

# RECORRÊNCIA DE NOMINALIZAÇÕES DEVERBAIS EM RESUMOS DE CARTAS CIENTÍFICAS EM LÍNGUA PORTUGUESA E A INDEXAÇÃO TEMÁTICA

Deverbal nominalization recurrence in the abstracts of scientific letters in  
portuguese language and the subject indexing

VÂNIA LISBÔA DA SILVEIRA GUEDES

Universidade Federal do Rio de Janeiro

vanielisboa@facc.ufrj.br

MARIA JOSÉ VELOSO DA COSTA SANTOS

Universidade Federal do Rio de Janeiro

msantos1402@gmail.com

## Resumo

Este estudo investiga a recorrência de nominalizações deverbais em resumo de cartas científicas em língua portuguesa, do arquivo da cientista Bertha Lutz, com intuito de estimar os graus de produtividade das nominalizações e observar suas funções de índice. A pesquisa situa-se entre a Linguística e a Ciência da Informação e objetiva contribuir para o refinamento da indexação. Estabeleceu-se a hipótese de que as nominalizações em *-ção* prevalecem, em cartas científicas, e exercem funções de índices relevantes. Consideram-se como referencial teórico a análise de gêneros na Sociolinguística, a teoria lexical e a indexação na Ciência da Informação. Os resumos foram processados pelo software *RankWords2.0* e produzida uma lista de frequência de palavras. A aplicação das leis de Zipf e Ponto T confirmou sua viabilidade na indexação dos resumos. Finalmente, observa-se a recorrência de nominalizações relevantes, na

Recibido  
8/7/12  
Aceptado  
21/12/12

região de concentração de palavras-chave, e ressalta-se que os dados obtidos corroboram a hipótese estabelecida.

**Palavras-chave:** Cartas Científicas. Nominalização Deverbal. Indexação Temática.

### Abstract

The study investigates the recurrence of deverbal nominalization in the abstracts of scientific letters in Portuguese language from the files of the scientist Bertha Lutz. The study aims to estimate the degree of productivity of nominalizations and to observe their index role. The study is situated between the Linguistics and Information Science and intends to contribute to the indexing refinement. The hypothesis is that the nominalizations in *-ção* prevail in the lexicon of scientific letters as relevant index. The theoretical framework used was the genre analysis, academic discourse and lexical theory, in the Sociolinguistics, and indexing within Information Science. The abstracts were processed by RankWords 2.0 software. A list of word frequency was obtained. It was confirmed the adequacy of Zipf's Laws and Goffman Transition Point. The results point out to the recurrence of nominalizations in *-ção* with high semantic content in Goffman T Region and confirm the established hypothesis.

**Keywords:** Scientific Letters. Deverbal Nominalization. Subject Indexing.

## 1. INTRODUÇÃO

A presente pesquisa volta-se para a interdisciplinaridade entre a Ciência da Informação e a Linguística. Desenvolve uma análise léxico-morfológica em cartas científicas sobre herpetologia<sup>1</sup>, área de atuação da cientista Bertha Lutz e seus correspondentes, investigando processos produtivos de nominalização deverbal e suas funções de índices relevantes para indexação temática da informação científica.

---

1 Campo da zoologia dedicado ao estudo dos répteis e anfíbios.

A indexação temática na Ciência da Informação consiste na classificação do conteúdo tanto do documento quanto das consultas enunciadas pelos usuários, com o propósito de reter as ideias mais significativas e representá-las por termos de indexação. Os termos de indexação podem ser extraídos da linguagem natural, utilizada pelos autores, ou de um vocabulário controlado, selecionado *a priori* (Mendez Rodriguez e Moreira Gonzalez 1999).

Nesse sentido, o estudo se vale (a) da teoria sobre análise de gêneros na Sociolinguística, para a observação da produtividade da nominalização de verbos com diferentes sufixos, sobretudo o sufixo *-ção*. (b) de recursos quantitativos utilizados na indexação temática para o exame dos dados, sob uma perspectiva bibliométrica na Ciência da Informação, para a análise da frequência de palavras.

O desenvolvimento de uma área científica e sua maturidade é caracterizado pela publicação de trabalhos escritos, sendo estes considerados por muitos autores, tão importantes quanto a própria pesquisa que lhes deu origem.

