

Análise léxica automatizada em administração de empresas

JOÃO CUSTÓDIO PINTO

UNINOVE – Universidade Nove de Julho
joaocustodiopinto2013@gmail.com

MARCOS ROGÉRIO MAZIERI

UNINOVE – Universidade Nove de Julho
m_mazzieri@hotmail.com

LEONARDO VILS

UNINOVE – Universidade Nove de Julho
vilsleo@gmail.com



ANÁLISE LÉXICA AUTOMATIZADA EM ADMINISTRAÇÃO DE EMPRESAS

Resumo

A pesquisa acadêmica é um processo que origina a produção do conhecimento. O conhecimento produzido tende a se propagar por meio de difusão orientada à comunidade acadêmica e pesquisadores, evidenciando a importância da pesquisa e o seu produto final, o conhecimento. A produção literária realizada na pesquisa acadêmica, se formaliza através da divulgação escrita, variados recursos e símbolos para suportar o conteúdo produzido. Neste contexto, a linguagem natural se situa como meio de difusão do pensamento, sensações, valores e ideias da sociedade. Em uma revisão de literatura como metodologia, apresentamos uma retrospectiva dos fundamentos que originaram o desenvolvimento da ferramenta computacional denominado IRaMuTeq (*Interface de R pour les Analyses Multidimensionnelles de Textes et de Questionnaires*), como objetivo de identificar os aspectos implícitos da lexicometria, mas que podem auxiliar a melhoria da interpretação dos resultados das análises de conteúdo. Os resultados sugerem a existência de oito aspectos implícitos relevantes nos artigos publicados entre 1945 e 2016 sobre lexicometria. Os achados contribuem para a melhoria das interpretações das análises de conteúdo para os pesquisadores da área de administração de empresas.

Palavras-chave: Lexicografia, Produção do Conhecimento, Análise Semântica, Análise de Similitude, Estatísticas Textuais Clássicas.

Abstract

Academic research is a process that originates the production of knowledge. The knowledge produced tends to propagate through diffusion oriented to the academic community and researchers, highlighting the importance of research and its final product, knowledge. The literary production carried out in academic research is formalized through written dissemination, varied resources and symbols to support the content produced. In this context, natural language it is a means of diffusion of thought, sensations, values and ideas of society. In a review of literature as methodology, we present a retrospective of the foundations that led to the development of the computational tool called IRaMuTeq (*Interface de R pour les Analyses Multidimensionnelles de Textes et de Questionnaires*), in order to identify the implicit aspects of lexicometry, but which can to help improve the interpretation of content analysis results. The results suggest the existence of eight implied relevant aspects in the articles published between 1945 and 2016 on lexicometrics. The findings contribute to the improvement of the interpretations of content analysis for researchers in the area of business administration.

Keywords: Lexicography, knowledge production, Semantic Analysis, Similarity analysis, Classical Textual Statistics.



1 Introducao

A utilizacao de software para analise de conteudo dos *corpus* textuais tem se mostrado uma pratica crescente no campo da descoberta do conhecimento (Camargo, B. V., & Justo, A. M. 2013). Neste contexto, apresenta-se o *software* IRaMuTeQ (*Interface de R pour les Analyses Multidimensionnelles de Textes et de Questionnaires*), desenvolvido pelo professor Pierre Ratinaud, do laboratorio de estudos e pesquisas em ciencias sociais aplicadas (LERASS), da Universidade de Toulouse. O IRaMuTeQ, tem como principio a lexicometria, possibilitando analises lexicais como, por exemplo, o agrupamento hierarquico descendente (Reinert, 1991) e analise de similaridade de segmentos de textos (Benzecri, 1973).

O software IRaMuTeQ, desenvolvido originalmente em lingua francesa, e ancorado pelo software R, que tambem e de distribuicao gratuita, vem sendo utilizado no Brasil desde 2013. Existe atualmente em seu banco de dados, um dicionario experimental na lingua portuguesa. O IRaMuTeQ possibilita a construcao de analises textuais como, estatisticas textuais classicas, classificacao hierarquica descendente, pesquisa de especificidade de grupos de variaveis, analise de similitude e nuvem de palavras, sendo o seu rigor estatistico uma das caracteristicas que o torna um aplicativo de relevancia no tratamento das informacoes analisadas. Todavia, ha necessidade de preparacao previa dos *corpus* textuais a serem analisados. Devido a simplicidade de sua interface, apresenta-se como uma ferramenta para auxiliar na pesquisa de conteudo automatizado em administracao de empresas.

