

A FUNÇÃO DE ÍNDICE TEMÁTICO DA NOMINALIZAÇÃO DEVERBAL NA ESCRITA CIENTÍFICA: uma análise léxico-morfológica

Vânia Lisbôa da Silveira Guedes*

RESUMO

Analisa a função de índice temático de nominalizações deverbais na escrita científica, em português do Brasil, na área de Vinicultura. Objetiva desenvolver uma análise linguística léxico-morfológica da escrita científica na área de Vinicultura e, assim, contribuir para o refinamento de modelos de sistemas semiautomáticos de indexação. Considera como referencial teórico a análise de discurso científico, a teoria de gêneros, a teoria lexical e as leis bibliométricas voltadas para a indexação temática. A amostra é composta por quatro artigos sobre Vinicultura, publicados na *Scientific Electronic Library Online*. Os artigos foram processados pelo *Software RankWords 2.0* e foi produzida uma lista de frequência de palavras, para cada artigo. Observa-se a aplicação das leis de Zipf e do Ponto de Transição de Goffman, assim como a produtividade e frequência de uso de nominalizações em *-ção*, *-mento*, *-ncia*, *-agem* e *-da*, e os graus de relevância de suas funções de índice temáticos, em cada lista. Para a comparação dos dados, o método é aplicado a dois artigos sobre Economia. Os resultados obtidos apontam a adequação do método, especialmente aos artigos sobre Economia, e evidenciam a importância das abordagens teóricas e descritivas à nominalização para a indexação na Ciência da Informação.

Palavras-chave: Análise de gêneros discursivos. Leis de Zipf. Ponto de Transição de Goffman. Discurso científico. Nominalização verbal. Indexação temática.

* Doutora em Linguística pela Universidade Federal do Rio de Janeiro, Brasil. Professora do Curso de Biblioteconomia e Gestão de Unidades de Informação da Universidade Federal do Rio de Janeiro, Brasil.
E-mail: vanialisboa@facc.ufrj.br.

I INTRODUÇÃO

Estudos linguísticos e bibliométricos acerca da função de índice temático da nominalização verbal na escrita científica vêm revelando a importância dessas abordagens teóricas e metodológicas para a identificação do conteúdo informativo de artigos científicos, na Ciência da Informação (CI). Este estudo propõe uma análise bibliométrica e léxico-morfológica do discurso científico, na área de Vinicultura, com o objetivo de estimar a produtividade e a frequência de uso de nominalizações deverbais

na escrita científica e suas funções de índices temáticos na indexação.

A motivação para a escolha do tema de pesquisa prende-se aos seguintes objetivos: (i) verificar o grau alto de informatividade da escrita científica pela regularidade de ocorrência de nominalizações deverbais; (ii) determinar a produtividade e a frequência de uso de nominalizações em *-ção*, *-mento*, *-ncia*, *-agem* e *-da*, do ponto de vista da Linguística; (iii) investigar (i) à luz de modelos bibliométricos utilizados na indexação temática na CI.

A pesquisa baseia-se sobretudo em estudos bibliométricos, voltados para a

indexação temática, e nas teoria lexical e teoria de gêneros discursivos, no âmbito da análise crítica do discurso. Para o seu desenvolvimento impõem-se os seguintes postulados: (i) o padrão de nominalização $[X]v \rightarrow [[X]v -ção]$ N é predominante na escrita científica de artigos sobre Vinicultura; (ii) a frequência do item e produtividade de processos léxico-morfológicos de nominalizações deverbais mantém relação com o gênero científico; (iii) do ponto de vista bibliométrico, as nominalizações deverbais com função de índices temáticos ocorrem na Região de Transição (T) de Goffman.

Em suma, a análise léxico-morfológica e bibliométrica é de grande importância, sobretudo como aporte teórico e metodológico para o processo de reconhecimento do grau de informatividade da escrita científica contida em gêneros discursivos de pesquisa, no contexto da Organização do Conhecimento na Ciência da Informação (CI).

2 PRESSUPOSTOS TEÓRICOS: MORFOLOGIA, LÉXICO E PRODUTIVIDADE LEXICAL

A Morfologia é uma subárea da Linguística, que versa sobre a estrutura das palavras. Todas as línguas possuem palavras e, em todas as línguas, as palavras apresentam uma estrutura interna e são formadas por um ou mais morfemas. (SPENCER, 2003).