Antes da publicação dos resultados, a pesquisa científica é exaustivamente discutida, criticada e avaliada pela comunidade científica que corrobora ou refuta os resultados, ou ainda, estabelece novas perspectivas para esse ou aquele campo de interesse. A pesquisa, depois de legitimada pelos pares, torna-se conhecimento científico propriamente dito. Para Ziman (1969) não existe ciência sem sua comunicação e Meadows (1999) preconiza que os processos de comunicação são básicos para a natureza e prática da ciência, tanto que considera a comunicação o “coração da ciência”.

Assim, considera-se que a comunicação é vital para a ciência porque garante a divulgação dos resultados de pesquisas; a proteção da propriedade intelectual; a aceitação dos resultados pelos pares e a consolidação do conhecimento. Sob essa perspectiva, os cientistas são potencialmente consumidores e produtores de informações de natureza científica.

Entende-se por comunicação científica como a troca de informações entre os membros da comunidade científica, incluindo, segundo Garvey (1979, p.ix) em seu livro intitulado “*Communication, the essence of science*”:

*facilitating information exchange among librarians, scientists, engineers, and students”*,

as atividades associadas à produção, disseminação e uso da informação, desde o momento em que o cientista concebe uma idéia para pesquisar, até que os resultados de sua pesquisa sejam aceitos como constituintes do conhecimento científico.

Conceito que vem ao encontro do pensamento de Mikhailov *et al.* (1984) que definem comunicação científica como processos combinados de representação, transferência e recepção de informações de caráter científico na sociedade humana.

As cartas desempenharam papel crucial na comunicação científica. No passado, circulavam entre filósofos naturais com o propósito de trocar informações sobre suas pesquisas. Foi a partir das cartas informativas que os cientistas escreviam para seus pares que se desenvolveram outros gêneros textuais acadêmicos como, por exemplo, o gênero artigo científico.

Santos (2009) também defende que as cartas tiveram um papel basilar para a comunicação na ciência. Utilizadas para a transmissão de conhecimento e difusão de ideias, eram trocadas entre pesquisadores e cientistas para relatar pesquisas originais para serem discutidas, opinadas e avaliadas pelos pares. A autora acrescenta que:

a rede de comunicação de correspondentes é conhecida pela expressão colégio invisível utilizada pela primeira vez, de acordo com Merton (1968), por Boyle e reutilizada por Solla Price referindo-se a um grupo de pesquisadores que mantinha correspondência entre si (Santos 2009: 318).

Swales (1990) menciona 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 março de 1665. Opinião referendada por Bazerman (2006), Hyland (2009) entre outros quando mencionam que o primeiro artigo científico publicado no *Philosophical Transactions* originou-se da correspondência trocada entre o alemão Henry Oldenburg, secretário da *Royal*

*Society of London desde 1662*, e outros pesquisadores. Como secretário, essa correspondência foi intensificada e utilizada para a edição de textos publicados nesse periódico. O *Philosophical Transactions of the Royal Society of London* junto com o *Journal des Savants*<sup>2</sup>, editado na França em janeiro de 1665, são considerados os primeiros periódicos científicos publicados (Guedes 2010). Bazerman (1988) esclarece que a organização de artigos relatando experimentos serve como uma manifestação visível de tendências discutidas sobre o tema.

Bazerman (1997) ressalta que os gêneros auxiliam na navegação pelos complexos mundos da comunicação escrita e de atividades simbólicas, na medida em que reconhecer um gênero textual permite, por exemplo, o conhecimento de muitos aspectos e fenômenos inerentes à situação social e institucional, às atividades propostas, aos papéis disponíveis ao leitor e autor, aos motivos, às ideias, à ideologia e ao conteúdo esperado do documento.

Ao analisar a escrita científica no gênero acadêmico e seu propósito comunicativo Hyland (2009) afirma que os autores buscam *encaixar* sua escrita em um mundo social particular, por meio de um discurso aprovado por seus pares. Como resultado, os gêneros acadêmicos têm atraído crescente atenção como fontes de informação sobre as práticas das comunidades de discurso acadêmico. Nesse sentido, Hyland e Salager-Meyer (2008) defendem que os textos são socialmente produzidos em comunidades particulares e dependem dessas comunidades para a obtenção de sentido, logo pelo estudo da escrita científica aprendemos muito mais sobre as questões ligadas às áreas de assunto e ainda como o conhecimento é construído, negociado e persuasivo. Bhatia (2004, *apud* Guedes 2010) menciona que, nos últimos anos, a teoria sobre gênero discursivo tem contribuído para a compreensão de como o discurso é usado no contexto acadêmico, profissional e em outros contextos institucionais. Na escrita de textos científicos, por meio da nominalização, processos normalmente representados por verbos e propriedades normalmente representadas por adjetivos surgem como substantivos (Halliday 1994).

