=en&nrm=iso)  
Acesso em: 27/05/2009

O artigo 2 reúne um total de 3256 palavras, contendo 843 palavras distintas que equivalem a aproximadamente 26 (25,89) % do total de palavras. A frequência média por palavra, isto é, o índice médio de repetição de cada palavra no *corpus* analisado é de aproximadamente 4 (3,86).

A análise das estruturas léxico-morfológicas do artigo 2 revelou a superioridade no padrão de formação de nominalizações deverbais com o sufixo *-ção* sobre o número de ocorrências de cada padrão investigado. Com o sufixo *-ção*, foram identificadas 47 bases distintas que ocorreram 112 vezes, com a frequência média aproximada igual a 2 (2,38), correspondendo a cerca de 81 (80,57) % do total de ocorrências (139) dos diferentes sufixos e a cerca de 72 (72,30) % do total de bases distintas (65).

O sufixo *-mento* foi ligado a 9 bases distintas que ocorreram 11 vezes, com a frequência média aproximada igual a 1 (1,22), correspondendo a cerca de 8 (7,91)% do total de ocorrências (139) e a cerca de 14 (13,84)% do total de bases distintas (65). O sufixo *-agem*, com 3 bases, ocorreu 6 vezes, com frequência média 2, correspondendo a cerca de 4 (4,31) % do total de ocorrências (139) e a cerca de 5 (4,61)% do total de bases distintas (65). O sufixo *-ncia*, com 4 bases distintas, foi empregado 6 vezes, com frequência média 1,5, correspondendo a cerca de 4 (4,31) % do total de ocorrências (139) e a cerca de 6 (6,15)% do total de bases distintas (65). O sufixo *-da* foi ligado a 2 bases que ocorreram 4 vezes, com frequência média 2, correspondendo a cerca de 3 (2,87)% do total de ocorrências (139) e a cerca de 3 (3,07)% do total de bases distintas (65).

Segue a Tabela 2 que sintetiza os sufixos nominalizadores no artigo 2 e suas frequências relativas de ocorrência, assim como o grau de produtividade dos padrões de formação de palavras estimados pela observação da ocorrência de bases/produtos distintos de nominalização. A tabela também demonstra a força lexical relativa dos distintos padrões de produção de nominalizações.

Tabela 2 Síntese da produtividade dos sufixos nominalizadores do artigo 2 e da frequência relativa de ocorrência.

Rank	Sufixos	Frequência Total por Sufixo	Frequência Relativa do Sufixo/ Frequência Total de Sufixos	Bases Distintas	Bases Distintas/ Total de Bases Distintas	Frequência Média das Bases Distintas
1	-ção	112	80,57%	47	72,30%	2,38
2	-mento	11	7,91%	9	13,84%	1,22
3	-agem	6	4,31%	3	4,61%	2
4	-ncia	6	4,31%	4	6,15%	1,5
5	-da	4	2,87%	2	3,070%	2

A seguir apresenta-se o Quadro 5 com a relação das nominalizações deverbais investigadas no artigo 2 e suas frequências relativas.

-ÇÃO			-ÇÕES		
1	Maturação	15	1	concentrações	2
2	Evolução	8	2	avaliações	1
3	Redução	7	3	determinações	1
4	Vinificação	6	4	informações	1
5	Fermentação	5	5	microvinificações	1
6	Produção	5	6	modificações	1

7	Relação	5
8	avaliação	3
9	determinação	3
10	maceração	3
11	salificação	3
12	variação	3
13	absorção	2
14	composição	2
15	diminuição	2
16	estabilização	2
17	formação	2
18	transformação	2
19	utilização	2
20	acumulação	1
21	aplicação	1
22	caracterização	1
23	conservação	1
24	diluição	1
25	elaboração	1
26	eluição	1
27	evaporação	1
28	extração	1
29	graduação	1
30	inserção	1
31	introdução	1
32	ionização	1
33	liberação	1
34	mobilização	1
35	oxirredução	1
36	percepção	1
37	pulverização	1
38	seleção	1
39	separação	1
<b>Totais</b>	<b>39</b>	<b>100</b>
<b>-MENTO</b>		
1	crescimento	2
2	esmagamento	2
3	abastecimento	1
4	acompanhamento	1
5	engarrafamento	1
6	envelhecimento	1
7	espaçamento	1
8	pegamento	1
9	tratamento	1
<b>Totais</b>	<b>9</b>	<b>11</b>
<b>-NCIA</b>		
1	decorrência	2
2	predominância	2
3	influência	1
4	tendência	1
<b>Totais</b>	<b>4</b>	<b>6</b>
<b>-AGEM</b>		
7	operações	1
8	precipitações	1
9	reações	1
10	turvações	1
11	variações	1
<b>Totais</b>	<b>11</b>	<b>12</b>
<b>-MENTOS</b>		
<b>Totais</b>	<b>0</b>	<b>0</b>
<b>-NCIAS</b>		
<b>Totais</b>	<b>0</b>	<b>0</b>
<b>-AGENS</b>		

1	pesagem	2	1	amostragens	1
2	prensagem	2	<b>Totais</b>	<b>1</b>	<b>1</b>
3	amostragem	1			
<b>Totais</b>	<b>3</b>	<b>5</b>			
-DA			-DAS		
1	medida	1	1	medidas	2
2	poda	1	<b>Totais</b>	<b>1</b>	<b>2</b>
<b>Totais</b>	<b>2</b>	<b>2</b>			

Quadro 5 Relação das nominalizações deverbais investigadas no artigo 2.

A análise bibliométrica do artigo indicou que, no *rank* 1, uma única palavra (de) foi empregada 189 vezes, equivalendo aproximadamente a 6 (5,80) % do total de palavras. Em contrapartida, 492 itens ocorreram uma única vez ( $I_1 = 492$ ), equivalendo aproximadamente a 15 (15,11)% do total de palavras do artigo 2 e a cerca de 58 (58,36)% do total de palavras distintas.

A aplicação da Primeira Lei de Zipf, considerando como palavras de alta frequência as que se situam na ordem de serie (  $r$  ) igual ou menor que 5 ( $r \leq 5$ ), apresentou desvios da média de 343,4, para mais, de 30,46% ( $r=4$ ), de 45,60% ( $r = 5$ ), e, para menos, de 44,96% ( $r=1$ ), de 28,94% ( $r = 2$ ) e de 2,15% ( $r=3$ ). A aplicação da Segunda Lei de Zipf ( $n \leq 5$ ), onde  $n = f$ , apresentou desvios, para menos, de cerca de 20,12% ( $n = 2$ ), 7,31% ( $n = 3$ ), 36,99% ( $n = 4$ ), e 35,97% ( $n = 5$ ) (Tabela 9).

O Ponto T de Goffman localiza-se aproximadamente na frequência 31 (30,87). Se projetarmos para baixo o mesmo número de palavras que ocorre acima da frequência 31, um total de 30 palavras é identificado. Nessa região, ocorrem 21 operadores sintáticos, tais como: preposições, artigos, conjunções, pronomes e verbos auxiliares. Essas palavras foram excluídas na medida em que o objetivo do estudo é a investigação da representação temática dos termos na Região de Transição de Goffman. Assim, foram identificadas as seguintes palavras: *vinho, Merlot, uva, mosto, teor, peso, acidez, safras e baga*, que integram o título e/ou resumo e/ou palavras-chave e são de alto teor semântico em relação ao artigo em análise.

Logo após, são identificadas a sigla *pH* (frequência 18), as palavras *bagas* (frequência 17), *ácido* (frequência 15), *maturação* (frequência 15), *cacho* (frequência 14), *vinhos* (frequência 12), entre outras, que constam da terminologia da área da indústria de vinhos. Foram identificadas as seguintes nominalizações deverbais relevantes, do ponto de vista da indexação: *acidez* (23), *maturação* (frequência 15), *vinificação* (frequência 6), *fermentação*

(frequência 5), *produção* (frequência 5), *maceração* (frequência 3), *salificação* (frequência 3), *esmagamento* (frequência 2), relativas à indústria de vinhos.

Há de se ressaltar que o artigo tem como objetivo avaliar o potencial da *cultivar vinífera* Merlot para elaboração de vinho tinto, analisando a maturação e a caracterização da uva, do cacho, do mosto e do vinho.

Segue o quadro 6 com o *rank*, a palavra, a frequência e o produto da ordem de série vezes a frequência, da Região de Transição de Goffman.

Rank	Word	Frequency	C		
1	De	189	189		
2	Do	122	244		
3	A	112	336		
4	Da	112	448		
5	E	100	500		
6	O	89	534		
7	Vinho	51	357		
8	Em	47	376		
9	Merlot	44	396		
10	Se	42	420		
11	Uva	42	462		
12	Que	39	468		
13	Na	34	442		
14	No	33	462		
15	Com	31	465		
16	Para	31	496		
17	Mosto	30	510		
18	Os	30	540		
19	Teor	28	532		
20	Foi	27	540		
21	Peso	26	546		
22	As	24	528		
23	Acidez	23	529		
24	Ao	23	552		
25	Por	23	575		
26	Das	22	572		
27	Foram	21	567		
28	Safras	21	588		
29	Baga	19	551		
30	É	19	570		
				PONTO T DE GOFFMAN: 31 (30,87)	
				TOTAL DE PALAVRAS: 3256	
				I <sub>1</sub>	492
				I <sub>2</sub>	131
				I <sub>3</sub>	76
				I <sub>4</sub>	31
				I <sub>5</sub>	21

Quadro 6 Representação da Região de Transição de Goffman no artigo 2.

### Artigo 3

RIZZON, Luiz Antenor; MIELE Alberto. Avaliação da cv. Cabernet Sauvignon para elaboração de vinho tinto. . **Ciência e Tecnologia de Alimentos**, Campinas, v. 22, n. 2, p. 192-198, maio/ago. 2002.  
Disponível em: <http://www.scielo.br/pdf/cta/v22n2/a15v22n2.pdf>  
Acesso em: 27/05/2009

O artigo 3 reúne um total de 4007 palavras, contendo 1024 palavras distintas que equivalem a aproximadamente 26 (25,55) % do total de palavras. A frequência média por palavra, isto é, o índice médio de repetição de cada palavra no *corpus* analisado é de aproximadamente 4 (3,91).

A análise das estruturas léxico-morfológicas do artigo 3 revelou a superioridade no padrão de formação de nominalizações deverbais com o sufixo *-ção* sobre o número de ocorrências de cada padrão investigado. Com o sufixo *-ção*, foram identificadas 46 bases distintas que ocorreram 148 vezes, com a frequência média aproximada igual a 3,21, correspondendo a cerca de 82 (81,76) % do total de ocorrências (181) dos diferentes sufixos e a cerca de 68 (67,64) % do total de bases distintas (68).

O sufixo *-mento* foi ligado a 10 bases distintas que ocorreram 15 vezes, com a frequência média aproximada igual a 1,5, correspondendo a cerca de 8 (8,28) % do total de ocorrências (181) e a cerca de 15 (14,70) % do total de bases distintas (68). O sufixo *-agem*, com 4 bases, ocorreu 5 vezes, com frequência média 1,25, correspondendo a cerca de 3 (2,76) % do total de ocorrências (181) e a cerca de 6 (5,88)% do total de bases distintas (68). O sufixo *-ncia*, com 6 bases distintas, foi empregado 8 vezes, com frequência média 1,33, correspondendo a cerca de 4 (4,41) % do total de ocorrências (181) e a cerca de 9 (8,82) % do total de bases distintas (68). O sufixo *-da*, com 2 bases distintas, foi empregado 5 vezes, com frequência média 2,5, correspondendo a cerca de 3 (2,76) % do total de ocorrências (181) e a cerca de 3 (2,94) % do total de bases distintas (68).

Segue a tabela 3, que sintetiza os sufixos nominalizadores no artigo 3 e suas frequências relativas de ocorrência, assim como o grau de produtividade dos padrões de formação de palavras, estimados pela observação da ocorrência de bases/produtos distintos de nominalização. A tabela também demonstra a força lexical relativa dos distintos padrões de produção de nominalizações.

Tabela 3 Síntese da produtividade dos sufixos nominalizadores do artigo 3 e da frequência relativa de ocorrência.

Rank	Sufixos	Frequência Total por Sufixo	Frequência Relativa do Sufixo/ Frequência Total de Sufixos	Bases Distintas	Bases Distintas/ Total de Bases Distintas	Frequência Média das Bases Distintas
1	-ção	148	81,76%	46	67,64%	3,21
2	-mento	15	8,28%	10	14,70%	1,50
3	-agem	5	2,76%	4	5,88%	1,25
4	-ncia	8	4,41%	6	8,82%	1,33
5	-da	5	2,76%	2	2,94%	2,50

A seguir apresenta-se o Quadro 7 com a relação das nominalizações deverbais investigadas no artigo 3 e suas frequências relativas.