O uso de softwares para auxilio nas analises de dados nao pode ser confundido com o procedimento metodologico da pesquisa (Camargo, B. V., & Justo, A. M. 2013). Da mesma forma, o relatorio de saida de um software de analise nao e *per se*, suficiente para comunicar resultados (Cross, M. C., & Hohenberg, P. C. 1993). Desta forma qualquer software de analise deve ser posicionado como uma ferramenta de apoio a pesquisa, cujos pressupostos e relatorios de saida precisam estar fundamentados na teoria do campo em estudo. Especialmente o relatorio de saida, precisa ser interpretado, usando as tecnicas de triangulacao (Yin, R. K. 2015), contraste, tabelas de contingencia nao exaustivamente.

Neste contexto, emergem questoes relativas ao funcionamento dos algoritmos de analise usados nos softwares de analise, em especial, neste caso, o IRaMuTeQ. O objetivo deste Relato Tecnico e submeter ao IRaMuTeQ, as principais literaturas que definiram os conceitos de analise Lexica, visando identificar quais os elementos originais constituintes destas literaturas, podem ser identificados.

O procedimento tem contribuicoes teoricas, uma vez que proporciona a verificabilidade do desenvolvimento da lexicografia ao longo do tempo, dos elementos originais da lexicografia, mostrando aqueles elementos que permanecem em uso e aqueles elementos que nao foram implantados na automatizacao.

Este relato tecnico, tambem oferece contribuicoes praticas, uma vez que foi possivel identificar limitacoes nao documentadas do IRaMuTeQ, propiciando aos pesquisadores compreender a aplicabilidade dos resultados que venham alcançar por meio de analise automatizada em pesquisas futuras, melhorando sua capacidade de interpretacao de resultados.

Referencial Teorico

A construcao social do significado a partir da analise semantica esta presente nos fundamentos dos estudos lexicos metricos. A lexicometria, constitui-se de organizacao e sumarizacao das estruturas dos textos, decompostas em palavras e as distancias entre estas palavras, com o objetivo de simplificacao e sumarizacao. Com a analise da estrutura de um



texto, 3 poss3vel identificar um conjunto de ocorr3ncias e co-ocorr3ncias de palavras e este conjunto de informa3o3es l3xicas s3o mensur3veis. Estas mensura3o3es s3o realizadas utilizando-se do m3todo estat3stico para a an3lise de texto em seus lexemas co-ocorrentes, existente nos textos segmentados (Achard, P. 1997).

A distribui3o estat3stica das frequ3ncias de palavras em um determinado texto, como uma lei, foi objeto de estudo h3 mais de um s3culo, conforme estudos realizados por Jean Baptiste Estoup em 1916 (Lelu, A. 2014). As pesquisas no campo da lingu3stica se originam nas observa3o3es do comportamento existente nos sistemas textuais de natureza variada encontradas na lingu3stica. O estado desta observa3o3e 3 definidor para formular uma lei de natureza emp3rica, tendo como base o relacionamento de classifica3o3o not3vel na lingu3stica, a lei George Kingsley Zipf. A lei de Zipf prop3e-se a confirmar esta particularidade (Zipf, G. K. 1945).

A estimativa de Zipf nas observa3o3es realizadas no campo da psicologia comportamental, definem a dimens3o, import3ncia e frequ3ncias de elementos em uma lista ordenada de palavras. A partir deste ponto tem-se uma lei de pot3ncia sobre a distribui3o de valores, estruturados em uma ordem presente em uma lista de palavras, sendo que desta lista o membro n teria uma rela3o3o de valor com o 1º da lista segundo $1/n$.

A frequ3ncia de palavras ou outros recurso presentes na palavra, 3 definidora de algumas leis aplicadas na lingu3stica, a lei de signific3ncia e significado, que define a tend3ncia para que as palavras sejam mais poliss3micas; a lei da abreviatura situa3o3e em que as palavras mais frequentes tornam mais curtas.

No ponto em que Zipf prop3e com sua lei uma abordagem estat3stica para o campo da lingu3stica, os estudos avan3am para an3lise do discurso (Harris, Z. S., & Dubois-Charlier, F.1969), an3lise de correspond3ncia (Benz3cri, J. P. 1973), an3lise do conte3do (Bardin, 1977), an3lise de probabilidades (Peirce, C. S. 1978). O IRaMuTeQ 3 um software de an3lise l3xica e, portanto, n3o se prop3e a analisar o discurso.