---

2 Grafia antiga: *Journal des Sçavants*.

Guedes (2010), ao investigar morfologicamente a escrita científica em textos editados no gênero artigo de periódico, observa que a formação de nominalização representada por  $[X] v \rightarrow [ [X] v \text{-ção} ] N$  é predominante no campo semântico da indústria de vinhos, indicando os caminhos mais adequados para a indexação temática em sistemas semi-eletrônicos de representação e recuperação da informação, na Ciência da Informação<sup>3</sup>. O estudo volta-se para os processos produtivos de nominalização deverbal e se vale ainda de recursos quantitativos utilizados na indexação temática para a identificação do conteúdo informativo de artigos, sob uma perspectiva bibliométrica, na Ciência da Informação.

Basílio (2007: 78) conceitua nominalização como o “conjunto de processos que formam substantivos a partir de adjetivos e, sobretudo, de verbos”. Basílio acrescenta 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.

Hyland e Salager-Meyer (2008: 298) defendem que na análise da escrita científica, sob a perspectiva da Ciência da Informação, a preocupação central é com os sistemas de recuperação da informação objetivando a comunicação científica. Os autores acrescentam a análise de citações na Bibliometria como metodologia de análise da escrita científica. Swales (1986) em seu artigo intitulado *Citation analysis and discourse analysis* aponta aspectos interdisciplinares nos estudos em Ciência da Informação, como por exemplo a análise de citações que utiliza indicadores bibliométricos e métodos de análise discursiva.

A Ciência da Informação possui em seu arcabouço um campo de pesquisa denominado Bibliometria que contribui para o estabelecimento de seus fundamentos epistemológicos, formado por um conjunto de técnicas, leis e princípios empíricos que utilizam métodos matemáticos e estatísticos para avaliar e quantificar os processos de comunicação escrita. A análise bibliométrica é um importante instrumento para o estabeleci-

---

3 Ciência social apoiada em tecnologia que tem por objeto o estudo das propriedades gerais da informação, a análise dos seus processos e a concepção dos seus produtos e sistemas (Le Coadic 1996); voltando-se, sobretudo para a produção e a organização do conhecimento.

mento de indicadores em uma área de conhecimento porque apresenta os aspectos quantitativos de produção, disseminação e uso da informação científica registrada (Guedes e Borschiver 2005; Santos 2009).

O presente estudo, com a intenção de ligar os estudos de aplicação das leis bibliométricas a estudos morfológicos na Linguística e à análise de gêneros do ponto de vista da Sociolinguística, investiga a recorrência de nominalizações deverbais na escrita de cartas científicas, em língua portuguesa, com vistas a estimar a produtividade dessas nominalizações na comunicação científica e observar suas funções de índice mais relevantes na indexação temática da informação, baseado na pesquisa de doutoramento de Guedes (2010), voltada para a análise léxico-morfológica e bibliométrica de artigos científicos, e no estudo de Santos (2009) que desenvolve uma análise bibliométrica de resumos de cartas científicas com vistas à indexação temática.

Para tal, foram utilizadas cartas do arquivo pessoal da cientista formada em biologia pela Universidade de Sorbonne e professora emérita da UFRJ, Bertha Maria Júlia Lutz (1894-1976), especialista em herpetologia, em particular, os anfíbios.

## 2. OBJETIVOS

Neste item serão apresentados os objetivos gerais e específicos que norteiam o desenvolvimento do presente estudo.

### 2.1. Objetivo geral

Contribuir para o desenvolvimento de modelos de sistemas semi-automáticos de indexação de resumos de cartas científicas pela aplicação da análise linguística e bibliométrica.

## 2.2. Objetivos específicos

- ✦ Analisar a produtividade de nominalizações deverbais na escrita de resumos do gênero cartas científicas baseado em modelos quantitativos empregados na indexação no âmbito da Bibliometria na Ciência da Informação;
- ✦ Investigar as funções de índice mais relevantes das nominalizações deverbais em *-ção* na indexação temática, baseada em suas frequências de ocorrência ou de uso;
- ✦ Contribuir para o enriquecimento das abordagens teóricas e práticas da indexação da informação e
- ✦ Fortalecer os pontos de interface da Linguística com a Bibliometria na Ciência da Informação.