-ÇÃO			-ÇÕES		
1	Maturação	23	1	concentrações	4
2	Vinificação	9	2	variações	3
3	Concentração	8	3	alterações	1
4	Evolução	8	4	avaliações	1
5	Relação	8	5	determinações	1
6	Fermentação	6	6	diminuições	1
7	Salificação	6	7	informações	1
8	avaliação	5	8	microvinificações	1
9	precipitação	5	9	pulverizações	1
10	produção	5	10	relações	1
11	maceração	4	11	transformações	1
12	redução	4	12	turvações	1
13	composição	3	<b>Totais</b>	<b>12</b>	<b>17</b>
14	elaboração	3			
15	utilização	3			
16	absorção	2			
17	floração	2			
18	liberação	2			
19	variação	2			
20	antecipação	1			
21	brotação	1			
22	chaptalização	1			
23	correção	1			
24	correlação	1			
25	determinação	1			
26	diluição	1			
27	diminuição	1			
28	eluição	1			
29	extração	1			
30	implantação	1			
31	indicação	1			
32	introdução	1			
33	ionização	1			

34	microvinificação	1
35	modificação	1
36	oxidorredução	1
37	participação	1
38	percepção	1
39	pulverização	1
40	respiração	1
41	separação	1
42	valorização	1
<b>Totais</b>	<b>42</b>	<b>131</b>
<b>-MENTO</b>		
1	amadurecimento	4
2	crescimento	2
3	envelhecimento	2
4	abastecimento	1
5	acompanhamento	1
6	desenvolvimento	1
7	dessecamento	1
8	esmagamento	1
9	espaçamento	1
10	pegamento	1
<b>Totais</b>	<b>10</b>	<b>15</b>
<b>-NCIA</b>		
1	decorrência	2
2	resistência	2
3	incidência	1
4	influência	1
5	interferência	1
6	predominância	1
<b>Totais</b>	<b>6</b>	<b>8</b>
<b>-AGEM</b>		
1	pesagem	2
2	amostragem	1
3	prensagem	1
<b>Totais</b>	<b>3</b>	<b>4</b>
<b>-DA</b>		
1	demanda	2
2	medida	1
<b>Totais</b>	<b>2</b>	<b>3</b>
<b>-MENTOS</b>		
<b>Totais</b>	<b>0</b>	<b>0</b>
<b>-NCIAS</b>		
<b>Totais</b>	<b>0</b>	<b>0</b>
<b>-AGENS</b>		
1	amostragens	1
2	remontagens	1
<b>Totais</b>	<b>2</b>	<b>1</b>
<b>-DAS</b>		
1	medidas	2
<b>Totais</b>	<b>1</b>	<b>2</b>

Quadro 7 Relação das nominalizações deverbais investigadas no artigo 3.

A análise bibliométrica do artigo indicou que, no *rank* 1, a preposição *de* foi empregada 227 vezes, equivalendo aproximadamente a 6 (5,66) % do total de palavras. Em contrapartida, 593 itens ocorreram uma única vez ( $I_1 = 593$ ), equivalendo aproximadamente a 15 (14,79)% do total de palavras do artigo 3 e a cerca de 58 (57,91)% do total de palavras distintas .

A aplicação da Primeira Lei de Zipf, considerando como palavras de alta frequência as que se situam na ordem de serie (  $r$  ) igual ou menor que 5 ( $r \leq 5$ ), apresentou desvios da média de



363, para mais, de 21,21% ( r=4), de 46,00% ( r = 5), e, para menos, de 37,46% (r=1) , de 27,27%( r = 2 ) e de 2,47% (r=3). A aplicação da Segunda Lei de Zipf (n ≤ 5), onde n = f, apresentou desvios, para menos, de 25,12% ( n = 2), 20,06% ( n = 3), 29,17% ( n = 4).e 1,34% ( n = 5) (Tabela 9).

O Ponto T de Goffman localiza-se aproximadamente na frequência 34 (33,94). Se projetarmos para baixo o mesmo número de palavras que ocorre acima da frequência 34, um total de 32 palavras é identificado. Nessa região, ocorrem 22 formas do tipo preposições, artigos, conjunções, pronomes e verbos auxiliares. Essas formas foram excluídas na medida em que o objetivo do estudo é a investigação da representação temática dos termos na Região de Transição de Goffman. Assim, foram identificadas as palavras *uva*, *vinho*, *Sauvignon*, *Cabernet*, *peso*, *vinhos*, *teor*, *mosto*, *maturação* e *baga*, que apresentam alta carga semântica em relação ao artigo em análise e, excluindo-se *peso* e *maturação*, ocorrem no título e/ou no resumo e/ou nas palavras-chave.

Logo após são identificadas as seguintes palavras: *acidez* (frequência 20), *cacho* (frequência 16), *colheita* (frequência 15), *açúcar* (frequência 15), *safras* (frequência 13), *bagas* (frequência 13), *teores* (frequência 12) e a sigla *pH* (frequência 14), entre as demais que são adstritas à indústria de vinhos. Foram identificadas as seguintes nominalizações relevantes do ponto de vista da indexação: *maturação* (frequência 23), *acidez* (frequência 20), *vinificação* (frequência 9), *fermentação* (frequência 6), *salificação* (frequência 6), *maceração* (frequência 4), *floração* (frequência 2), *amadurecimento* (frequência 4), relativas à produção de vinhos.

Há de se ressaltar que o artigo tem como objetivo a avaliação da cv. *Cabernet Sauvignon* para a elaboração de vinho tinto, por meio da caracterização do cacho, do mosto e do vinho.

Segue o quadro 8 com o *rank*, a palavra, a frequência e o produto da ordem de série vezes a frequência, da Região de Transição de Goffman.

Rank	Word	Frequency	C
1	De	227	227
2	Da	132	264
3	A	118	354
4	E	110	440
5	O	106	530
6	Do	83	498

7	Em	60	420		
8	Na	48	384		
9	Que	47	423		
10	Se	46	460		
11	Uva	45	495		
12	Com	39	468		
13	Vinho	38	494		
14	Os	36	504		
15	No	36	540		
16	l	36	576		
17	Foi	33	561		
18	É	32	576		
19	Sauvignon	30	570		
20	Cabernet	30	600		
21	Por	29	609		
22	Para	29	638		
23	Peso	28	644		
24	As	28	672		
25	Um	27	675		
26	Vinhos	26	676		
27	Teor	25	675		
28	Mosto	24	672		
29	Maturação	23	667		
30	Dos	22	660		
31	Baga	22	682		
32	À	21	672		
				PONTO T DE GOFFMAN: 34 (33,94)	
				TOTAL DE PALAVRAS: 4007	
				I <sub>1</sub>	593
				I <sub>2</sub>	148
				I <sub>3</sub>	79
				I <sub>4</sub>	42
				I <sub>5</sub>	39

Quadro 8 Representação da Região de Transição de Goffman no artigo 3

#### Artigo 4

MAMEDE, Maria Eugênia de Oliveira; PASTORE, Gláucia Maria. Avaliação da produção dos compostos majoritários da fermentação de mosto da uva por leveduras isoladas da região da “Serra Gaúcha” (RS). **Ciência e Tecnologia de Alimentos**, Campinas, v. 24, n. 3, p. 453-458, jul./set. 2004.  
Disponível em: <http://www.scielo.br/pdf/cta/v24n3/21942.pdf>  
Acesso em: 27/05/2009

O artigo 4 reúne um total de 2369 palavras, contendo 636 palavras distintas que equivalem a aproximadamente 27 (26,84)% do total de palavras. A frequência média por palavra, isto é, o índice médio de repetição de cada palavra no *corpus* analisado é de aproximadamente 4 (3,72).

A análise das estruturas léxico-morfológicas do artigo 4 revelou a superioridade no padrão de formação de nominalizações deverbais com o sufixo *-ção* sobre o número de ocorrências de cada padrão investigado. Com o sufixo *-ção*, foram identificadas 33 bases distintas que ocorreram 157 vezes, com a frequência média aproximada igual a 4,75, correspondendo a cerca de 89 (88,70)% do total de ocorrências (177) dos diferentes sufixos e a cerca de 83 (82,50)% do total de bases distintas (40).

O sufixo *-mento* foi ligado a 3 bases distintas que ocorreram 16 vezes, com a frequência média aproximada igual a 5 (5,33), correspondendo a cerca de 9 (9,03)% do total de ocorrências (177) e a cerca de 8 (7,50)% do total de bases distintas (40). O sufixo *-agem*, com 1 base, ocorreu 1 vez, com frequência média 1, correspondendo a cerca de 1 (0,56)% do total de ocorrências (177) e a cerca de 3 (2,50)% do total de bases distintas (40). O sufixo *-ncia*, com 3 bases distintas, foi empregado 3 vezes, com frequência média 1, correspondendo a cerca de 2 (1,69)% do total de ocorrências (177) e a cerca de 8 (7,50)% do total de bases distintas (40). O sufixo *-da* não ocorreu na formação de nominalizações deverbais no artigo em análise.

Segue a Tabela 4, que sintetiza os sufixos nominalizadores no artigo 4 e suas frequências relativas de ocorrência, assim como o grau de produtividade dos padrões de formação de palavras estimados pela observação da ocorrência de bases/produtos distintos de nominalização. A tabela também demonstra a força lexical relativa dos distintos padrões de produção de nominalizações.

Tabela 4 Síntese da produtividade dos sufixos nominalizadores do artigo 4 e da frequência relativa de ocorrência.

Rank	Sufixos	Frequência Total por Sufixo	Frequência Relativa do Sufixo/ Frequência Total de Sufixos	Bases Distintas	Bases Distintas/ Total de Bases Distintas	Frequência Média das Bases Distintas
1	-ção	157	88,70%	33	82,50%	4,75
2	-mento	16	9,03%	3	7,50%	5,33
3	-agem	1	0,56%	1	2,50%	1
4	-ncia	3	1,69%	3	7,50%	1
5	-da	0	0%	0	0%	0,0

A seguir apresenta-se o Quadro 9 com a relação das nominalizações deverbais investigadas no artigo 4 e suas frequências relativas.

-ÇÃO			-ÇÕES		
1	fermentação	59	1	concentrações	9
2	concentração	21	2	fermentações	2
3	produção	21	3	produções	1
4	identificação	6	4	variações	1
5	variação	4	<b>Totais</b>	<b>4</b>	<b>13</b>
6	avaliação	2			

7	desorção	2
8	ebulição	2
9	injeção	2
10	programação	2
11	adaptação	1
12	adição	1
13	aplicação	1
14	assimilação	1
15	comparação	1
16	contaminação	1
17	contribuição	1
18	descarboxilação	1
19	detecção	1
20	fabricação	1
21	formação	1
22	interação	1
23	interpretação	1
24	introdução	1
25	investigação	1
26	ionização	1
27	obtenção	1
28	padronização	1
29	quantificação	1
30	reação	1
31	relação	1
32	seleção	1
33	titulação	1
<b>Totais</b>	<b>33</b>	<b>144</b>
<b>-MENTO</b>		
1	crescimento	13
2	isolamento	2
3	comportamento	1
<b>Totais</b>	<b>3</b>	<b>16</b>
<b>-NCIA</b>		
1	influência	1
2	transferência	1
3	variância	1
<b>Totais</b>	<b>3</b>	<b>3</b>
<b>-AGEM</b>		
1	secagem	1
<b>Totais</b>	<b>1</b>	<b>1</b>
<b>-DA</b>		
<b>Totais</b>	<b>0</b>	<b>0</b>
<b>-MENTOS</b>		
<b>Totais</b>	<b>0</b>	<b>0</b>
<b>-NCIAS</b>		
<b>Totais</b>	<b>0</b>	<b>0</b>
<b>-AGENS</b>		
<b>Totais</b>	<b>0</b>	<b>0</b>
<b>-DAS</b>		
<b>Totais</b>	<b>0</b>	<b>0</b>

Quadro 9 Relação das nominalizações deverbais investigadas no artigo 4.

A análise bibliométrica do artigo indicou que, no *rank* 1, a preposição *de* ocorreu 197 vezes, equivalendo aproximadamente a 8 (8,31)% da totalidade dos itens processados. Em contrapartida, 380 itens ocorreram uma única vez ( $I_1 = 380$ ), equivalendo aproximadamente a

16 (16,04)% do total de palavras do artigo 4 e a cerca de 60 (59,74)% do total de palavras distintas.

A aplicação da Primeira Lei de Zipf, considerando como palavras de alta frequência as que se situam na ordem de serie (  $r$  ) igual ou menor que 5 (  $r \leq 5$  ), apresentou desvios da média de 224,4, para mais, de 5,16% (  $r=4$  ), de 13,63% (  $r = 5$  ), e, para menos, de 12,21% (  $r=1$  ), de 2,85% (  $r = 2$  ) e de 3,74% (  $r=3$  ). A aplicação da Segunda Lei de Zipf (  $n \leq 5$  ), onde  $n = f$ , apresentou desvios, para menos, de cerca de 28,15% (  $n = 2$  ), 32,10% (  $n = 3$  ), 23,68% (  $n = 4$  ). e 40,78% (  $n = 5$  ) (Tabela 9).