A Morfologia, segundo Aronoff & Anshen (2001), trata da estrutura interna das palavras potencialmente compostas de uma língua e encontra as bases sobre as quais opera no léxico. Basílio (2007) ressalta que o léxico é um sistema de formas simbólicas que evocam significados, utilizadas na construção de enunciados, para fins de comunicação. Em Aronoff & Anshen (2001), a produtividade léxico-morfológica é concebida como a extensão na qual determinado afixo é provavelmente utilizado na produção de palavras. Aronoff (1976) menciona que a noção de produtividade lexical é considerada basilar no âmbito da Morfologia Derivacional. Ele defende que uma nova palavra é gerada pela aplicação de

uma Regra de Formação de Palavra a um item pré-existente no léxico e introduz noções de graus de produtividade e de bloqueio. Bybee (2007) destaca a frequência do item como vetor facilitador para o seu acesso, diretamente relacionado ao grau de força lexical das relações morfológicas do item.

Câmara Jr. (1985) compreende que a palavra morfológica abrange palavras lexicais, como nome, adjetivo e verbo, classes abertas, e palavras funcionais, como preposição, conjunção e determinativos, classes fechadas. Sob essa perspectiva, a literatura aponta que a escrita científica tende a ser: (1) nominalizada, representando eventos preferencialmente por nomes em substituição a verbos; (2) mais densa lexicalmente, apresentando maior índice de palavras com conteúdo semântico, porém, simples gramaticalmente e (3) explícita, com clara sinalização de relações semânticas (cf. Swales, 1990, Hyland, 2009, Eggs, 2004).

Neste estudo, defende-se a importância da associação do estudo de nominalizações deverbais à análise de indicadores bibliométricos, enunciados por Zipf e Goffman, para a identificação da faixa de ocorrência de palavras significativas em artigos científicos e seleção de termos de indexação. Para Basílio (1979, p. 1), os “[...]estudos lexicais constituem um caminho promissor para o desenvolvimento de pesquisas interdisciplinares em Linguística e Ciência da Informação”. Nesse sentido, as abordagens teóricas e descritivas à nominalização na Linguística são importantes para a indexação temática, especialmente na Web Semântica idealizada por Berners-Lee. Berners-Lee; Hendler; Lassila (2001) afirmam que, na representação do conhecimento, os computadores devem ter acesso a coleções estruturadas de informação e regras de inferências semânticas que possibilitem deduções automáticas, objetivando melhorar resultados de busca, em domínios especializados do conhecimento, e a interpolaridade dos sistemas. Basílio (2007, p.78) define, do ponto de vista léxico-morfológico, a nominalização como “conjunto de processos que formam substantivos a partir de adjetivos e,

sobretudo, de verbos”, que desempenha funções textuais e semânticas, sendo que a semântica é a de denominação. Basílio (1999) menciona que a nominalização deverbal pode ter função designadora, pela utilização do significado básico do verbo para designação de conceitos, processos etc., como ocorre na linguagem científica.

3 A ANÁLISE DE DOMÍNIO NA CI E AS LEIS DE ZIPF

Na Organização do Conhecimento, Hjørland (2004) menciona que a análise de domínio oferece uma perspectiva teórica e metodológica relevante para o desenvolvimento do arcabouço teórico e prático da CI e salienta, nesse campo de pesquisa, a análise de discurso, a indexação, a análise de gêneros, a Bibliometria entre outros. Hjørland (1995) considera que o horizonte mais profícuo para a CI é o estudo dos domínios do conhecimento ou comunidades discursivas, do ponto de vista da Sociolinguística. Ressalta ainda a relevância dos estudos linguísticos para as pesquisas sobre o significado, destacando as contribuições de autores como Bakhtin e Halliday.