## 3. REFERENCIAL TEÓRICO

Para fundamentar a presente pesquisa utilizaram-se como referencial teórico a análise de gêneros na Sociolinguística e a teoria lexical no campo da Linguística; a indexação e recuperação da informação no âmbito da Bibliometria no campo da Ciência da Informação. Para tal, foram destacadas as abordagens teóricas e descritivas de Swales, Bazerman, Hylland e Basílio que discorrem sobre os fenômenos inerentes à escrita científica em gêneros acadêmicos. Na área da Bibliometria foram selecionados trabalhos de aplicação das leis de Zipf e Ponto de Transição de Goffman, tais como os trabalhos de Zipf, Goffman e Pao.

### 3.1. Análise de gênero, léxico e nominalização deverbal

Bazerman (2006) 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.

Swales (1990) defende que as comunidades de discurso detêm convenções próprias e tradições para as quais diversas atividades verbais convergem produzindo relatórios e publicação de atividades. As classes recorrentes de eventos comunicativos são os gêneros que regem a vida verbal de uma comunidade de discurso.

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.

A nominalização, segundo Basílio (2007: 78), refere-se ao “conjunto de processos que formam substantivos a partir de adjetivos e, sobretudo, de verbos”. Ao discorrer sobre a representação e o acesso lexical por palavras previamente existentes, Basílio (1999) menciona que a nominalização deverbal pode ter função designadora, usando o significado básico do verbo para designar conceitos, processos etc., como ocorre na linguagem científica. “Palavras como neutralização, derivação, referência e outras, da terminologia linguística, são formadas com propósitos designadores” (Basílio 1999: 63). No que se refere ao processo de formação de nominalizações, Basílio (2004) afirma que o léxico não é somente um conjunto de palavras, mas sim um sistema dinâmico que também engloba estruturas de expansão, dentre as quais as mais importantes são as que correspondem à formação de palavras (Guedes 2010).

### 3.2. Leis de Zipf e Ponto de Transição (Ponto T) de Goffman

A área de Bibliometria na Ciência da Informação apresenta um conjunto de leis e princípios empíricos baseados em métodos matemáticos e estatísticos de investigação, acesso e quantificação de processos de comunicação escrita. Miranda Pao (1989) refere-se à Bibliometria como um termo introduzido por Allan Pritchard, em seu artigo *Statistical Bibliography or Bibliometrics*, editado em 1969, para designar a área do conhecimento, que utiliza métodos matemáticos e estatísticos, com o objetivo de analisar os processos de comunicação escrita. Pao (1989) considera os estudos bibliométricos da literatura como instrumentos de quantificação, diagnóstico, e prognóstico dos 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 científica é a manifestação constituinte do conhecimento. Okubo esclarece que os dados e indicadores bibliométricos podem servir como ferramentas ao menos para a descrição e expressão de questões que surgem no mundo da Ciência (Guedes 2012).

As principais leis bibliométricas são: Lei de Bradford, (produtividade de periódicos), Lei de Lotka (produtividade científica de autores) e Leis de Zipf (frequência de palavras). Bybee e Hopper (2001: 1), ao discutirem a relação da frequência de uso (*token frequency*) e da frequência de tipo (*type frequency*) com a emergência de estruturas linguísticas, afirmam que George K. Zipf foi o pioneiro, ao editar a obra "*The Psycho-biology of Language*", em 1935, baseada no *Princípio do Menor Esforço*. Segundo os autores, Zipf antecipou muitos temas de pesquisa investigados recentemente sobre a frequência de tipo e de uso das estruturas linguísticas na fala e na escrita (Guedes 2010).

George Kingsley Zipf (1902-1950) observou certas regularidades na estrutura linguística de textos. Com base nessas observações e na análise de frequência de ocorrência de palavras, em um texto que ele classifica de suficientemente longo, enunciou duas leis (Zipf 1935, 1949). A primeira

lei está relacionada às palavras de alta frequência em um texto e a segunda às de baixa frequência.

A primeira lei de Zipf está assim formulada: “o produto da ordem de série( $r$ ) de uma palavra pela sua frequência ( $f$ ) é aproximadamente constante ( $c$ ).

$$r \times f = c$$

A segunda lei enuncia que “em um exto, várias palavras de baixa frequência de ocorrência têm a mesma frequência”.

Essas leis foram complementadas por diversos estudos, destacando-se a modificação proposta por Booth para a segunda lei, conhecida como lei Zipf-Booth representada matematicamente da seguinte forma:

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

Onde:

- $I_1$  é o número de palavras que têm frequência 1;
- $I_n$  é o número de palavras que tem frequência  $n$ ; e
- 2 é uma constante empírica válida para a língua inglesa.