O Ponto T de Goffman localiza-se aproximadamente na frequência 27 (27,07). Se projetarmos para baixo o mesmo número de palavras que ocorre acima da frequência 27, um total de 20 palavras é identificado. Nessa região, preposições, artigos, conjunções, pronomes e verbos auxiliares ocorreram 11 vezes. Essas palavras foram excluídas dos dados na medida em que o objetivo do estudo era a investigação temática dos termos na Região de Transição de Goffman. Assim, foram identificadas as seguintes palavras: *fermentação*, *Kloeckera*, *Apiculata*, *Saccharomyces*, *leveduras*, *mostos*, *Cerevisiae*, *concentração* e *produção*, que são de alto teor semântico em relação ao artigo em análise e, excluindo-se *mostos*, que ocorre no título no singular, integram o título e/ou resumo e/ou palavras-chave.

Logo após são identificadas as palavras *mosto* (frequência 20), *Chardonnay* (frequência 19), *Pinot* (frequência 17), *Noir* (frequência 17), *compostos* (frequência 16), *amostras* (frequência 13), *crescimento* (frequência 13), *voláteis* (frequência 12), *vinho* (frequência 11), *temperatura* (frequência 10), *concentrações* (frequência 09), *uva* (frequência 09), dentre as demais referentes à indústria de vinhos. Foram identificadas nominalizações deverbais relevantes do ponto de vista da indexação: *fermentação* (frequência 59), *concentração* (frequência 21), *produção* (frequência 21), *ebulição* (frequência 2), *crescimento* (frequência 13), *isolamento* (frequência 2), entre outras, do domínio contextual da produção de vinhos.

Há de se ressaltar que o artigo tem como objetivo estudar o crescimento e a produção de compostos voláteis durante a fermentação de mosto de uva pelas leveduras *Kloeckera apiculata* e *Saccharomyces cerevisiae*.

Segue o quadro 10 com o *rank*, a palavra, a frequência e o produto da ordem de série vezes a frequência, da Região de Transição de Goffman.

Rank	Word	Frequency	C		
1	De	197	197	PONTO T DE GOFFMAN: 27 (27,07)	
2	A	109	218		
3	E	72	216	TOTAL DE PALAVRAS:	2369
4	Fermentação	59	236	I <sub>1</sub>	380
5	Da	51	255	I <sub>2</sub>	91
6	O	41	246	I <sub>3</sub>	43
7	Em	37	259	I <sub>4</sub>	29
8	Foi	33	264	I <sub>5</sub>	15
9	Do	32	288		
10	As	29	290		
11	Kloeckera	25	275		
12	Apiculata	23	276		
13	Saccharomyces	23	299		
14	Com	22	308		
15	Leveduras	22	330		
16	Mostos	22	352		
17	Cerevisiae	21	357		
18	Concentração	21	378		
19	Foram	21	399		
20	Produção	21	420		

Quadro 10 Representação da Região de Transição de Goffman no artigo 4.

## Artigo 5

CRISÓSTOMO, Vicente Lima. Dificuldades das empresas brasileiras para financiar seus investimentos em capital físico e em inovação. **Revista de Economia Contemporânea**, Rio de Janeiro, v. 13, n. 2, p. 259-280, maio/ago., 2009.

Disponível em: <http://www.scielo.br/pdf/rec/v13n2/v13na04.pdf>

Acesso em: 27/01/2010

O artigo 5 reúne um total de 4414 palavras, contendo 1106 palavras distintas que equivalem a aproximadamente 25 (25,0566)% do total de palavras. A frequência média por palavra, isto é, o índice médio de repetição de cada palavra no *corpus* analisado é de aproximadamente 4 (3,9909).

A análise das estruturas léxico-morfológicas do artigo 5 revelou a superioridade no padrão de formação de nominalizações deverbais com o sufixo *-ção* sobre o número de ocorrências de cada padrão investigado. Com o sufixo *-ção*, foram identificadas 47 bases distintas que ocorreram 162 vezes, com a frequência média aproximada igual a 3 (3,44), correspondendo a cerca de 48 (47,78)% do total de ocorrências (339) dos diferentes sufixos e a cerca de 68 (68,11)% do total de bases distintas (69).

O sufixo *-mento* foi ligado a 12 bases distintas que ocorreram 151 vezes, com a frequência média aproximada igual a 13 (12,58), correspondendo a cerca de 45 (44,54)% do total de ocorrências (339) e a cerca de 17 (17,39)% do total de bases distintas (69). O sufixo *-agem* não ocorreu na formação de nominalizações deverbais no artigo em análise. O sufixo *-ncia*, com 10 bases distintas, foi empregado 26 vezes, com frequência média 2,6, correspondendo a cerca de 8 (7,66)% do total de ocorrências (339) e a cerca de 14 (14,49)% do total de bases distintas (69). O sufixo *-da* também não ocorreu na formação de nominalizações deverbais no artigo em análise.

Segue a Tabela 5, que sintetiza os sufixos nominalizadores no artigo 5 e suas frequências relativas de ocorrência, assim como o grau de produtividade dos padrões de formação de palavras estimados pela observação da ocorrência de bases/produtos distintos de nominalização. A tabela também demonstra a força lexical relativa dos distintos padrões de produção de nominalizações.

Tabela 5 Síntese da produtividade dos sufixos nominalizadores do artigo 5 e da frequência relativa de ocorrência.

Rank	Sufixos	Frequência Total por Sufixo	Frequência Relativa do Sufixo/ Frequência Total de Sufixos	Bases Distintas	Bases Distintas/ Total de Bases Distintas	Frequência Média das Bases Distintas
1	-ção	162	47,78%	47	68,11%	3,44
2	-mento	151	44,54%	12	17,39%	12,58
3	-agem	0	0%	0	0%	0,0
4	-ncia	26	7,66%	10	14,49%	2,6
5	-da	0	0%	0	0%	0,0

A seguir apresenta-se o Quadro 11 com a relação das nominalizações deverbais investigadas no artigo 5 e suas frequências relativas.

-ÇÃO			-ÇÕES		
1	Inovação	42	1	observações	6
2	Relação	17	2	restrições	6
3	Correlação	14	3	estimações	5
4	Produção	6	4	informações	2
5	Variação	6	5	condições	1
6	Estimação	3	6	contribuições	1
7	Obtenção	3	7	depreciações	1
8	Aplicação	2	8	interpretações	1
9	aproximação	2	9	motivações	1
10	contribuição	2	10	variações	1

11	Depreciação	2	<b>Totais</b>	<b>10</b>	<b>25</b>
12	interpretação	2			
13	Introdução	2			
14	maximização	2			
15	Proteção	2			
16	Restrição	2			
17	Acumulação	1			
18	Amortização	1			
19	argumentação	1			
20	Associação	1			
21	autocorrelação	1			
22	classificação	1			
23	comunicação	1			
24	concentração	1			
25	contrastação	1			
26	Distribuição	1			
27	Dotação	1			
28	Exacerbação	1			
29	Formação	1			
30	Implantação	1			
31	Informação	1			
32	Introdução	1			
33	investigação	1			
34	Limitação	1			
35	manifestação	1			
36	Manutenção	1			
37	Proposição	1			
38	Realização	1			
39	Rejeição	1			
40	reorganização	1			
41	Resolução	1			
42	Revelação	1			
43	Seleção	1			
44	sobreidentificação	1			
<b>Totais</b>	<b>44</b>	<b>137</b>			
<b>-MENTO</b>			<b>-MENTOS</b>		
1	investimento	66	1	investimentos	13
2	financiamento	39	2	questionamentos	1
3	endividamento	12	<b>Totais</b>	<b>2</b>	<b>14</b>
4	Faturamento	8			
5	desenvolvimento	6			
6	comportamento	1			
7	Crescimento	1			
8	fortalecimento	1			
9	racionamento	1			
10	sobreinvestimento	1			
11	Tratamento	1			
<b>Totais</b>	<b>11</b>	<b>137</b>			
<b>-NCIA</b>			<b>-NCIAS</b>		
1	Ausência	10			
2	Existência	4	<b>Totais</b>	<b>0</b>	<b>0</b>
3	Importância	3			



4	Influência	2			
5	Persistência	2			
6	dependência	1			
7	Falência	1			
8	Ocorrência	1			
9	Preferência	1			
10	Relevância	1			
<b>Totais</b>	<b>10</b>	<b>26</b>			
<b>-AGEM</b>			<b>-AGENS</b>		
<b>Totais</b>	<b>0</b>	<b>0</b>	<b>Totais</b>	<b>0</b>	<b>0</b>
<b>-DA</b>			<b>-DAS</b>		
<b>Totais</b>	<b>0</b>	<b>0</b>	<b>Totais</b>	<b>0</b>	<b>0</b>

Quadro 11 Relação das nominalizações deverbais investigadas no artigo 5.

A análise bibliométrica do artigo indicou que, no *rank* 1, uma única palavra (de), que foi empregada 351 vezes, equivalendo aproximadamente a 8 (7,95)% do total de palavras. Em contrapartida, 663 itens ocorreram uma única vez ( $I_1 = 663$ ), equivalendo aproximadamente a 15 (15,02)% do total de palavras do artigo 5 e a cerca de 60 (59,94)% do total de palavras distintas.

A aplicação da Primeira Lei de Zipf, considerando como palavras de alta frequência as que se situam na ordem de serie ( $r$ ) igual ou menor que 5 ( $r \leq 5$ ), apresentou desvios da média de 386,80, para mais, de 14,78% ( $r=4$ ), de 16,33% ( $r = 5$ ), e, para menos, de 9,25% ( $r=1$ ), de 15,71% ( $r = 2$ ) e de 6,15% ( $r=3$ ). A aplicação da Segunda Lei de Zipf ( $n \leq 5$ ), onde  $n = f$ , apresentou desvios, para menos, de cerca de 23,52% ( $n = 2$ ), 30,31% ( $n = 3$ ), 50,22% ( $n = 4$ ), e 29,86% ( $n = 5$ ) (Tabela 10).

O Ponto T de Goffman localiza-se aproximadamente na frequência 36 (35,91). Se projetarmos para baixo o mesmo número de palavras que ocorre acima da frequência 36, um total de 40 palavras é identificado. Nessa região, são identificados 27 itens como preposições, artigos, conjunções, pronomes, símbolos matemáticos e verbos auxiliares. Essas palavras foram excluídas na medida em que o objetivo do estudo é a investigação temática dos termos na Região de Transição de Goffman. Assim, foram verificadas os itens *investimento, empresa, inovação, capital, financiamento, mercado, físico, externo, pesquisa, empresas, dificuldades, relação e resultados*, que são de alta carga semântica em relação ao artigo em análise e, excluindo-se *mercado*, integram o título e/ou o resumo e/ou as palavras-chave.

Logo após são identificadas as palavras como *dificuldade* (frequência 15), *gastos* (frequência 14), *investimentos* (frequência 13), referentes ao campo semântico da área da Economia. São as seguintes nominalizações deverbais relevantes, do ponto de vista da indexação: *inovação* (frequência 42), *variação*(frequência 6), *estimação* (frequência 3), *aplicação* (frequência 2), *depreciação* (frequência 2), *maximização* (frequência 2), *investimento* (frequência 66), *financiamento* (frequência 39), *endividamento*(frequência 12), *faturamento* (frequência 8), dentre as que são mais utilizadas no campo da Economia.

Há de se ressaltar que o artigo tem como objetivo analisar as dificuldades das empresas brasileiras para financiar seus investimentos em capital físico e em inovação.

Segue o quadro 12 com o *rank*, a palavra, a frequência e o produto da ordem de série vezes a frequência, da Região de Transição de Goffman.

Rank	Word	Frequency	C		
1	De	351	351		
2	E	163	326		
3	A	121	363		
4	Em	111	444		
5	O	90	450		
6	Que	85	510		
7	Investimento	66	462		
8	Se	61	488		
9	Da	55	495		
10	Para	54	540		
11	Os	46	506		
12	Empresa	45	540		
13	Como	44	572		
14	Do	43	602		
15	Inovação	42	630		
16	Capital	42	672		
17	Com	41	697		
18	É	40	720		
19	Financiamento	39	741		
20	Na	37	740		
21	No	33	693		
22	Mercado	32	704		
23	Físico	31	713		
24	Uma	29	696		
25	Um	28	700		
26	Por	28	728		
27	Externo	28	756		
28	Também	26	728		
29	As	24	696		
30	Pesquisa	23	690		
31	Entre	23	713		
				PONTO T DE GOFFMAN: 36 (35,91)	
				TOTAL DE PALAVRAS: 4414	
				I <sub>1</sub>	663
				I <sub>2</sub>	169
				I <sub>3</sub>	77
				I <sub>4</sub>	33
				I <sub>5</sub>	31

32	Empresas	22	704
33	Dificuldades	22	726
34	Ao	20	680
35	2003	20	700
36	Relação	19	684
37	Mais	19	703
38	À	19	722
39	2	19	741
40	Resultados	18	720

Quadro 12 Representação da Região de Transição de Goffman no artigo 5.

