

The academic discourse in scientific letters abstracts from the personal archives of the Brazilian scientist Bertha Lutz: a bibliometric and linguistic study of deverbal nominalization in the subject indexing.

Vânia Lisbôa da Silveira Guedes¹ and Maria José Veloso da Costa Santos²

¹vanialisboa@facc.ufrj.br e ²msantos1402@gmail.com

Federal University of Rio de Janeiro, Av. Pasteur, 250- Fundos - Anexo do CCJE - Urca
Cep: 22290-902, Rio de Janeiro (Brazil)

Introduction

The letters had a crucial role for communication in science. Used for knowledge transmission and information diffusion, were changed among the researchers and scientists to show original ideas for being discussed, judged and evaluated by them.

Bazerman (2006) believes that the letters contributed to the emergence of academic genre. The scientific paper genre, he said, emerged from the correspondence of the German Henry Oldenburg, secretary of the Royal Society of London, with other scholars. The earliest numbers from the Philosophical Transactions of the Royal Society of London were published like abstracts of correspondence and meetings of the Royal Society.

Bibliometric analysis is an important instrument to establish indicators in an area of knowledge because it presents the quantitative aspects of production, dissemination and use of scientific information recorded.

The nominalization, according to Basilio (2007), refers to the set of processes that form nouns from adjectives, and especially from verbs. The author explains that the nominalization contains aspects syntactic and semantic textual and play functions of designation of process, action, state etc. Basilio (2007), Swales (1990), Hyland (2009) and other authors emphasize that the use of nominalizations characterize heavily the formal writing discourse and particularly the scientific discourse.

This study consists of a bibliometric and a linguistic analysis of the academic discourse of scientific letters abstract of the Brazilian scientist Bertha Lutz (1894-1976) private file. It investigates the recurrence of deverbal nominalizations in the lexicon of these abstract in Brazilian Portuguese language in order to verify their relevant subject index functions in the information indexing.

The specific research question to be answered is how a systematic analysis of deverbal nominalizations in abstract academic genre of scientific letters can contribute to the semi-automatic subject indexing based on Bibliometrics, making information retrieval electronic systems more precise, intelligent and scientifically established.

Objectives

Central Objective

The aim of this study is the establishment of patterns to developed semiautomatic subject indexing of the academic discourse of scientific letters abstracts by applying bibliometric and linguistic analysis

Specific Objectives

(a) to analyze scientific letters abstracts discourse based on quantitative models used in automatic indexing within Bibliometrics in Information Science;

(b) to contribute to the theoretical and practical approaches to information indexing;

(c) to strengthen the links of interface of the Linguistics with Bibliometrics in the Information Science.

Hypothesis

The hypothesis raised in this study is that deverbal nominalization described by $[X] v \rightarrow [[X] v -\text{ção}] N$ is predominant in the lexicon of the scientific letters abstracts studied as the most important index functions based on their frequencies of occurrences.

Theoretical framework

Bibliometrics is the science that presents a set of empirical laws and principles based on mathematical and statistical methods to investigate, assess and quantify the written communication processes. Among the various and more used bibliometrics laws there are the two laws of Zipf used for information indexing related to frequency of occurrence of words in a given text: Zipf's First Law of words of high frequency of occurrences and Zipf's Second Law of words of low frequency of occurrences. Zipf's Second Law, modified by Booth (1967), was enriched with the Goffman Transition (T) Point, a method of selecting index terms directly from a word frequency suggested by Goffman (apud Pao, 1978). This method indicates a region from the list of words used in a scientific text that likely to concentrate most thematically significant words of a scientific text. Goffman T Point is represented mathematically by the expression above, where n represents the Point T; I and i is the number of words that has a frequency 1.

$$\text{Guedes (2010)} \quad n = \frac{-1 + \sqrt{1 + 8I_1}}{2}$$

demonstrate the importance of morphological productivity analysis of nominalizations deverbal in academic papers genre, showing especially that the nominalizations in $-\text{ção}$ presents highest density of information, highlighting most appropriate processes for the subject indexing within controlled language electronic systems in Information Science.

Therefore, the theoretical framework used was the information subject indexing, Zipf's Laws and Goffman Transition Point (Pao, 1978) in the context of bibliometrics, as well as the textual genre analysis (Bazerman, 2006; Hyland, 2009; Swales, 1990) and the lexical theory (Basílio, 2007; Chomsky, 1970) within sociolinguistics.

Methodology

The study *employed a bibliometric analysis* based on Zipf's Laws and Goffman T Point related to information indexing. This method points to the region of the words frequency list that the most content-bearing words of a given text would occur. Then, in this region, nominalizations in $-\text{mento}$ and $-\text{ção}$ were investigated, based on lexical theory.

The methodology consisted of the following steps:

(1) sample definition - were selected in correspondence of Bertha Lutz, Series Scientific,) abstracts of 100 (one hundred) earliest letters, available in Santos (2009). These abstracts were compiled into a single text, a text considered in this study "long enough" to be analyzed according to Zipf and Goffman;

(2) count of words - these abstracts were processed by *Software Rank Words 2.0* available in http://download.cnet.com/Rank-Words3000-2279_4-10909564.html.

(3) listing and ranking of words - the software produced a table in 3 columns, distributed as follows: words, order of decreasing frequency and the rank of words;

(4) applying Zipf's Laws and the formula of Goffman T Point - it was investigated the Zipf's Laws and Goffman T Point adequacy to identify the frequency, where the transition from high frequency words for low frequency occurs;

(5) defining Goffman T region - it was identified in the ranking the region that concentrates the most content-bearing words that should be used for indexing and

provide greater precision in information retrieval systems;

(6) verification of nominalizations recurrence in *-ção* and *-mento*, within Goffman T region, and their index terms functions.

Conclusion

The results found in this study point to recurrence of nominalizations in *-ção* with high semantic content in Goffman Transition Region and confirm the established hypothesis and also the importance of theoretical and descriptive approaches of deverbal nominalization and bibliometrics for the subject indexing of scientific letters abstracts within information research systems.

Finally, the recurrent pattern of deverbal nominalization formation in *-ção* contributes to the knowledge of lexical-morphological features more productive in scientific texts, making it of great importance to the process of analyzing and identifying the information content of scientific letters.

References

- Basilio, M. M. de P. (2007). *Teoria lexical*. 8. ed. revista e atualizada. São Paulo: Ática.
- Bazerman, C. (2006). *Gêneros textuais, tipificação e interação*. São Paulo: Editora Cortez.
- Booth, A. D. A (1967) "Law" of occurrences for words of low frequency, *Information and Control*, 10, p. 386-393.
- Chomsky, N. (1970) Remarks on nominalization. In: R. Jacobs & P. Rosenbaum (Org). *Readings in English transformational grammar*. Waltham, MA: Ginn.
- Guedes, V. L. da S. (2010). *Nominalizações deverbais em artigos científicos: uma contribuição para a análise e a indexação temática da informação*, 2010. Tese (Doutorado em Linguística) – UFRJ, Rio de Janeiro.
- Hyland, K. (2009) *Academic discourse: english in a global context*. New York: Continuum International Publishing Group.
- Pao, M. L. (1978). Automatic text analysis based on transition phenomena of word occurrences, *JASIS*, 29, p. 121-124.
- Santos, M. J. V. da C. (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, p.317-323.
- Swales, J. M. (1990). *Genre analysis: English in academic and research settings*. Cambridge: Cambridge University Press.