Ao investigar a validade da Lei de Zipf-Booth para a escrita científica na área de Mecânica dos Solos, Guedes e Valois (1988) inferem que a expressão  $n(n+1)/2$  corresponde à soma dos  $n$  primeiros números naturais, conforme defendido por Carl-Friedrich Gauss (1777-1855), considerado na literatura um dos mais importantes matemáticos de todos os tempos. Por exemplo, a soma de  $1+2+3+4+5 = 15$  poderia ser calculada, levando-se em conta que  $n = 5$ , por  $5(5+1)/2 = 15$ . Assim entende-se que a Segunda Lei de Zipf pode ser enunciada alternativamente como: a relação entre o número de palavras que ocorrem uma só vez ( $I_1$ ) e o número de palavras que ocorrem  $n$  vezes ( $I_n$ ), na escrita de textos científicos, é igual à soma dos  $n$  primeiros números naturais, representada matematicamente pela fórmula

$$I_1/I_n = 1 + 2 + 3 + \dots + n$$

Goffman, a partir da segunda lei de Zipf modificada por Booth, admitiu que havia uma região na lista de palavras, localizada entre as palavras de alta frequência e as de baixa frequência, a qual ele denominou de “região de transição” com probabilidade de concentrar as palavras de alto conteúdo semântico, ou seja, palavras mais significativas e representativas do conteúdo temático ou intelectual de um texto. Para se chegar a essa região, este desenvolveu uma fórmula denominada de Ponto de Transição (T) de Goffman.

O Ponto T de Goffman é representado matematicamente pela expressão:

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

Onde:

- ♦ n representa o ponto T;
- ♦  $I_1$  é o número de palavras que tem frequência 1;
- ♦ 8 é uma constante derivada da língua inglesa;
- ♦ 2 é um número matemático da fórmula de Bhaskara, para resolução de equações de 2º grau (Guedes 2010; Santos 2009).

As leis de Zipf, enriquecidas com o ponto T de Goffman, relacionam-se diretamente com a representação da informação e maior precisão na sua recuperação.

Pao (1978), ao verificar a hipótese de que a Região de Transição de Goffman produziria as palavras de maior conteúdo semântico de um dado texto, delimitou a região projetando para baixo o mesmo número de palavras que ocorreu acima da frequência que corresponde ao Ponto T. Nessa região, foram excluídas as palavras sem conteúdo semântico, tais como: preposições, artigos, conjunções, pronomes e verbos auxiliares. Ao analisar os resultados obtidos, 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 (Guedes 2010).

É importante destacar que as ideias lançadas por Zipf fundamentam a moderna Linguística Quantitativa, no entanto, sua influência não se restringe a essa área, sendo também observada em outros domínios do conhecimento, como na Bibliometria, área de pesquisa da Ciência da Informação.

## 4. MATERIAL E MÉTODO

Nessa pesquisa, conforme visto anteriormente propõe-se analisar a produtividade de nominalizações deverbais na escrita de resumos do gênero cartas científicas, com o propósito comunicativo, e investigar a função de índices relevantes das nominalizações em *-ção*, na indexação temática das cartas do arquivo pessoal de Bertha Lutz.

### 4.1. Bertha Lutz

Bertha Maria Júlia Lutz nasceu em São Paulo no ano de 1894 e morreu no Rio de Janeiro, em 1976. É nacionalmente conhecida por suas pesquisas científicas e sua atuação como líder feminista. Filha da enfermeira inglesa Amy Fowler e do cientista pioneiro da medicina tropical no Brasil, Adolpho Lutz, especializou-se em anfíbio anuros e também se dedicou a pesquisas em outras especialidades da biologia. Trabalhou em instituições de renome como o Museu Nacional da UFRJ, onde ingressou por concurso público, e o Instituto Oswaldo Cruz, ambos no Rio de Janeiro. Foi a segunda mulher a ingressar no serviço público brasileiro. Bertha Lutz foi também presidente do Conselho Federal de Expedições Artísticas e Científicas, dentre outras atividades (Santos 2009).

Foram selecionados 100 resumos de cartas da Série Produção Científica, especificamente a correspondência científica do arquivo pessoal de Bertha Lutz.