## Artigo 6

VASCONCELOS, Cláudio Roberto Fóffano; VASCONCELOS, Silvinha Pinto. Medidas “antidumping” e resultados colativos: o caso do PEBDL na economia brasileira. *Nova Economia*, Belo Horizonte, v.15, n.3, p.117-141, set./dez. de 2005.  
Disponível em: <http://www.scielo.br/pdf/neco/v15n3/v15n3a05.pdf>  
Acesso em: 27/01/2010

O artigo 6 reúne um total de 5775 palavras, contendo 1340 palavras distintas que equivalem a aproximadamente 23 (23,20) % do total de palavras. A frequência média por palavra, isto é, o índice médio de repetição de cada palavra no *corpus* analisado é de aproximadamente 4 (4,30).

A análise das estruturas léxico-morfológicas do artigo 6 revelou a superioridade no padrão de formação de nominalizações deverbais com o sufixo *-ção* sobre o número de ocorrências de cada padrão investigado. Com o sufixo *-ção*, foram identificadas 61 bases distintas que ocorreram 274 vezes, com a frequência média aproximada igual a 4 (4,49), correspondendo a cerca de 67 (66,99) % do total de ocorrências (409) dos diferentes sufixos e a cerca de 64 (64,21) % do total de bases distintas (95).

O sufixo *-mento* foi ligado a 15 bases distintas que ocorreram 78 vezes, com a frequência média aproximada igual a 5 (5,20), correspondendo a cerca de 19 (19,07) % do total de ocorrências (409) e a cerca de 16 (15,78) % do total de bases distintas (95). O sufixo *-agem* foi ligado a 3 bases distintas que ocorreram 10 vezes, com a frequência média aproximada igual a 3 (3,33), correspondendo a cerca de 2 (2,44) % do total de ocorrências (409) e a cerca de 3 (3,15) % do total de bases distintas (95). O sufixo *-ncia*, com 12 bases distintas, foi empregado 20 vezes, com frequência média 1,66, correspondendo a cerca de 5 (4,88) % do total de ocorrências (409) e a cerca de 13 (12,63) % do total de bases distintas (95). O sufixo *-da* com 4 bases distintas, foi empregado 27 vezes, com frequência média 6,75,

correspondendo a cerca de 7 (6,60) % do total de ocorrências (409) e a cerca de 4 (4,21) % do total de bases distintas (95).

Segue a Tabela 6, que sintetiza os sufixos nominalizadores no artigo 6 e suas frequências relativas de ocorrência, assim como o grau de produtividade dos padrões de formação de palavras estimados pela observação da ocorrência de bases/produtos distintos de nominalização. A tabela também demonstra a força lexical relativa dos distintos padrões de produção de nominalizações.

Tabela 6 Síntese da produtividade dos sufixos nominalizadores do artigo 6 e da frequência relativa de ocorrência.

Rank	Sufixos	Frequência Total por Sufixo	Frequência Relativa do Sufixo/ Frequência Total de Sufixos	Bases Distintas	Bases Distintas/ Total de Bases Distintas	Frequência Média das Bases Distintas
1	-ção	274	66,99%	61	64,21%	4,49
2	-mento	78	19,07%	15	15,78%	5,20
3	-agem	10	2,44%	3	3,15%	3,33
4	-ncia	20	4,88%	12	12,63%	1,66
5	-da	27	6,60%	4	4,21%	6,75

A seguir apresenta-se o Quadro 13 com a relação das nominalizações deverbais investigadas no artigo 6 e suas frequências relativas.

<b>-ÇÃO</b>			<b>-ÇÕES</b>		
1	investigação	57	1	importações	24
2	importação	20	2	investigações	18
3	aplicação	19	3	informações	7
4	redução	12	4	exportações	2
5	produção	10	5	alterações	1
6	integração	8	6	aplicações	1
7	utilização	7	7	combinações	1
8	participação	5	8	implicações	1
9	punição	5	9	negociações	1
10	correção	4	10	observações	1
11	geração	4	11	preocupações	1
12	imposição	4	<b>Totais</b>	<b>11</b>	<b>58</b>
13	competição	3			
14	sustentação	3			
15	verificação	3			
16	concentração	2			
17	contribuição	2			
18	coordenação	2			
19	correlação	2			

20	elevação	2
21	implicação	2
22	informação	2
23	modificação	2
24	proteção	2
25	reestruturação	2
26	regulamentação	2
27	adoção	1
28	aquisição	1
29	caracterização	1
30	classificação	1
31	comparação	1
32	comprovação	1
33	concepção	1
34	determinação	1
35	diferenciação	1
36	diminuição	1
37	especificação	1
38	evolução	1
39	fabricação	1
40	fixação	1
41	formalização	1
42	formulação	1
43	intenção	1
44	interrupção	1
45	introdução	1
46	limitação	1
47	manifestação	1
48	negociação	1
49	parametrização	1
50	promulgação	1
51	reformulação	1
52	regulação	1
53	relação	1
54	restrição	1
55	suposição	1
56	transformação	1
<b>Totais</b>	<b>56</b>	<b>216</b>
<b>-MENTO</b>		
1	arquivamento	36
2	encerramento	11
3	comportamento	9
4	crescimento	3
5	procedimento	3
6	desenvolvimento	2
7	entendimento	2
8	aperfeiçoamento	1
9	aprofundamento	1
10	cumprimento	1
11	deferimento	1
12	julgamento	1
13	prosseguimento	1

<b>-MENTOS</b>		
1	comportamentos	2
2	arquivamentos	1
3	procedimentos	1
4	revestimentos	1
<b>Totais</b>	<b>4</b>	<b>5</b>

14	questionamento	1			
<b>Totais</b>	<b>14</b>	<b>73</b>			
<b>-NCIA</b>			<b>-NCIAS</b>		
1	existência	6	1	audiências	1
2	tendência	3	2	evidências	1
3	ocorrência	2	<b>Totais</b>	<b>2</b>	<b>2</b>
4	ausência	1			
5	experiência	1			
6	importância	1			
7	incidência	1			
8	resistência	1			
9	significância	1			
10	transparência	1			
<b>Totais</b>	<b>10</b>	<b>18</b>	<b>-AGENS</b>		
<b>-AGEM</b>			1	defasagens	1
1	abordagem	5	<b>Totais</b>	<b>1</b>	<b>1</b>
2	defasagem	3			
3	modelagem	1			
<b>Totais</b>	<b>3</b>	<b>9</b>	<b>-DAS</b>		
<b>-DA</b>			1	medidas	11
1	demanda	9	2	vendas	1
2	medida	5	<b>Totais</b>	<b>2</b>	<b>12</b>
3	rodada	1			
<b>Totais</b>	<b>3</b>	<b>15</b>			

Quadro 13 Relação das nominalizações deverbais investigadas no artigo 6.

A análise bibliométrica do artigo indicou que, no *rank* 1, a preposição *de* foi empregada 431 vezes, equivalendo aproximadamente a 7 (7,46) % do total de palavras (5775). Em contrapartida, 762 itens ocorreram uma única vez ( $I_1 = 762$ ), equivalendo aproximadamente a 13 (13,19)% do total de palavras do artigo 7 e a cerca de 57 (56,86)% do total de palavras distintas (1340).

A aplicação da Primeira Lei de Zipf, considerando como palavras de alta frequência as que se situam na ordem de serie ( $r$ ) igual ou menor que 5 ( $r \leq 5$ ), apresentou desvios da média de 507, para mais, de 29,38 % ( $r=4$ ), de 11,43% ( $r = 5$ ), e, para menos, de 14,99% ( $r=1$ ), de 23,47% ( $r = 2$ ) e de 2,36% ( $r=3$ ). A aplicação da Segunda Lei de Zipf ( $n \leq 5$ ), onde  $n = f$ , apresentou desvios, para menos, de cerca de 15,35% ( $n = 2$ ), 22,83% ( $n = 3$ ), 25,19% ( $n = 4$ ) e 33,07% ( $n = 5$ ) (Tabela 10).

O Ponto T de Goffman localiza-se aproximadamente na frequência 39 (38,54). Se projetarmos para baixo o mesmo número de palavras que ocorre acima da frequência 39, um total de 44 palavras é identificado. Nessa região, são 28 as ocorrências de preposições, artigos,

conjunções, pronomes, símbolos matemáticos e verbos auxiliares. Essas palavras foram excluídas na medida em que o objetivo do estudo é a investigação temática na Região de Transição de Goffman. Assim, foram identificadas as seguintes palavras: *investigação*, *antidumping*, *dumping*, *arquivamento*, *período*, *países*, *variáveis*, *preço*, *quantidade*, *importações*, *resultados*, *casos*, *acordo*, *importação*, *firmas* e a sigla *PEBDL*, que são de alto teor semântico em relação ao artigo em análise e, dentre as quais *investigação*, *antidumping*, *dumping*, *arquivamento*, *PEBDL*, *acordo* e *firmas*, integram o título e/ou o resumo e/ou as palavras-chave.

Logo após são identificadas as palavras *petição* (frequência 19), *aplicação* (frequência 19), *preços* (frequência 18), *investigações* (frequência 18), *comércio* (frequência 17), frequentes em textos acadêmicos sobre Economia. As nominalizações deverbais relevantes do ponto de vista da indexação foram: *investigação* (frequência 57), *importação* (frequência 20), *aplicação* (frequência 19), *redução* (frequência 12), *punição* (frequência 5), *correção* (frequência 4), *imposição* (frequência 4), *competição* (frequência 3), *sustentação* (frequência 3), *regulamentação* (frequência 2), *arquivamento* (frequência 36), *encerramento* (frequência 11), como exemplo de formas empregadas em artigos na área de Economia.

Há de se ressaltar que o artigo tem como objetivo examinar o pedido de arquivamento da investigação de *dumping* do produto PEBDL (polietileno de baixa densidade linear).

Segue o Quadro 14 com o *rank*, a palavra, a frequência e o produto da ordem de série vezes a frequência, da região de transição de Goffman.

Rank	Word	frequency	C		
1	De	431	431		
2	A	194	388		
3	O	165	495		
4	E	164	656		
5	Da	113	565		
6	Para	91	546		
7	Do	91	637		
8	Que	87	696		
9	Se	70	630		
10	Em	64	640		
11	Dos	63	693		
12	Das	63	756		
13	Os	59	767		
14	Investigação	57	798		
				PONTO T DE GOFFMAN: 39 (38,54)	
				TOTAL DE PALAVRAS: 5775	
				I <sub>1</sub>	762
				I <sub>2</sub>	215
				I <sub>3</sub>	98
				I <sub>4</sub>	57
				I <sub>5</sub>	34

15	No	55	825
16	As	48	768
17	Um	47	799
18	Na	43	774
19	Antidumping	43	817
20	Dumping	40	800
21	Como	40	840
22	É	40	880
23	Por	36	828
24	Arquivamento	36	864
25	Não	35	875
26	Ln	33	858
27	Período	31	837
28	Com	31	868
29	Entre	27	783
30	Países	26	780
31	Variáveis	25	775
32	Preço	25	800
33	Quantidade	24	792
34	Importações	24	816
35	Resultados	23	805
36	PEBDL	23	828
37	Ou	23	851
38	Casos	23	874
39	Acordo	23	897
40	Foi	21	840
41	Ao	21	861
42	Importação	20	840
43	Firmas	20	860
44	São	19	836

Quadro 14 Representação da Região de Transição de Goffman no artigo 6.

Nos textos examinados sobre a indústria de vinhos e sobre a Economia, foram selecionadas algumas estruturas nominalizadas mais frequentes, de acordo com a descrição  $[X] v \rightarrow [ [X] v -ção ] N$  e  $[X] v \rightarrow [ [X] v -mento$  e investigada a regularidade dessas formas. A sistematicidade das regras, como previsto pelo quadro teórico da Teoria Lexical de Chomsky (1965, 1970), Aronoff (1976), Jackendoff (1975), Bybee (1988), dentre outros, e pelos resultados voltados para o português sobretudo verificados em Basílio (1980), é atestada novamente em textos acadêmicos apontando uma contribuição importante para a teoria e prática da indexação, na Ciência da Informação. A título de exemplo, foram analisadas quatro nominalizações extraídas dos textos da amostra que ilustram também possível correspondência entre estruturas nominais e verbais com significados assemelhados.



(1) Com o sufixo nominalizador *-ção*, foi selecionada a palavra *vinificação* (artigo 1, frequência 17), que se refere à designação do processo de vinificar. Segue a correlação entre a forma nominalizada e o verbo correspondente nos trechos a seguir:

“[...] este trabalho teve o objetivo de avaliar as diferenças na evolução do PH, acidez total, ácido tartárico e K na *vinificação* de uvas tintas de três regiões vitícolas do Rio Grande do Sul.”

[...] este trabalho teve o objetivo de avaliar as diferenças na evolução do PH, acidez total, ácido tartárico e K enquanto as uvas tintas de três regiões vitícolas do Rio Grande do Sul *são vinificadas*.