Como aporte teórico e metodológico ao processo de indexação, utilizam-se conhecimentos linguísticos e bibliométricos na construção de referenciais voltados para a análise de *corpora* textuais, com o objetivo de contribuir para o desenvolvimento de sistemas semiautomáticos de indexação. Nesse sentido, os pressupostos teóricos das Leis de Zipf e Ponto de T de Goffman, na Bibliometria, aproximam a frequência de uso da palavra e seu potencial de representação temática de um *corpus* em análise. Hřebicek (2002) salienta que, do ponto de vista estatístico, nenhuma objeção é feita às Leis de Zipf; entretanto, do ponto de vista linguístico, existe um *gap* entre unidades lexicais e *corpora* que indica a ausência de fenômenos relevantes ligados às estruturas lexicais. Ele explica que as propriedades semânticas das palavras podem ser observadas quando elas são testadas como unidades lexicais em textos. Hřebicek (2002)

propõe, entre outras, as seguintes questões de pesquisa: (1) a abordagem mencionada está baseada em *corpora* capazes de expressar a real estrutura lexical de uma linguagem? (2) existem abordagens teóricas, associadas às Leis de Zipf, suficientemente fortes para suportar o impacto causado pela mudança de análise de *corpora* para textos?

Bybee & Hopper (2001) ao discutirem a relação da frequência com a emergência de estruturas linguística mencionam que Zipf¹ foi pioneiro nos anos 30. *The Psycho-biology of Language* editado em 1935 se baseia no *Princípio do Menor Esforço*. Zipf antecipou muitos temas recentes de investigação sobre a relação entre a frequência e a estruturas linguísticas.

4 METODOLOGIA

A linha teórico-metodológica baseia-se na Sociolinguística e na Bibliometria. Como amostra, optou-se por um *corpus* composto por quatro artigos sobre Vinicultura. Visando à confrontação dos dados, foram selecionados dois artigos sobre Economia. A seleção dos textos segue os seguintes critérios:

- (a) quatro artigos de periódicos científicos sobre Vinicultura e dois sobre Economia indexados na Base de Dados da *Scielo Brazil Library Online*;
- (b) artigos científicos e tecnológicos com textos integrais disponíveis na *web*, com 2.000 a 6.000 palavras;
- (c) artigos científicos e tecnológicos publicados em língua portuguesa, no período de 1998 a 2008.

As referências dos seis artigos selecionados estão elencadas no Quadro 1, a seguir.

¹ 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).

Quadro 1: Referências

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&tlng=es Acesso em: 27/03/2010
2. 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: 27/03/2010
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: 27/03/2010
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: 27/03/2010.
5. 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: 27/03/2010.
6. 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: 27/03/2010

Fonte: Amostra

4.1 Procedimento de Análise

- (a) seleção de quatro artigos sobre a indústria de vinhos e dois sobre Economia;
- (b) contagem das palavras pela utilização do *software RankWords2.0*;
- (c) ordenação das palavras operando o *software* mencionado em (b) que produziu três colunas: 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: total de palavras, total de palavras distintas, palavras com frequência 1 e *ranking* das nominalizações deverbais;

- (e) identificação dos seguintes padrões de nominalizações: (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 de produtividade das nominalizações, 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) identificação do Ponto T de Goffman, com o objetivo de delimitar a Região de Transição que, segundo Goffman, reúne os itens lexicais de alto conteúdo semântico;
- (h) 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.

4.2 Bibliometria: leis de Zipf e Ponto T de Goffman

O termo Bibliometria na CI foi introduzido por Pritchard (1969), para designação da ciência que utiliza métodos matemáticos e estatísticos com o objetivo de analisar os fenômenos inerentes aos processos de comunicação escrita. Okubo (1997) afirma que as abordagens bibliométricas são baseadas na noção de que a produção de conhecimento é a essência da pesquisa científica e, nesse contexto, a literatura é a manifestação constituinte do conhecimento. As principais leis bibliométricas são: Lei de Bradford, (produtividade de periódicos), Lei de Lotka (produtividade científica) e Leis de Zipf (frequência de palavras). (Guedes 2012).

No que se refere às leis de Zipf, Bybee e Hopper (2001: 1), ao discutirem a relação da frequência com a emergência de estruturas linguísticas, afirmam que Zipf foi o pioneiro, ao editar a obra *"The Psycho-biology of Language"*, em 1935, com base no Princípio do Menor Esforço. Segundo eles, Zipf antecipou muitos temas de pesquisa investigados recentemente sobre a frequência de tipo e de uso de estruturas linguísticas na fala e na escrita. Ele observou certas regularidades na estrutura linguística de textos e enunciou duas leis (Zipf, 1935, 1945, 1949). A

primeira lei está relacionada às palavras de alta frequência e a segunda às de baixa frequência.