## 4.2. Procedimentos metodológicos

Neste item, são apresentados o procedimento de análise e os critérios que nortearam o processamento dos resumos das 100 cartas que compuseram a amostra, seguindo-se então as seguintes etapas:

- α) delimitação da amostra - foram selecionadas 100 cartas no Série Produção Científica do arquivo pessoal de Bertha Lutz;
- β) contagem das palavras - foi utilizado o *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: excluir títulos do artigo, nomes, afiliações e títulos dos autores, resumos, palavras-chave, sumários, referências, notas de rodapé, bibliografias, agradecimentos, tabelas, gráficos, ilustrações, diagramas, equações e pontuação; considerar citações como parte dos textos e formas flexionadas das palavras como palavras distintas;
- χ) listagem e ordenação das palavras - o *software* produziu uma listagem em 3 colunas assim distribuídas: palavras, frequência em ordem decrescente de ocorrência e o *rank* (ordem de série) das palavras;
- δ) cálculos matemáticos e estatísticos - os dados foram plotados em planilhas utilizando-se o *Software Microsoft Office Excel 2003* e foram obtidos os seguintes produtos:  $r \times f = C$ , total de palavras, total de palavras distintas e *ranking* das nominalizações deverbais;
- ε) identificação dos padrões de nominalizações - foram investigados os 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)  $[X] v \rightarrow [ [X] v \text{-ncia} ] N$ ;
- φ) cálculo da produtividade das nominalizações - foi calculada a produtividade relativa dos padrões de formação de nominalizações deverbais, considerando-se a frequência relativa a cada nominalização;

- γ) controle de formas no singular e plural - a título de experiência foram somadas as frequências da nominalização publicação e sua forma no plural, publicações;
- η) aplicação da fórmula relativa ao Ponto de Transição (T) de Goffman - visa à identificação da região de transição, onde, segundo Goffman, encontram-se as palavras de alto conteúdo semântico, com probabilidade de serem utilizadas na indexação;
- ι) investigação das nominalizações - foram investigadas as nominalizações deverbais que se situam na região de concentração de palavras de alto conteúdo semântico;

Os resumos das cartas foram processados pelo *Software Rank Words 2.0*<sup>4</sup> e após, foi produzida uma lista de frequência de ocorrência de palavras, verificadas a aplicação das leis de Zipf e da Região de Transição de Goffman, a recorrência de nominalizações deverbais, as frequências e a relevância relativas de suas funções de índice.

## 5. RESULTADOS

Os resultados foram analisados por meio do quadro de frequência de palavras gerado pelo *software Rank Words* e pela aplicação da fórmula do Ponto de Transição de Goffman e delimitação da Região de Transição aos resultados desse quadro.

Esse quadro apresenta 3 colunas, contendo as seguintes informações: palavras, frequência em ordem decrescente e o *rank* (ordem de série) das palavras.

Encontrou-se no quadro o total de 1903 palavras, que foram retiradas do conteúdo do texto dos resumos das 100 cartas do arquivo privado de Bertha Lutz. Desse total de 1903 palavras, observou-se que 618 palavras foram repetidas de uma (1) a 237 vezes. Em média, cada palavra foi repetida 3 vezes.

---

4 Disponível em: <[http://download.cnet.com/Rank-Words3000-2279\\_4-10909564.html](http://download.cnet.com/Rank-Words3000-2279_4-10909564.html)>

O Ponto de T de Goffman foi calculado aplicando-se a fórmula matemática a seguir:

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

Aplicando-se a fórmula aos resultados encontrados no quadro do *Rank Words*, tem-se:

$$n = \frac{-1 + \sqrt{1 + (8 \times 405)}}{2}$$

$$n = \frac{-1 + \sqrt{1 + 3240}}{2} = \frac{-1 + \sqrt{3241}}{2} = \frac{-1 + 57}{2} = 56 = 27,96$$

O Ponto de Transição de Goffman ocorreu na frequência 27,96, sendo que a frequência 27 corresponde a LUTZ, sobrenome de Bertha Lutz, considerado de alto conteúdo semântico em relação à correspondência, já que essa é a autora das cartas científicas.

Na margem de frequências de 20 (*Rank* 15) a 237 (*Rank* 1), considera-se a região onde se concentram as palavras de alto conteúdo semântico (palavras-chave, descritores) tais como: publicações, Lutz, pedidos, informações, sapos e Bertha. Importante ressaltar que as cartas científicas eram utilizadas por Bertha Lutz para a publicação de informações sobre sapos, tema da pesquisa desenvolvida pela cientista.

Na Região de Transição de Goffman, foram encontradas duas nominalizações em *-ção*: publicações (*Rank* 6 = Frequência 33) e informações (*Rank* 11 = Frequência 26).

No *ranking* total de palavras, foram identificadas 40 nominalizações em *-ção*, 12 em *-mento* e 5 em *-ncia*, além de outras poucas formas que não foram objetos de análise desse estudo, como por exemplo *pedido* e *permuta*.