(2) Com o sufixo nominalizador *-mento*, destacou-se a forma *esmagamento* (artigo 1, frequência 3) que reporta ao processo de esmagar. Observa-se, nos trechos abaixo, a equivalência com o verbo correspondente:

“A retirada das amostras para análise foi feita na seguinte sequência: 1. imediatamente após o *esmagamento* da uva; [...]”

A retirada das amostras para análise foi feita na seguinte sequência: 1. imediatamente após a uva *ser esmagada* [...]

(3) Com o sufixo nominalizador *-ção*, considerou-se também a forma *elaboração* (artigo 3, frequência 6). Sua função mostra-se correlata à do verbo correspondente, conforme demonstrado nos trechos a seguir.

“[...] realizou-se o presente trabalho para avaliar o potencial dessa cultivar para *elaboração* de vinho tinto.”

[...] realizou-se o presente trabalho para avaliar o potencial dessa cultivar para que o vinho tinto *seja elaborado*.

(4) Outro exemplo com *-mento* é o de *arquivamento* (artigo 6, frequência 36), que se refere ao processo de arquivar e que apresenta a seguinte correlação da nominalização com o verbo correspondente:

“[...] objetiva-se verificar empiricamente a hipótese de que pedidos de *arquivamentos* de petições de *dumping* servem como mecanismo de sustentação de resultados colusivos no Brasil.”

[...] objetiva-se verificar empiricamente a hipótese de que pedidos para que *sejam arquivadas* as petições de *dumping* servem como mecanismo de sustentação de resultados colusivos no Brasil.

Os exemplos corroboram a hipótese comprovada em Albino (1992, p. 36), segundo a qual a produção de estruturas nominais é mais frequente com o propósito de “nomear, de fazer referência a atos e processos como algo abstrato ou como resultado concreto”. Basílio (2007) esclarece que o uso em terminologias científicas e tecnológicas é mais uma vertente da função denominadora das nominalizações deverbais.

## 6 SÍNTESE DOS RESULTADOS

A investigação do grau relativo de produtividade de nominalizações deverbais nos artigos em foco e o cálculo da frequência relativa de cada padrão de formação de palavras indicaram os seguintes resultados:

(1) apenas o sufixo *-ção*, corresponde a cerca de 88% (T1), 72% (T2), 68% (T3), 83% (T4), 68% (T5) e 64% (T6) do total de bases distintas. Com o sufixo *-ção* foram identificadas 36 (T1), 47 (T2), 46 (T3), 33 (T4), 47 (T5) e 61 (T6) bases distintas que ocorreram 98 (T1), 112 (T2), 148 (T3), 157 (T4), 162 (T5) e 274 (T6) vezes, com a frequência média aproximada igual a 2,72 (T1), 2,38 (T2), 3,21 (T3), 4,75 (T4), 3,44 (T5) e 4,49 (T6);

(2) o sufixo *-mento* corresponde a cerca de 5% (T1), 14% (T2), 15% (T3), 8% (T4), 17% (T5) e 16% (T6) do total de bases distintas. Com o sufixo *-mento* foram identificadas 2 (T1), 9 (T2), 10 (T3), 3 (T4), 12 (T5) e 15 (T6) bases distintas que ocorreram 3 (T1), 11 (T2), 15 (T3), 16 (T4), 151 (T5) e 78 (T6) vezes, com a frequência média aproximada igual a 1,5 (T1), 1,22 (T2), 1,50 (T3), 5,33 (T4), 12,58 (T5) e 5 (T6);

(3) o sufixo *-ncia* corresponde a cerca de 5% (T1), 6% (T2), 9% (T3), 8% (T4), 14% (T5) e 13% (T6) do total de bases distintas. Com o sufixo *-ncia* foram identificadas 2 (T1), 4 (T2), 6 (T3), 3 (T4), 10 (T5) e 12 (T6) bases distintas que ocorreram 2 (T1), 6 (T2), 8 (T3), 3 (T4), 26 (T5) e 20 (T6) vezes, com a frequência média aproximada igual a 1 (T1), 1,50 (T2), 1,33 (T3), 1 (T4), 2,6 (T5) e 2 (T6);

(4) o sufixo *-agem* corresponde a cerca de 0% (T1), 5% (T2), 6% (T3), 3% (T4), 0% (T5) e 3% (T6) do total de bases distintas. Com o sufixo *-agem* foram identificadas 0 (T1), 3 (T2), 4 (T3), 1 (T4), 0 (T5) e 3 (T6) bases distintas que ocorreram 0 (T1), 6 (T2), 5 (T3), 1 (T4), 0 (T5) e 10 (T6) vezes, com a frequência média aproximada igual a 0 (T1), 2 (T2), 1,25 (T3), 1 (T4), 0 (T5) e 3 (T6);

(5) o sufixo *-da* corresponde a cerca de 2% (T1), 3% (T2), 3% (T3), 0% (T4), 0% (T5) e 4% (T6) do total de bases distintas. Com o sufixo *-da* foram identificadas 1 (T1), 2 (T2), 2 (T3), 0 (T4), 0 (T5) e 4 (T6) bases distintas que ocorreram 1 (T1), 4 (T2), 5 (T3), 0 (T4), 0 (T5) e

27 (T6) vezes, com a frequência média aproximada igual a 1 (T1), 2 (T2), 2,5 (T3), 0 (T4), 0 (T5) e 7 (T6);

Para efeito de análise comparativa, seguem abaixo, nas tabelas 7 e 8, os resultados acima apontados que demonstram a gradação da produtividade dos sufixos nominalizadores nos artigos 1, 2, 3, 4, 5 e 6, bem como suas frequências relativas de ocorrência.

Tabela 7 Síntese da produtividade dos sufixos nominalizadores em análise nos artigos 1, 2, 3 e 4, bem como suas frequências relativas de ocorrência.

Sufixo	Frequência Total por Sufixo				Frequência Relativa do Sufixo/ Frequência Total de Sufixos				Bases Distintas				Bases Distintas/ Total de Bases Distintas				Frequência Média das Bases Distintas			
	T1	T2	T3	T4	T1	T2	T3	T4	T1	T2	T3	T4	T1	T2	T3	T4	T1	T2	T3	T4
<b>-cão</b>	98	112	148	157	94,23	80,57	81,76	88,70	36	47	46	33	87,80	72,30	67,64	82,50	2,72	2,38	3,21	4,75
<b>-mento</b>	3	11	15	16	2,88	7,91	8,28	9,03	2	9	10	3	4,87	13,84	14,70	7,50	1,50	1,22	1,50	5,33
<b>-agem</b>	0	6	5	1	0	4,31	4,41	0,56	0	3	4	1	0	4,61	5,88	2,50	0	2	1,25	1
<b>-ncia</b>	2	6	8	3	1,92	4,31	2,76	1,69	2	4	6	3	4,87	6,15	8,82	7,50	1	1,50	1,33	1
<b>-da</b>	1	4	5	0	0,96	2,87	2,76	0	1	2	2	0	2,43	3,07	2,94	0	1	2	2,50	0
<b>Totais</b>	104	139	181	177	100	100	100	100	41	65	68	40	100	100	100	100	6,22	9,10	9,79	12,08

Tabela 8 Síntese da produtividade dos sufixos nominalizadores em análise nos artigos 5 e 6, bem como suas frequências relativas de ocorrência.

Sufixo	Frequência Total por Sufixo		Frequência Relativa do Sufixo/ Frequência Total de Sufixos		Bases Distintas		Bases Distintas/ Total de Bases Distintas		Frequência Média das Bases Distintas	
	T5	T6	T5	T6	T5	T6	T5	T6	T5	T6
<b>-cão</b>	162	274	47,78	66,99	47	61	68,11	64,21	3,44	4,49
<b>-mento</b>	151	78	44,54	19,07	12	15	17,39	15,78	12,58	5,20
<b>-agem</b>	0	10	0	2,44	0	3	0	3,15	0	3,33
<b>-ncia</b>	26	20	7,66	4,88	10	12	14,49	12,63	2,60	1,66
<b>-da</b>	0	27	0	6,60	0	4	0	4,21	0	6,75
<b>Totais</b>	339	409	100	100	69	95	100	100	18,62	21,43

A síntese dos resultados, incluídos nas tabelas 7 e 8, evidencia alto grau de produtividade do padrão  $[X] v \rightarrow [ [X] v -\text{ção} ] N$  nos textos em análise. Nas tabelas, em contrapartida, verifica-se baixo grau de produtividade do padrão  $[X] v \rightarrow [ [X] v -\text{da} ] N$ , tendendo a zero em alguns artigos. Observa-se que esses dois padrões descritos de nominalização parecem situar-se em pontos polarizados em um *continuum* projetado, marcado por um gradiente de produtividade de formação de nominalizações deverbais.

Constata-se na pesquisa de tese que o gênero artigo de periódico científico prevê frequência alta de produtividade na ocorrência dos padrões de formação de nominalizações deverbais  $[X] v \rightarrow [ [X] v -\text{ção} ] N$  e  $[X] v \rightarrow [ [X] v -\text{mento} ] N$ . Em função das chances de emergência, tais formas bloqueiam e enfraquecem, conseqüentemente, o processamento dos demais padrões possíveis, conforme fica comprovado na pesquisa desenvolvida. Desse modo, atesta a relevância do conceito de força lexical postulado em Bybee (1988, p. 121). Nesse sentido, a frequência de uso dos itens influencia diretamente no grau de produtividade dos distintos padrões de nominalização deverbais investigados.

Além disso, no gráfico 1, foram plotados os resultados indicadores dos graus distintos de produtividades das cinco Regras de Formações de Palavras investigadas nos artigos 1, 2, 3, 4, 5 e 6, que constituem a amostra desta tese.

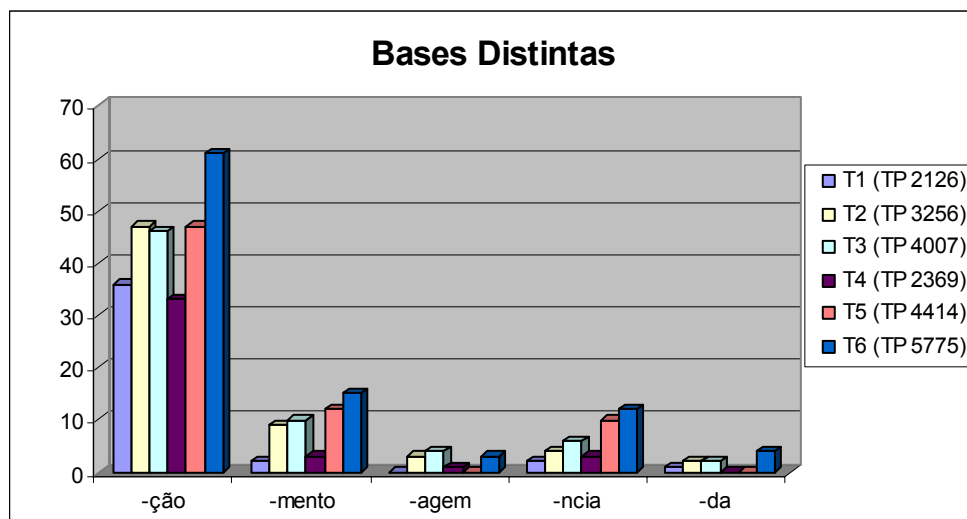


Gráfico 1 Representação espacial dos graus distintos de produtividade das Regras de Formações de Palavras (RFP) nos textos (T) analisados, com o número total de palavras (TP) de cada artigo.

Para efeito de indexação, o gráfico 2 representa espacialmente resultados que indicam a frequência total de ocorrência de cada sufixo nominalizador analisado na pesquisa, nos artigos 1, 2, 3, 4, 5 e 6, com o número total de palavras de cada artigo.

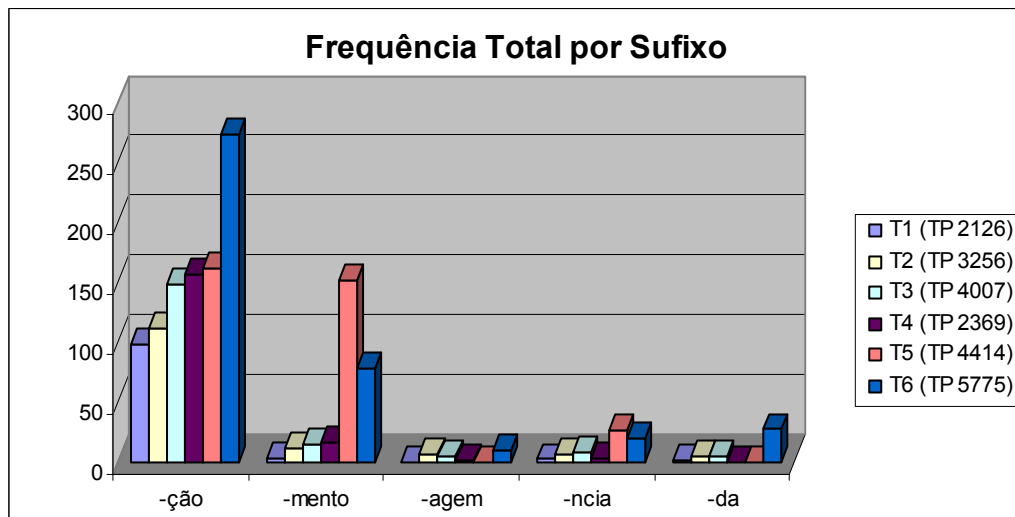


Gráfico 2 Representação espacial da frequência de ocorrência de cada sufixo nominalizador deverbal analisado, com o número total de palavras de cada artigo.