A primeira lei de Zipf enuncia que "o produto da ordem de série(r) de uma palavra pela sua frequência (f) é aproximadamente constante (C):

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

A segunda lei enuncia que "em um texto, várias palavras de baixa frequência de ocorrência têm a mesma frequência". Essa leis foram complementadas por diversos estudos, destacando-se a modificação proposta por Booth (1967) para a segunda lei, representada matematicamente da seguinte forma:

$$(2) I_1 = n(n + 1) I_n$$

Goffman, ao orientar Pao (1978), explica que a delimitação da região de transição de palavras de baixa frequência para as de alta frequência, onde teoricamente ocorrem palavras com alta carga semântica, é calculada a partir da aplicação da equação do Ponto T de Goffman a seguir:

$$(3) n = \frac{-1 + \sqrt{1 + 8I_1}}{2}$$

Ao valor de n assim determinado corresponde o Ponto T de Goffman. Existe uma região, ao redor desse ponto, com probabilidade de concentrar palavras de alta carga semântica. Pao (1978), ao testar a hipótese e seguindo a orientação de Goffman, delimitou a região de transição projetando para baixo o mesmo número de palavras que ocorreu acima da frequência ligada ao Ponto T. Nessa região, foram analisadas as palavras e, assim, excluídas as palavras sem conteúdo semântico, tais como: preposições, artigos, conjunções, pronomes e verbos auxiliares, já que o objetivo do estudo era a investigação da carga semântica de termos contidos na região. A autora conclui que os resultados alcançados foram satisfatórios, para os textos analisados, e parecem indicar que a seleção de termos de indexação, a partir de uma lista de frequência de uso de palavras, é uma perspectiva promissora para a indexação automática.

5 ANÁLISE, SÍNTESE E DISCUSSÃO DE RESULTADOS

A síntese dos resultados apresentada nas tabelas 7 e 8 evidencia alto grau de produtividade do padrão $[X] v \rightarrow [[X]v -\text{ção}]N$, seguido de $[X] v \rightarrow [[X] v -\text{mento}] N$ nos textos em análise. Em contrapartida, verifica-se baixo grau de produtividade do padrão $[X]v \rightarrow [[X]v -\text{da}]N$, tendendo a zero em alguns artigos.

Assim, constata-se que o gênero artigo científico prevê frequência alta de produtividade e uso de nominalizações $[X] v \rightarrow$

$[[X] v -\text{ção}] N$ e $[X] v \rightarrow [[X] v -\text{mento}] N$. Em função das chances de emergência, tais formas bloqueiam e enfraquecem o processamento dos demais padrões possíveis. Nesse sentido, a frequência de uso e a força lexical dos itens influenciam diretamente o grau de produtividade dos padrões de nominalização deverbiais investigados (BYBEE, 1988, p. 121).

A investigação do grau relativo de produtividade e o cálculo da frequência de uso de cada padrão de nominalizações indicaram os resultados plotados nas tabelas 1 e 2, revelando um gradiente de produtividade e uso de sufixos nominalizadores, nos artigos 1, 2, 3, 4, 5 e 6.

Tabela 1: 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 T5	T1	T2	T3	T4 T5	T1	T2	T3	T4 T5	T1	T2	T3	T4 T5	T1	T2	T3	T4 T5
-ção	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
-mento	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
-agem	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
-ncia	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
-da	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
Totais	104	139	181	177	100	100	100	100	41	65	68	40	100	100	100	100	6,22	9,10	9,79	12,08

Fonte: Própria

Tabela 2- Síntese da produtividade dos sufixos nominalizadores em análise nos artigos 5 e 6, bem como suas frequências relativas de ocorrência.