No extrato do quadro 1, a seguir, gerado pelo *Software RankWord 2.0* e *Software Microsoft Office Excel 2003*, pode-se observar a delimitação da



zona de concentração de palavras de alto conteúdo semântico, com as nominalizações em *-ção* destacadas em vermelho.

Rank	Word	Frequency	
1	de	237	
2	e	116	
3	do	39	
4	a	39	
5	envio	37	
6	publicações	33	
7	sobre	33	
<b>8</b>	<b>Lutz</b>	<b>27</b>	<b>Ponto T = frequência 27</b>
9	informações	26	
10	para	26	
11	pedidos	25	
12	das	25	
13	em	22	
14	sapos	21	
15	Bertha	20	

Quadro 1 Representação da Região de Transição de Goffman.

## 6. CONSIDERAÇÕES FINAIS

Verificou-se que o objetivo principal da pesquisa foi alcançado considerando-se válida a aplicação das referidas leis aos documentos de arquivo selecionados e analisados neste estudo, uma vez que os resultados apontaram uma zona de concentração de palavras de alto conteúdo semântico que podem ser utilizadas na indexação temática da correspondência da cientista, conforme formulação das leis.

Pode-se constatar também que a aplicação das leis a um número reduzido de cartas não modificou os resultados frequentemente obtidos por pesquisas que utilizam o processamento e análise de textos, fundamentadas nas abordagens teórico-metodológicas adotadas nesta pesquisa.

Observou-se que duas nominalizações deverbais em *-ção*, publicações e informações, exercem funções de índice mais relevantes na indexação temática, baseada em suas frequências de ocorrência ou de uso.

Finalmente, observa-se que os dados obtidos corroboram os objetivos estabelecidos e apontam para a importância das abordagens teóricas e descritivas da nominalização verbal, sob a perspectiva da análise lexicomorfológica do discurso em gêneros científicos, no contexto da Sociolinguística, para a indexação temática de cartas científicas, conforme já verificado anteriormente na análise do gênero artigo científico por Guedes (2010).

Recomenda-se a investigação em um número maior de cartas trocadas por outros cientistas que pesquisam nesse campo do conhecimento. Recomenda-se ainda que outros estudos sejam realizados com outros tipos de documentos arquivísticos de fundos específicos, para confirmar a viabilidade de aplicação das leis em material de arquivo.

Espera-se, com esse estudo, contribuir para a pesquisa na área de Bibliometria e obter, em estudos futuros, resultados que apontem para um maior grau de precisão na indexação de material de arquivo, precisamente em correspondências científicas.

## 7. REFERÊNCIAS

- Basílio, Margarida Maria de Paula. 2007. *Teoria lexical*, 8ª ed. São Paulo, Ática.
- Basílio, Margarida Maria de Paula. 2004. Polissemia sistemática em substantivos deverbais, *Ilha do Desterro*, 47: 49-71 [en línea]. Disponible en: <http://www.ilhadodesterro.ufsc.br/pdf/47%20A/basilio%20A.pdf>
- Basílio, Margarida Maria de Paula. 1999. A Morfologia no Brasil: indicadores e questões, *Delta*, 15: 53-70. [en línea]. Disponible en: <http://www.scielo.br/pdf/delta/v15nspe/4011.pdf>
- Bazerman, Charles. 2006. Cartas e a base social de gêneros diferenciados, en A. P. Dionísio e J. C. Hoffnagel, *Gêneros textuais, tipificação e interação*, São Paulo, Cortez.