No que se refere à análise bibliométrica, as tabelas 9 e 10 sintetizam os resultados verificados na aplicação da Segunda Lei de Zipf nos seis artigos investigados.

Tabela 9 Síntese dos resultados verificados na aplicação da Segunda Lei de Zipf aos artigos sobre a indústria de vinho

N ≤ 5	Valor Observado				Valor Computado				Desvio			
	Textos				2ª Lei de Zipf				Valor Observado/ Valor Computado (%)			
	T1	T2	T3	T4	T1	T2	T3	T4	T1	T2	T3	T4
I <sub>1</sub>	380	492	593	380	-	-	-	-	-	-	-	-
I <sub>2</sub>	89	131	148	91	126,66	164	197,66	126,66	29,21	20,12	25,12	28,15
I <sub>3</sub>	30	76	79	43	63,33	82	98,83	63,33	52,62	7,31	20,06	32,10
I <sub>4</sub>	15	31	43	29	38	49,2	59,30	38	60,52	36,99	29,17	23,68
I <sub>5</sub>	17	21	39	15	25,33	32,8	39,53	25,33	32,88	35,97	1,34	40,78

Tabela 10 Síntese dos resultados verificados na aplicação da Segunda Lei de Zipf aos artigos sobre a Economia

N ≤ 5	Valor Observado		Valor Computado		Desvio	
	Textos		2ª Lei de Zipf		Valor Observado/ Valor Computado (%)	
	T5	T6	T5	T6	T5	T6
I <sub>1</sub>	663	762	-	-	-	-
I <sub>2</sub>	169	215	221	254	23,52	15,35
I <sub>3</sub>	77	98	110,5	127	30,31	22,83
I <sub>4</sub>	33	57	66,3	76,2	50,22	25,19
I <sub>5</sub>	31	34	44,2	50,8	29,86	33,07

Observa-se, ainda, do ponto de vista bibliométrico, que, em relação aos seis textos da amostra, o índice médio de repetição de cada palavra é de aproximadamente 4.

Finalmente, em relação à regularidade das estruturas nominalizadas mais frequentes, de acordo com a descrição  $[X] v \rightarrow [ [X] v -\text{ção} ] N$  e  $[X] v \rightarrow [ [X] v -\text{mento} ]$ , analisaram-se duas nominalizações em *-ção* e duas nominalizações em *-mento*, extraídas dos textos da amostra. Os resultados obtidos indicaram, nos quatro exemplos selecionados, a regularidade da regra. A sistematicidade da regras se deve à designação de processos, atos, produtos e conceitos.

## 7 CONCLUSÕES E SUGESTÕES PARA PESQUISAS FUTURAS

Neste capítulo, pontuam-se as principais conclusões, consideram-se algumas reflexões e sugerem-se direções possíveis que permitem dar continuidade à pesquisa de doutoramento aqui empreendida.

A análise das conexões lexicais, formuladas nas regras, e suas recorrências na formação de nominalizações nos artigos analisados permitem afirmar que o modelo morfológico  $[X] v \rightarrow [ [X] v -\text{ção} ] N$  tende a intensificar ainda mais sua força lexical. O grau de produtividade da regra funciona como estratégia de bloqueio para outros padrões de nominalização. Operando o bloqueio, a tendência aponta ainda maior frequência de sufixos em *-ção*. Constatou-se em contrapartida que os padrões de formação de nominalizações deverbais  $[X] v \rightarrow [ [X] v -\text{ncia} ] N$ ,  $[X] v \rightarrow [ [X] v -\text{agem} ] N$ ,  $[X] v \rightarrow [ [X] v -\text{da} ] N$  são improdutivo no domínio de discurso em foco. O total de nominalizações deverbais segundo o padrão  $[X] v \rightarrow [ [X] v -\text{mento} ] N$  apresentou grau de produtividade menor comparado ao grau de produtividade do padrão de formação de nominalizações deverbais  $[X] v \rightarrow [ [X] v -\text{ção} ]$ .

No entanto, cotejados os resultados da análise dos artigos sobre indústria de vinhos com os obtidos pela análise dos artigos sobre Economia, constata-se diferença significativa na relação entre o número de bases distintas e de frequências relativas de uso, associado aos sufixos nominalizadores *-ção* e *-mento*. Os textos sobre Economia apresentaram menor diferença entre as frequências relativas de uso dos sufixos nominalizadores *-ção* (162) e *-mento* (151), apesar de terem tido ocorrência relevante em relação às bases conectadas aos sufixos *-ção* (47) e *-mento* (12) na formação de nominalizações deverbais.

Para os textos analisados, a aplicação das Leis de Zipf não se verificou satisfatoriamente, conforme apresentado nos quadros 4, 6, 8, 10, 12 e 14 e nas tabelas 9 e 10. Entretanto, observa-se que os baixos *ranks* (altas frequências) estão associados às palavras de baixo conteúdo semântico como artigos, preposições, conjunções, verbos auxiliares etc. como mencionado em Rouault (1987), Eggins (2004), dentre outros.

A aplicação da fórmula do Ponto de T e do mecanismo de delimitação da Região de Transição de Goffman revelou-se satisfatória. Os resultados obtidos pela observação dos itens lexicais



frequentes na Região de Transição de Goffman, em 100% dos textos, indicaram a ocorrência de palavras de alta carga semântica. Ou seja, para os textos analisados, a região concentra frequências associadas a palavras de alto conteúdo semântico. Essa assertiva é reforçada pelo fato dos termos incluírem-se nos títulos e/ou nos resumos e/ou nas palavras-chave dos seis textos analisados. A investigação das frequências de uso das nominalizações deverbais na Região de Transição de Goffman revelou-se satisfatória em 100% da amostra (quadros 4, 6, 8, 10, 12 e 14). Contudo, os resultados obtidos no processamento dos textos 3, 4, 5 e 6 (quadros 8 e 10) sobre a indústria de vinhos e a Economia verificaram-se ainda mais satisfatórios, principalmente na análise dos textos sobre a Economia. (quadros 12 e 14).

Os achados de pesquisa demonstram a importância da análise morfológica da produtividade de nominalizações deverbais no gênero acadêmico, na busca de evidências empíricas, sobretudo para comprovar que o sufixo *-ção* nas nominalizações em artigos científicos apresenta alta densidade de informação. Além disso, as nominalizações encontram-se no campo semântico dos domínios especializados do conhecimento científico e tecnológico. Nesse sentido, procede a postulação de gradiente de probabilidade de ocorrência de determinados padrões de construções e, também, da frequência de uso de certas regras, nos textos analisados, marcados por graus relativos de produtividade de padrões de nominalização verbal. Numa perspectiva léxico-morfológico, a formação de nominalização formalmente representada por  $[X] v \rightarrow [ [X] v -ção ] N$  é predominante no campo semântico da indústria de vinhos, apontando os caminhos mais apropriados para os processos de indexação temática em sistemas eletrônicos de linguagem controladas, na Ciência da Informação.

Constata-se ainda que os processos analisados apresentam função denominadora predominante. A formação de nominalizações resulta na constituição de rótulos que têm uma representação conceptual nos campos de conhecimento.

Finalmente, os resultados confirmam a teoria de que o gênero acadêmico é caracterizado por alto grau de informatividade, de tal modo que a produtividade de nominalizações deverbais em artigos é esperada.

Especialmente no que se refere ao procedimento de delimitação da Região de Transição de Goffman, foi verificada a aplicabilidade do mecanismo, que indicou nessa região a ocorrência de termos com alto teor semântico em relação aos temas abordados nos artigos e, portanto,

indicados como palavras-chave no processo de indexação da informação. Na área de Economia os resultados se mostraram ainda mais acentuados, porque apresentaram maior número de ocorrência de nominalizações deverbais relevantes, em relação ao conteúdo do texto. Do ponto de vista bibliométrico, as nominalizações deverbais de maior relevância ocorrem na Região de Transição de Goffman.

Há de se ressaltar que tentativas de delimitação da região de concentração de termos de indexação, calculando-se um desvio de 10% e de 15%, com base em estudos publicados na literatura sobre o tema, não se mostraram satisfatórias para os artigos em foco. Em contrapartida, a aplicação do procedimento comunicado informalmente por Goffmann à Miranda Pao, considerado o de maior relevância do ponto de vista teórico-metodológico, foi satisfatória para os artigos analisados nesta pesquisa de doutoramento.

Esta tese aponta indicadores relevantes que comprovam a necessidade de mais estudos assemelhados para o estabelecimento de padrões de produtividade de nominalizações deverbais em textos acadêmicos. O estudo das nominalizações deverbais contribui para a adequação do método bibliométrico semi-automático de indexação em análise, sobretudo no que se refere à identificação de bases e sufixos de alto conteúdo informativo. Sugere-se estender esta investigação para um maior número de textos sobre indústria de vinhos e sobre Economia, com o intuito de verificar a validação do método em análise. Nessa investigação, pode-se optar pela análise comparativa de gêneros acadêmicos distintos, tais como: artigos e resumos. É procedente também considerar o desenvolvimento de mecanismos de identificação de padrões de formação de nominalizações deverbais, sobretudo na região de concentração de palavras-chave, com base em um maior número de artigos em outras áreas do conhecimento.

## REFERÊNCIAS

ALBINO, José Mauro Branco. As condições de produção dos sufixos nominalizadores -ção e -mento no português escrito formal. 1993. 70 f. Dissertação (Mestrado em Linguística) – Programa de Pós-graduação em Letras, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 1993.

ARONOFF, Mark. **Word formation in generative grammar**. Cambridge, Massachusetts: the MIT Press, 1976. (Linguistic Inquiry Monographs, 1).

\_\_\_\_\_, Mark. **Morphology by itself: stems and inflectional classes**. Cambridge, Massachusetts: the MIT Press, 1994. (Linguistic Inquiry Monographs, 22).

Disponível em: [http://books.google.com.br/books?id=PFMH1qg-Z5kC&printsec=frontcover&source=gbs\\_navlinks\\_s#v=onepage&q=&f=false](http://books.google.com.br/books?id=PFMH1qg-Z5kC&printsec=frontcover&source=gbs_navlinks_s#v=onepage&q=&f=false)  
Acesso em: 27/06/2007.

ARONOFF, Mark; ANSHEN, Frank. Morphology and the Lexicon: lexicalization and productivity. 2001. In: SPENCER, Andrew; ZWICKY, Arnold M. (Ed.) **The Handbook of Morphology**. Oxford: Wiley-Blackwell, 2001.

Disponível em: [http://books.google.com.br/books?id=xrdbNnbufqAC&dq=The+Handbook+of+Morphology+Por+Andrew+ARONOFFSpencer,+Arnold+M.+Zwicky&printsec=frontcover&source=bn&hl=pt-BR&ei=NwMcStjCDOi\\_twfY\\_smODA&sa=X&oi=book\\_result&ct=result&resnum=4](http://books.google.com.br/books?id=xrdbNnbufqAC&dq=The+Handbook+of+Morphology+Por+Andrew+ARONOFFSpencer,+Arnold+M.+Zwicky&printsec=frontcover&source=bn&hl=pt-BR&ei=NwMcStjCDOi_twfY_smODA&sa=X&oi=book_result&ct=result&resnum=4)  
Acesso em: 26/05/2009.

BAKHTIN, Mikhail Mikhaïlovich et al. **Marxismo e filosofia da linguagem: problemas fundamentais do método sociológico na ciência da linguagem**. Tradução Michel Lahud e Yara Frateschi Vieira. São Paulo: Hucitec, 1979. 182 p.

BAKHTIN, Mikhail Mikhaïlovich. **Estética da criação verbal**. Tradução do francês por Maria Ermantina Galvão G. Pereira. São Paulo: Martins Fontes, 2000.

BASÍLIO, Margarida Maria de Paula. **Estruturas lexicais do português: uma abordagem gerativa**. Petrópolis: Editora Vozes, 1980. 128 p.

\_\_\_\_\_, Margarida Maria de Paula. Produtividade e função dos processos de formação de palavras no português falado. In: CONGRESSO INTERNACIONAL DA ASSOCIAÇÃO DE LINGÜÍSTICA E FILOGIA DA AMÉRICA LATINA, 9., 1990, Campinas. **Anais...** Campinas: ALFAL/UNICAMP, 1993.