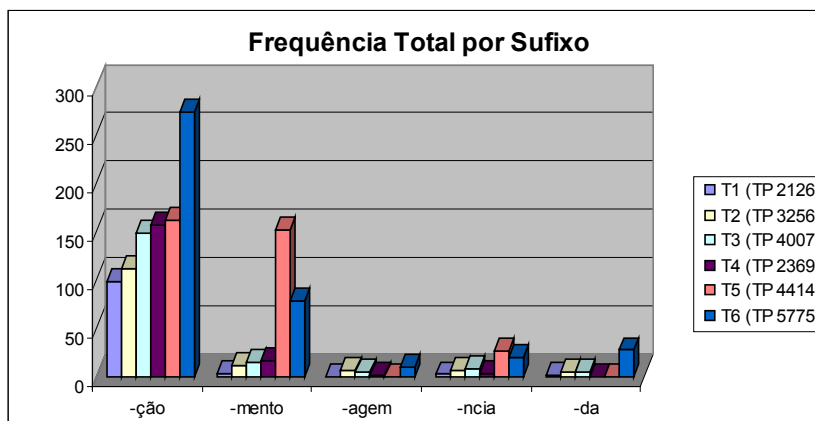
Sufixo	Frequência Total por Sufixo		Frequência Relativa do Sufixo/ Frequência Total de Sufixos		Bases Distintas		Bases Distintas/ Total de Bases Distintas		Frequência Média das Bases Distintas	
	T5	T6	T5	T6	T5	T6	T5	T6	T5	T6
-ção	162	274	47,78	66,99	47	61	68,11	64,21	3,44	4,49
-mento	151	78	44,54	19,07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

Fonte: Própria

Para efeito de indexação, o gráfico 1 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 1 - Representação espacial da frequência de ocorrência de cada sufixo nominalizador deverbal analisado, com o número total de palavras de cada artigo.



Fonte: Própria

No que se refere à análise bibliométrica, as tabelas 3 e 4 sintetizam os resultados verificados na aplicação da Segunda Lei de Zipf nos seis artigos investigados.

Tabela 3 - Síntese dos resultados verificados na aplicação da Segunda Lei de Zipf aos artigos sobre Vinicultura

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 ₁	380	492	593	380	-	-	-	-	-	-	-	-
I ₂	89	131	148	91	126,66	164	197,66	126,66	29,21	20,12	25,12	28,15
I ₃	30	76	79	43	63,33	82	98,83	63,33	52,62	7,31	20,06	32,10
I ₄	15	31	43	29	38	49,2	59,30	38	60,52	36,99	29,17	23,68
I ₅	17	21	39	15	25,33	32,8	39,53	25,33	32,88	35,97	1,34	40,78

Fonte: Própria

Tabela 4 - Síntese dos resultados verificados na aplicação da Segunda Lei de Zipf aos artigos sobre Economia

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

Fonte: Própria

A aplicação das Leis de Zipf apresentou desvios em 100% dos textos analisados, conforme apresentado nas tabelas 3 e 4. Entretanto, 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. Contudo, os resultados obtidos no processamento dos textos 3, 4, 5 e 6 sobre Vinicultura e Economia verificaram-se ainda mais satisfatórios, principalmente na análise dos textos sobre Economia. (tabelas 1 e 2).

Observa-se ainda 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; Antoniadis & Lallich-Boidin (1987), Eggins (2004), dentre outros. Além disso, do ponto de vista bibliométrico, verifica-se que o índice médio de repetição de cada palavra nos seis textos da amostra, 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.

(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 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. Basílio (2007) esclarece que o uso em terminologias científicas e tecnológicas é mais uma vertente da função denominadora das nominalizações deverbais.

7 CONCLUSÕES E SUGESTÕES PARA PESQUISAS FUTURAS

A análise das conexões lexicais, formuladas nas regras, e suas recorrências na formação de nominalizações nos artigos analisados permitem afirmar que o modelo morfológico $[X] v \rightarrow [[X] v -\text{ção}] N$ é predominante e 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 improdutivos 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 Vinicultura 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.

No entanto, 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 indicaram a ocorrência de palavras de alta carga semântica.

A investigação das frequências de uso das nominalizações deverbais na Região de Transição de Goffman revelou-se satisfatória em 100% da amostra. Contudo, os resultados obtidos no processamento dos textos sobre Economia verificaram-se ainda mais satisfatórios.

Os achados de pesquisa demonstram a importância da análise léxico-morfológica da produtividade de nominalizações deverbais na escrita científica, na busca de evidências empíricas de que o sufixo nominalizador *-ção* em artigos científicos apresenta alta densidade de informação, 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 discursivo científico 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.

Finalmente, sugere-se estender esta investigação para maior número de textos e áreas de assunto, com o intuito de validação do método.