- Bazerman, Charles. 1997. Discursively structured activities, *Mind, culture and activity*, 4: 296-308 [en línea]. Disponible en: <http://education.ucsb.edu/bazerman/articles/documents/5>. Bazerman1997 rtStructured.pdf
- Bazerman, Charles. 1988. Reporting the experiment: the changing account of scientific doings in the philosophical transactions of the royal society, en Charles Bazerman, *Shaping Written knowledge: the genre and activity of the experimental article in science*, Madison, Wisconsin, The University of Wisconsin Press [en línea]. Disponible en: [http://wac.colostate.edu/books/bazerman\\_shaping/chapter3.pdf](http://wac.colostate.edu/books/bazerman_shaping/chapter3.pdf)
- Bybee, Joan e Paul Hopper. 2001. *Frequency and the emergence of language structure*, Amsterdam, John Benjamins [en línea]. Disponible en: <http://books.google.com.br/books?id=6-bQUClxn1IC&pg=PA135&dq=e+books+B+ybee+linguistic+Morphology:+A+Study+of+the+Relation+between+Meaning+and+Form&cd=1#v=onepage&q=zipf&f=false>
- Garvey, William D. 1979. *Communication: the essence of science facilitating information among librarians, scientists, engineers and students*, Oxford, Pergamon Press.
- Guedes, Vânia Lisbôa da Silveira. 2012. A bibliometria e a gestão da informação e do Conhecimento científico e tecnológico: uma revisão da literature, *Ponto de acesso*, 6: 74-109 [en línea]. Disponible en: <http://www.portalseer.ufba.br/index.php/revistaici/article/view/5695/4591>
- Guedes, Vânia Lisbôa da Silveira. 2010. *Nominalizações deverbais em artigos científicos: uma contribuição para a análise e a indexação temática da informação*, Tese de Doutorado, Universidade Federal do Rio de Janeiro, Rio de Janeiro. Inédita.
- Guedes, Vânia Lisbôa da Silveira e Valois, Eliana Candeira. 1988. Adequação das Leis de Zipf (1ª e 2ª) e Ponto T de Goffman à indexação de documentos científicos: uma aplicação em Mecânica dos Solos (engenharia civil). 1988. Trabalho apresentado como requisito parcial para aprovação da disciplina de Bibliometria. Escola de Comunicação. Universidade Federal do Rio de Janeiro.
- Guedes, Vânia Lisbôa da Silveira e Suzana Borschiver. 2005. Bibliometria: uma ferramenta estatística para a gestão da informação e do conhecimento, em sistemas de informação, de comunicação e de avaliação científica e tecnológica, *VI Encontro Nacional de Ciência da Informação*, Salvador, UFBA [en

- línea]. Disponible en: [http://www.cinform.ufba.br/vi\\_anais/docs/VaniaLSGuedes.pdf](http://www.cinform.ufba.br/vi_anais/docs/VaniaLSGuedes.pdf)
- Halliday, Michael Alexander Kirkwood. 1994. *An introduction to functional grammar*. 2. ed. London, Edward Arnold.
- Hyland, Ken. 2009. *Academic discourse: english in a global context*, New York, Continuum International Publishing Group.
- Hyland, Ken e Françoise Salager-Meyer. 2008. Scientific writing, *Annual Review of Information Science and Technology*, 42: 297-338.
- Le Coadic, Yves-François. 2004. *A ciência da informação*, 2ª ed., Brasília, Briquet de Lemos.
- Meadows, Arthur Jack. 1999. Canais da comunicação científica, en A. J. Meadows, *A comunicação científica*, Brasília, Briquet de Lemos.
- Méndez Rodríguez, Eva Maria e José A. Moreiro González. 1999. Lenguaje natural e indización automatizada, *Ciencias de la Información*, 30: 11-24 [en línea]. Disponible en: <http://www.bib.uc3m.es/~mendez/publicaciones/articulos/indizacion99.pdf>
- Mikhaïlov, Aleksandr Ivanovich. 1984. Scientific communication, en Aleksandr Ivanovich Mikhaïlov y Arkadiï Ivanovich Chernyï y Rudzhero Sergeevich Giliarevskiï. *Scientific communication and informatics*, Arllington, Information Resources Press.
- Okubo, Yoshiko. 1997. *Bibliometric Indicators and Analysis of Research Systems: Methods and Examples*, Michigan, OECD.
- Pao, Miranda Lee. 1989. *Concepts of information retrieval*. Englewood: Libraries Unlimited.
- Pao, Miranda Lee. 1978. Automatic text analysis based on transition phenomena of word occurrences, *Journal of the American Society for Information Science*, 29: 121-124.
- Santos, Maria José Veloso da Costa. 2009. Correspondência científica de Bertha Lutz: um estudo de aplicação da lei de ZIPF e Ponto de Transição de Goffman em um arquivo pessoal, *Ponto de Acesso*, 3: 317-326 [en línea]. Disponible en: <http://www.portalseer.ufba.br/index.php/revistaici/article/view/3628/2751>.
- Swales, John M. 1990. *Genre analysis: English in academic and research settings*, Cambridge, Cambridge University Press.
- Swales, John M. 1986. Citation analysis and discourse analysis, *Applied Linguistics*, 7: 39-56.

Ziman, John Michael. 1969. Information, communication and knowledge, *Nature*, 224: 318-324.

Zipf, George Kingsley. 1949. *Human behavior and the Principle of Least Effort*, Cambridge, Addison-Wesley.

Zipf, George Kingsley. 1935. *The Psycho-Biology of Language*, Boston, Massachusetts, Houghton Mifflin.