\_\_\_\_\_, Margarida Maria de Paula. Formação e Uso da Nominalização Deverbal Sufixal no Português Falado. In: A.T. Castilho e M. Basilio (orgs.) **Gramática do Português Falado Volume IV- Estudos Descritivos**. São Paulo/Campinas: FAPESP/ UNICAMP: 23-33, 1996.

\_\_\_\_\_, Margarida Maria de Paula. A Morfologia no Brasil: indicadores e questões, **D.E.L.T.A.**, São Paulo, v.15, special issue, 1999. p. 53-70.  
Disponível em: <http://www.scielo.br/pdf/delta/v15nspe/4011.pdf>  
Acesso em: 26/05/2009.

\_\_\_\_\_, Margarida Maria de Paula. **Formação e classes de palavras no português do Brasil**. Rio de Janeiro: Editora Contexto, 2004a.  
Disponível em: [http://books.google.com.br/books?id=Bbi\\_m3TKpu8C&pg=PP3&dq=Forma%C3%A7%C3%A3o+e+classes+de+palavras+no+portugu%C3%AAs+do+Brasil](http://books.google.com.br/books?id=Bbi_m3TKpu8C&pg=PP3&dq=Forma%C3%A7%C3%A3o+e+classes+de+palavras+no+portugu%C3%AAs+do+Brasil)  
Acesso em: 26/05/2009.

\_\_\_\_\_, Margarida Maria de Paula. Polissemia sistemática em substantivos deverbais. **Ilha do Desterro**, Florianópolis, n. 47 p.49- 71, jul./dez. 2004b.  
Disponível em: <http://www.periodicos.ufsc.br/index.php/desterro/article/viewFile/7347/6769>  
Acesso em; 27/01/2010.

\_\_\_\_\_, Margarida Maria de Paula. **Teoria lexical**. 8. ed. revista e atualizada. São Paulo: Ática, 2007. 104 p. (Série Princípios, 88).

\_\_\_\_\_, Margarida Maria de Paula. Interface Linguística e Ciência da Informação: potencialidade na análise de estruturas lexicais. In: REUNIÃO BRASILEIRA DE CIÊNCIA DA INFORMAÇÃO, 2., 1979, Rio de Janeiro. **Trabalhos apresentados...** Rio de Janeiro: IBICT, 1979a.

BASÍLIO, Margarida Maria de Paula; BRAGA, Lilian Maria; PIEROTTI, Maria de Lourdes Carvalho. Estrutura de textos científicos em língua portuguesa: estudo bibliométrico-linguístico. In: REUNIÃO BRASILEIRA DE CIÊNCIA DA INFORMAÇÃO, 2., 1979, Rio de Janeiro. **Anais...** Rio de Janeiro: IBICT, 1979b.

BAZERMAN, Charles. **Gêneros textuais, tipificação e interação**. São Paulo: Editora Cortez, 2006.

\_\_\_\_\_, Charles. Atos de fala, gêneros textuais e sistemas de atividades: como os textos organizam atividades e pessoas. In: BAZERMAN, Charles. **Gêneros textuais, tipificação e interação**. São Paulo: Editora Cortez, 2006a. cap. 1. p. 19-46.

\_\_\_\_\_, Charles. Cartas e a base social de gêneros diferenciais. In: BAZERMAN, Charles. **Gêneros textuais, tipificação e interação**. São Paulo: Editora Cortez, 2006b. cap. 4. p. 83-99.

\_\_\_\_\_, Charles. Reporting the experiment: the changing account of scientific doings in the philosophical transactions of the royal society, 1665-1800. In: BAZERMAN, Charles. **Shaping Written knowledge: the genre and activity of the experimental article in science**. Madison, Wisconsin: The University of Wisconsin Press, 1988.  
Disponível em: [http://wac.colostate.edu/books/bazerman\\_shaping/chapter3.pdf](http://wac.colostate.edu/books/bazerman_shaping/chapter3.pdf)  
Acesso em: 01/06/2009.

BERNERS-LEE, Tim; HENDLER, James; LASSILA, Ora. The semantic web: a new form of Web content that is meaningful to computers will unleash a revolution of new possibilities. **Scientific American**, New York, no. 5, May 2001. Disponível em:  
[http://www.ryerson.ca/~dgrimsha/courses/cps720\\_02/resources/Scientific%20American%20The%20Semantic%20Web.htm](http://www.ryerson.ca/~dgrimsha/courses/cps720_02/resources/Scientific%20American%20The%20Semantic%20Web.htm)  
Acesso em: 02 dezembro de 2009.

BHATIA, Vijay Kumar. **Worlds of written discourse**. London: Continuum International Publishing Group, 2004.  
Disponível em:  
[http://books.google.com.br/books?id=\\_A8Hf5BHO\\_QC&pg=PA157&dq=BHATIA+\(2004\).+Words+of+Written+Discourse](http://books.google.com.br/books?id=_A8Hf5BHO_QC&pg=PA157&dq=BHATIA+(2004).+Words+of+Written+Discourse)  
Acesso em: 26/04/2009.

BLOOMFIELD, Leonard. A set of postulates for the science of language. **Language**, Chicago, v. 2, no. 3, p. 153-164, sep. 1926.  
Disponível em: <http://www.jstor.org/stable/pdfplus/408741.pdf>  
Acesso em: 04/12/2009

BOOTH, Andrew D. A "Law" of occurrences for words of low frequency, **Information and Control**, [s.l.], v. 10, n.4, p. 386-393, April 1967.

BRAGA, Gilda Maria. A representação da informação na reconstrução do texto. In: COLÓQUIO SOBRE ORGANIZAÇÃO DO CONHECIMENTO E REPRESENTAÇÃO DA INFORMAÇÃO: enfoques e perspectivas. **Resumos...** Rio de Janeiro: IBICT/CNPq, 1996.

BRAUN, Tibor; SCHUBERT, András. A quantitative view on the coming of age of interdisciplinarity in the sciences 1980-1999, **Scientometrics**, Budapest, v. 58, no. 1, p. 183-189, sept. 2003.  
Disponível em: <http://www.springerlink.com/content/q80v88u45642662l/fulltext.pdf>  
Acesso em: 24/11/2009.

BYBEE, J. L. **Frequency of use and the organization of language**. New York: Oxford University Press, 2007.

Disponível em:  
[http://books.google.com.br/books?id=W20t\\_5AXeaYC&printsec=frontcover&dq=bybee+frequency+of+use&cd=1](http://books.google.com.br/books?id=W20t_5AXeaYC&printsec=frontcover&dq=bybee+frequency+of+use&cd=1)  
 Acesso em: 10/02/2010.

\_\_\_\_\_, J. L. Morphology as lexical organization. In: HAMMOND, Michael; NOONAN, Michael P. **Theoretical morphology: approaches in modern linguistics**. San Diego: Academic Press. 1988. p.119-141.

BYBEE, J. L.; HOPPER, P. **Frequency and the emergence of language structure**. Amsterdam: John Benjamins, 2001.

Disponível em:  
<http://books.google.com.br/books?id=6-bQUClxn1IC&pg=PA135&dq=e+books+Bybee+linguistic+Morphology:+A+Study+of+the+Relation+between+Meaning+and+Form&cd=1#v=onepage&q=zipf&f=false>  
 Acesso em: 10/02/2010.

CABRÉ, María Teresa. **La terminología: teoría, metodología, aplicaciones**. Barcelona: Empúries, 1993. 529 p.

CAMARA JR, Joaquim Matoso. **História e estrutura da língua portuguesa**. 4. ed. Rio de Janeiro: Editora Padrão, 1985.

CHOMSKY, N. Remarks on nominalization. In: JACOBS, R.; ROSENBAUM, P. (Orgs.). **Readings in English transformational grammar**. Waltham, MA: Ginn, 1970. 277 p.

\_\_\_\_\_, N. Remarks on nominalization. In: -----, **Studies on semantics in generative grammar**. 3. ed. Netherlands: Mouton Publishers, 1980. p.11-61.

Disponível em:  
<http://books.google.com.br/books?id=E75lKtZ76rgC&printsec=frontcover&dq=chomsky&lr=&cd=13#v=onepage&q&f=false>  
 Acesso em: 20/02/2010.

\_\_\_\_\_, N. **Aspects of the Theory of Syntax**. Cambridge: MIT Press, 1965.

Disponível em:  
<http://books.google.com.br/books?hl=pt-BR&lr=&id=u0ksbFqagU8C&oi=fnd&pg=PR5&dq=CHOMSKY,+N.+Aspects+of+the+Theory+of+Syntax.+Cambridge:+MIT+Press,+1965&ots=5LyDpw6kTA&sig=VmXKAd2UBhbfswPSi43uuFIz-EM#v=onepage&q&f=false>  
 Acesso em: 15/02/2010.

CHOMSKY, Noan; MILLER, George A. **L'analyse formelle des langues naturelles**. Paris: Gauthier-Villars. 1971. Trad. Original Handbook of Mathematical Psychology, Volume II, Chapitre 11 e 12.

Disponível em:

[http://books.google.com.br/books?id=0u\\_DyqJ15QkC&pg=PP4&dq=Mouton,+Gauthier-Villars+chomsky+miller+1971&cd=1#v=onepage&q&f=false](http://books.google.com.br/books?id=0u_DyqJ15QkC&pg=PP4&dq=Mouton,+Gauthier-Villars+chomsky+miller+1971&cd=1#v=onepage&q&f=false)

Acesso em: 15/02/2010.

CINTRA, M. et al. **Para entender as linguagens documentárias**. 2ª ed. revista e ampliada. São Paulo: Polis, 2002.

CRANE, Diana. **Invisible colleges: diffusion of knowlegde in scientific communities**. Chicago: The University of Chicago Press, 1988. 224 p.

CRUSE, A. D. The lexicon. In: ARONOFF, M.; REES-MILLER, J. (Ed.) **The handbook of linguistics**. [s.l.]: Wiley-Blackwell, 2003. cap. 10. p. 238.

Disponível em:

[http://books.google.com.br/books?id=r18LDiR11nYC&dq=The+handbook+of+linguistics.&printsec=frontcover&source=bn&hl=pt-BR&ei=\\_usjStO8OMyLtgfi8KToAg&sa=X&oi=book\\_result&ct=result&resnum=4](http://books.google.com.br/books?id=r18LDiR11nYC&dq=The+handbook+of+linguistics.&printsec=frontcover&source=bn&hl=pt-BR&ei=_usjStO8OMyLtgfi8KToAg&sa=X&oi=book_result&ct=result&resnum=4)

Acesso em: 27/04/2009.

CUNHA, Celso; CINTRA, Luís Filipe Lindley. **Nova gramática do português contemporâneo**. 2 ed. Rio de Janeiro: Nova Fronteira, 1985.

DUBOIS, Jean et al. **Dicionário de linguística**. Tradução de Frederico Pessoa de Barros et al. São Paulo: Editora Cultrix, 2007. 653 p.

EGGINS, Suzanne. **An introduction to systemic functional linguistics**. 2. ed. Nova York: Continuum International Publishing Group, 2004. 365 p.

Disponível

em:

[http://books.google.com.br/books?id=sS7UXugIIg8C&printsec=frontcover&source=gbs\\_navlinks\\_s#v=onepage&q=&f=false](http://books.google.com.br/books?id=sS7UXugIIg8C&printsec=frontcover&source=gbs_navlinks_s#v=onepage&q=&f=false)

Acesso em: 21/02/2010.

FUJITA, Maria Ângela Spotti Lopes. **Linguagem documentária em odontologia: uma aplicação do sistema de indexação PRECIS**. v. 1. São Paulo: USP/Escola de Comunicações e Artes, 1992.

Disponível em:

[http://www.marilia.unesp.br/Home/Pos-Graduacao/CienciadaInformacao/Dissertacoes/fujita\\_msl\\_dr\\_mar.pdf](http://www.marilia.unesp.br/Home/Pos-Graduacao/CienciadaInformacao/Dissertacoes/fujita_msl_dr_mar.pdf)

Acesso em: 25/01/2010.

GIL, Isabel Teresa Morais. Algumas considerações sobre línguas de especialidade e seus processos lexicogénicos, **Máthesis**, [s.l.], n. 12, p. 113-130, 2003.

Disponível em: [http://www4.crb.ucp.pt/biblioteca/Mathesis/Mat12/Mathesis12\\_113.pdf](http://www4.crb.ucp.pt/biblioteca/Mathesis/Mat12/Mathesis12_113.pdf)  
Acesso em: 27/04/2009.

GUEDES, V.; BORSCHIVER, S. Bibliometria: uma ferramenta estatística para a gestão da informação e do conhecimento, em sistemas de informação, de comunicação e de avaliação científica e tecnológica. In: ENCONTRO NACIONAL DE CIÊNCIA DA INFORMAÇÃO, 6, 2005, Salvador. *Proceedings...* Salvador: UFBA, 2005.