FUNCTION INDEX THEME NOMINALISATION DEVERBAL WRITTEN IN SCIENTIFIC: one lexical- morphological analysis

ABSTRACTS This study analyses the index function of deverbal nominalizations in the scientific writing in Brazilian Portuguese language in the Viniculture area. The study aims to develop a statistical and lexical-morphological linguistic analysis of scientific writing in the Viniculture area and thus contribute to the refinement of models of semi-automatic indexing systems. The theoretical framework used were the scientific discourse analysis, the genre theory, the lexical theory and subject indexing based on Bibliometrics. To identify those lexico-morphological features in such terminology, four papers about the Viniculture were selected which were published in the Scientific Electronic Library Online. These papers were processed by Software RankWords 2.0. For each paper, a list of words frequencies was obtained. In each list it was investigated the Zipf's Laws and Goffman Transition Point adequacy, the productivity of deverbal nominalization in -ção, -ment, -ncia, -agem, -da and the use frequencies of these words and the relevance of their functions as an index term in the subject indexing. Aiming data confrontation, the method is applied into two text about Economics. The results show the adequacy of the method and highlight the importance of the theoretical and descriptive approaches to nominalization for the subject indexing within Information Science.

Keywords: Discourse genre Analysis. Zipf's Laws. Goffman's Transition Point. Scientific Discourse. Deverbal Nominalization. Subject Indexing.

REFERÊNCIAS

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

ARONOFF, M.; ANSHEN, F. 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=xrd bNbufqAC&dq=The+Handbook+of+Morphology+Por+Andrew+ARONOFFSpencer,+Arnold+M.+Zwicky&printsec=frontcover&source=bn&hl=pt-BR&ei=NwMcStj-CDOi_twfY_smODA&sa=X&oi=book_result&ct=result&resnum=4. Acesso em: 26 maio 2013.

BASÍLIO, M. M. de P. **Teoria lexical**. 8. ed. revista e atualizada. São Paulo: Ática, 2007. 104 p. (Série Princípios, 88).

_____. 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, 1979.

_____. 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 maio 2013.

BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. 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. Acesso em: 02 dezembro de 2013.

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

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. Acesso em: 10 fevereiro 2013.
- 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 fevereiro 2013.
- CAMARA JR, J. M. **História e estrutura da língua portuguesa**. 4. ed. Rio de Janeiro: Editora Padrão, 1985.
- CHOMSKY, N.; MILLER, G. 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. Acesso em: 15 fevereiro 2013.
- EGGINS, S. **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. Acesso em: 21 fevereiro 2013.
- GUEDES, V. L. da S. A bibliometria e a gestão da informação e do conhecimento científico e tecnológico: uma revisão da literatura. **Pontodeacesso**, v.6, n.2, 2012. Disponível em: <http://www.portalseer.ufba.br/index.php/revistaici/article/viewArticle/5695>. Acesso em: 04 março 2013.
- HJØRLAND, B. 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 junho 2013.
- _____. 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, L. Zipf's law and text, **Glottometrics**, Czech Republic, v. 3, n. 22, p. 23-38, 2002.
- HYLAND, K. **Academic discourse: english in a global context**. New York: Continuum International Publishing Group, 2009. 215p.
- OKUBO, Y. Bibliometric indicators and analysis of research systems: methods and examples. Paris: OECD Publishing, 1997. **Science, Technology and Industry Working Papers, 1997/01**. Disponível em: [http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=OCDE/GD\(9\)41&docLanguage=En](http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=OCDE/GD(9)41&docLanguage=En). Acesso em: 27 mar. 2012.
- PAO, M. L. 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.
- PRITCHARD, A. A statistical bibliography or bibliometrics? **Journal of Documentation**, [s.l.], v.25, n.4, p. 348-349, Dec. 1969.
- ROUAULT, J.; Antoniadis, G.; Lallich-Boidin, G. **Linguistique automatique: applications documentaires**. Berne: PETER LANG, 1987. 308 p.
- SPENCER, A. 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 maio 2009.

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

Zipf, G.K. **The psycho-biology of language** : an introduction to dynamic philology. Oxford, England: Houghton, Mifflin, 1935. 336 pp.

Zipf, G.K. (1945) The meaning frequency relationship of words. **The Journal of general psychology**, 33, p. 251-256.

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