Disponível em: [http://www.cinform.ufba.br/vi\\_anais/docs/VaniaLSGuedes.pdf](http://www.cinform.ufba.br/vi_anais/docs/VaniaLSGuedes.pdf)  
Acesso em: 27/04/2009.

HALLIDAY, Michael Alexander Kirkwood. **An introduction to functional grammar**. 2ª ed. London: Edward Arnold, 1994. 434 p.

HJØRLAND, Birger. Domain analysis in information science: eleven approaches - traditional as well as innovative, **Journal of Documentation**, v. 58, n. 4, p. 422-462, 2002.

Disponível em:

<http://www.emeraldinsight.com/Insight/viewPDF.jsp?contentType=Article&Filename=html/Output/Published/EmeraldFullTextArticle/Pdf/2780580404.pdf>

Acesso em: 08/06/2009.

\_\_\_\_\_, Birger. Domain analysis: a socio-cognitive orientation for information science research. **Bulletin of American Society of Information Science and Technology**, v. 30, n. 3, feb./mar, 2004.

Disponível em: <http://www.asis.org/Bulletin/Feb-04/hjorland.html>

Acesso em: 08/06/2009.

\_\_\_\_\_, Birger. Toward a new horizon in information science: Domain-analysis, **Journal of the American Society for Information Science**, v. 48, n. 6, p. 400-425, 1995.

HREBICEK, Ludek. Zipf's law and text, **Glottometrics**, Czech Republic, v. 3, n. 22, p. 23-38, 2002.

HUSTON, S. Evaluation and ideology in scientific writing. In GHADESSY, M. (Ed.). **Register analysis: theory and practice**. London: Pinter Publishers, 1994.

HYLAND, Ken. **Academic discourse: english in a global context**. New York: Continuum International Publishing Group, 2009. 215p.



\_\_\_\_\_, Ken. **Disciplinary discourses**: social interactions in academic writing. Ann Arbor: The University of Michigan Press, 2004. 211p.

JACKENDOFF, Ray. Morphological and Semantic Regularities in the Lexicon, **Language**, v. 51, no. 3, p. 639-671, sept. 1975.

Disponível em: <http://www.jstor.org/stable/412891>

Acesso em: 25/03/2010.

JOHNSON, Hugh. **A História do Vinho**. Tradução de Hildegard Feist. São Paulo: Companhia das Letras, 1999. 546 p.

KUHN, Thomas S. **A estrutura das revoluções científicas**. São Paulo: Perspectiva, 2007. 260 p.

LUHN, H. P. A statistical approach to mechanized encoding and searching of literary information. **IBM Journal of Research and Development**, New York, v. 1, n.4, p. 309- 317, Oct. 1957.

MACIAS-CHAPULA, Cesar A. O papel da Informetria e da Cienciometria e sua perspectiva nacional e internacional. **Ciência da Informação**, Brasília, v. 27, n. 2, p. 134-140, maio/ago. 1998.

Disponível em: <http://revista.ibict.br/index.php/ciinf/article/view/342/0>

Acesso em: 12/08/2009.

MAcNEIL, Karen. **The wine bible**. New York: Workman Publishing, 2001. 910 p.

MARCONDES, Carlos Henrique. Metadados: descrição e recuperação de informação na Web. In: MARCONDES, Carlos H. et al (Orgs.). **Bibliotecas digitais: saberes e práticas**. Salvador: Ed. UFBA, Brasília: IBICT, 2005. p. 97-114.

MATTHEWS, P. H. **Morphology**: an introduction to the theory of word-structure. Cambridge: Cambridge University Press, 1974.

MAUCERI, Christian. L'analyse de corpus textuels à l'aide de techniques linguistiques et statistiques au service de la construction de référentiels et du développement des systèmes d'indexation automatique. In : JOURNÉES D'ETUDES ORGANISÉES PAR LA SOCIÉTÉ FRANÇAISE DE BIBLIOMÉTRIE APPLIQUÉE AVEC LE CONCOURS. **Resumos...** Ilê Rouse: Société Française de Bibliométrie Appliquée, 1995.

Disponível em: <http://crrm.u-3mrs.fr/sfba/ile-rousse/1995/cd-rom.pdf>

Acesso em: 27/06/2007.

MEDEIROS, Alessandro Boechat de. A Molecada se Empanturrou de Bananada! - Considerações sobre as nominalizações em -ada não-eventivas. **Revista Virtual de Estudos da Linguagem**, v. 7, n. 12, p. 1-26, 2009.

Disponível em:

[http://www.revel.inf.br/site2007/\\_pdf/14/artigos/revel\\_12\\_a\\_molecada\\_se\\_empanturrou.pdf](http://www.revel.inf.br/site2007/_pdf/14/artigos/revel_12_a_molecada_se_empanturrou.pdf)

Acesso em: 27/03/2010.

MELLO, Loiva Maria Ribeiro de. Atuação do Brasil no mercado vitivinícola mundial - Panorama 2009, **Embrapa Uva e Vinho**, Rio Grande do Sul, 2010.

Disponível em: <http://www.cnpuv.embrapa.br/publica/artigos/mercextvi2009vf.pdf>

Acesso em: 23/05/2010.

MÉNDEZ RODRÍGUEZ, E. M.; MOREIRO GONZÁLEZ, J. A. Lenguaje natural e indización automatizada, **Ciencias de la Información**, v. 30, n. 3, p. 11-24, sept. 1999.

Disponível em: <http://www.bib.uc3m.es/~mendez/publicaciones/articulos/indizacion99.pdf>

Acesso em: 27/06/2005.

PAO, Miranda Lee. Automatic text analysis based on transition phenomena of word occurrences, **Journal of the American Society for Information Science**, New York, v. 29, n.3, p. 121-124, may 1978.

\_\_\_\_\_, Miranda Lee. **Concepts of information retrieval**. Englewood: Libraries Unlimited, Inc., 1989. 285 p.

POLANCO, Xavier.; GRIVEL, Luc.; RO AUTÉ, Jean. Infométrie et linguistique informatique: une approche linguistico-infométrique au service de la veille scientifique et technologique. In: JOURNÉES D'ETUDES ORGANISÉES PAR LA SOCIÉTÉ FRANÇAISE DE BIBLIOMÉRIE APPLIQUÉE AVEC LE CONCOURS. **Resumos...** Ilê Rouse: Société Française de Bibliométrie Appliquée, 1995.

Disponível em: <http://crrm.u-3mrs.fr/sfba/ile-rousse/1995/cd-rom.pdf>

Acesso em: 27/08/2007.

PRICE, Derek J. de Solla. **Little science, big science**. New York: Columbia University Press, 1963.

PRITCHARD, Allan. A statistical bibliography or bibliometrics? **Journal of Documentation**, [s.l.], v.25, n.4, p. 348-349, Dec. 1969.

RIZZON, Luiz Antenor; ZANUZ, Mauro C.; MIELE, Alberto. Evolução da acidez durante a vinificação de uvas tintas de três regiões vitícolas do Rio Grande do Sul. *Ciência e Tecnologia de Alimentos*, Campinas, v. 18, n. 2, p. 179-183, Maio/Jul.1998.

Disponível em:

[http://www.scielo.br/scielo.php?pid=S0101-20611998000200007&script=sci\\_arttext](http://www.scielo.br/scielo.php?pid=S0101-20611998000200007&script=sci_arttext)

Acesso em: 27/05/2009

ROUAULT, Jacques. **Linguistique automatique: applications documentaires**. Berne: PETER LANG, 1987. 308 p.

SPENCER, Andrew. Morphology. In: ARONOFF, M.; REES-MILLER, J. (Ed.) **The handbook of linguistics**. Oxford: Wiley-Blackwell, 2003. cap 9.

Disponível

em:

<http://books.google.com.br/books?id=omRXLtx83wgC&printsec=frontcover&dq=The+handbook+of+linguistics>

Acesso em: 26/05/2009.

SPENCER, Andrew; ZWICKY, Arnold M. (Ed.) **The handbook of Morphology**. Oxford: Wiley-Blackwell, 2001.

Disponível

em:

[http://books.google.com.br/books?id=xrdbNnbufqAC&dq=The+Handbook+of+Morphology+Por+Andrew+Spencer,+Arnold+M.+Zwicky&printsec=frontcover&source=bn&hl=pt-BR&ei=NwMcStjCDOi\\_twfY\\_smODA&sa=X&oi=book\\_result&ct=result&resnum=4](http://books.google.com.br/books?id=xrdbNnbufqAC&dq=The+Handbook+of+Morphology+Por+Andrew+Spencer,+Arnold+M.+Zwicky&printsec=frontcover&source=bn&hl=pt-BR&ei=NwMcStjCDOi_twfY_smODA&sa=X&oi=book_result&ct=result&resnum=4)

Acesso em: 26/05/2009.

SWALES, John M. Citation analysis and discourse analysis, **Applied Linguistics**, v. 7, n. 1, p. 39-56, 1986.

Disponível em: [http://applij.oxfordjournals.org/cgi/pdf\\_extract/7/1/39](http://applij.oxfordjournals.org/cgi/pdf_extract/7/1/39)

Acesso em: 01/06/2009.

\_\_\_\_\_, John M. **Genre analysis: English in academic and research settings**. Cambridge: Cambridge University Press, 1990. 260 p. (The Cambridge Applied Linguistics Series).

\_\_\_\_\_, John M. **Research genres: exploration and applications**. Cambridge: Cambridge University Press, 2004. 314 p. (The Cambridge Applied Linguistics Series).

SWALES, John M.; FEAK, Christine B. **Abstracts and the writing of abstracts**. Michigan, USA: University of Michigan Press, 2009. 104 p. (Series Michigan Series in English for Academic & Professional Purposes).

TAGUE-SUTCLIFFE, J. An introduction to informetrics, **Information Processing & Management**, v. 28, n. 1, p. 1-3, 1992.

VANTI, Nadia Aurora Peres. Da bibliometria à webometria: uma exploração conceitual dos mecanismos utilizados para medir o registro da informação e a difusão do conhecimento, **Ciência da Informação**, Brasília, v. 31, n. 2, p. 152-162, maio/ago. 2002.  
Disponível em: <http://revista.ibict.br/index.php/ciinf/article/view/171/150>  
Acesso em: 29/07/2009.

VILELA, M.; KOCH, I. **Gramática da língua portuguesa**: gramática da palavra, gramática da frase, gramática do texto/discurso. Rio de Janeiro: Almedina, 2002.

ZIMAN, J. M. Information, communication and knowledge, **Nature**, v. 224, no. 5217, p. 318-324, oct. 1969.

ZIPF, G.K. **Human behavior and the Principle of Least Effort**. Cambridge, MA: Addison-Wesley, 1949.

## GLOSSÁRIO

### *Cabernet Sauvignon*

É uma casta de uvas da família *Vitis vinifera*, originária da Região de Bordeaux, no Sudoeste da França, encontrada em zonas temperadas e quentes.

### *Dumping*

Designa uma situação em que um determinado produto é vendido por um preço inferior, em comparação ao preço do produto no mercado doméstico. A palavra também é usada para designar uma situação em que ocorre a venda de produtos a preços inferiores ao valor pago para produzi-los, com a finalidade de eliminar os concorrentes (Dicionário de Economia da UnB).

### Fermentação

Reação provocada pela levedura no vinho que dá à bebida um teor alcoólico. A fermentação é realizada por uma sequência de espécies distintas de leveduras, que fazem parte da microflora da uva, tais como: *Kloeckera Apiculata* e *Saccharomys Cerivisiae*.

### Fermentação alcoólica

Processo bioquímico pelo qual leveduras convertem o açúcar (glicose, frutose) em álcool e gás carbônico. Transforma o suco de uva em vinho.

### Fermentação maloláctica

Processo de fermentação secundária que ocorre quando há a transformação do ácido málico em ácido láctico, por ação das bactérias lácticas.

### *Hapax*

Forma, palavra ou expressão de que só se conhece um exemplo no “*corpus*” definido. É termo da linguagem filosófica e na sua forma completa é *hapax legómenom* “dito uma vez só” (DUBOIS ET AL., 2007).

*Kloeckera Apiculata*

Agente de fermentação do mosto da uva, que dá início ao processo tradicional de produção de vinhos.

## Mosto

Líquido resultante da prensagem das uvas; suco; sumo. (Glossário Enológico da Academia do Vinho).

## pH

Potencial Hidrogeniônico, símbolo ou medida que indica o grau de acidez, neutralidade ou alcalinidade de um líquido em uma escala de 0 a 14. pH superior a 7 representa alcalinidade, igual a 7 denota neutralidade e inferior a 7 indica acidez.

*Token*

Tipo ocorrência (DUBOIS ET AL., 2007).

## Vinho

Bebida alcoólica de amplo consumo, resultante da fermentação total ou parcial do mosto da uva, e produzida atualmente por aperfeiçoados processos tecnológicos (FERREIRA, s.d., 14 ed., p. 1462).

*Vitis vinífera*

Espécie botânica de uvas destinadas à produção de vinhos de qualidade, com milhares de variedades. Na prática, são utilizadas cerca de cinquenta.