A partir do ano de 1983 iniciam-se estudos para automatizar os algoritmos de an3lise l3xica. Para tanto, definiram-se algumas aplica3o3es que fundamentaram a cria3o3o do IRaMuTeQ. Inicialmente utilizando o m3todo de classifica3o3o hier3rquica descendente (Reinert, M. 1987), foi lan3ado o software em sua primeira vers3o3e e que atualmente apresenta-se na vers3o3e IRaMuTeQ 0.7 alpha 2. Conhecer as capacidades e os limites dos conte3dos textuais, podem auxiliar para a compreens3o3o do sentido epistemol3gico presente na pesquisa cient3fica, e com a produ3o3o de pesquisa potencializa-se a capacidade humana na constru3o3o e desenvolvimento do conhecimento, gerando novas perspectivas para a inven3o3o e inova3o3e, e sedimentando o caminho do aprendizado (Bregonje, M. 2005).

No campo das Ci3ncias Humanas e Sociais, a utiliza3o3o de softwares de an3lises l3xicas tem se firmado como ferramenta para administrar situa3o3es din3micas presentes na produ3o3o de pesquisas cient3ficas (Cross, M. C., & Hohenberg, P. C. 1993).

A an3lise l3xica usada na pesquisa acad3mica, possibilitada pelo uso do software IRaMuTeQ, pode ser 3til ao pesquisador, sendo que sua aplicabilidade consistida em plataforma simples, faz quantifica3o3es de dados, realiza c3lculos estat3sticos a partir da an3lise de textos (Benz3cri J.P. 1973).

As an3lises se iniciam ap3s tratamento do corpus textuais, consistido em um processo de analisar a entrada de linhas de caracteres, como, por exemplo, o c3digo-fonte de um programa de computador, em seguida, produz uma sequ3ncia de s3mbolos chamado s3mbolos l3xicos, que podem ser manipulados mais facilmente por um leitor de sa3da, a an3lise de conte3do se consolida pelo m3todo qualitativo no sentido de identificar as especificidades de um dado corpus de informa3o3e escrita (Lahlou, S. 1994).



Procedimentos Metodológicos

Na realização deste levantamento bibliográfico foram consultadas, nos meses de junho e julho de 2017, referências sobre a temática da análise léxica automatizada em pesquisa aplicada, na base de dados Scopus e Web Of Science. Foram utilizados os seguintes descritores: “automated and lexical and analysis and in and applied and research”, e seus respectivos correspondentes na língua portuguesa. A abordagem aplicada é qualitativa e foi realizada sob a forma de revisão narrativa de literatura, em sua etapa inicial. Após selecionar os artigos que tratam do tema da lexicometria, foram realizadas as análises de conteúdo dos resumos destes artigos. Seguindo o objetivo deste relato, a análise de conteúdo foi automatizada pelo IRaMuTeQ. Não se trata de uma meta análise, uma vez que não foram revisitados os dados usados para a concepção de tais artigos. Trata-se de uma análise dos resumos dos artigos que tiveram alguma influência da concepção dos próprios algoritmos usados para construir o IRaMuTeQ.

Desta forma, ao analisar os conteúdos dos resumos dos artigos que geraram os conceitos que, hipoteticamente, deveriam estar presentes no próprio IRaMuTeQ, foi possível identificar quais os elementos originais permanecem presentes e podem ser identificados pela análise automatizada. Justifica-se esta metodologia, por permitir estabelecer relações entre os elementos conceituais das produções anteriores, que definiram os conceitos de análise lexicométrica, com os próprios resultados obtidos pela ferramenta de software que se originou destes conceitos (no caso o software IRaMuTeQ).

No quadro 1, a seguir, estão apresentados os artigos cujos resumos foram analisados. Foram incluídos na amostra artigos de 1945 até 2016, que responderam a palavra chave de busca apresentada nesta seção de procedimentos metodológicos.



Autor	Ano	Titulo
Zipf, G.K.	1945	The meaning-frequency relationship of words. Journal of General Psychology,
Berger, E.,	1966	Problèmes de linguistique générale
Harris, Z. S., & Dubois-Charlier, F.	1969	Analyse du discours
Benzécri J.P.	1973	L'Analyse des Données, 2, L'analyse des correspondances
Peirce, C. S.	1978	Notes on the doctrine of chances.
Reinert M	1983	Une méthode de classification descendante hiérarchique
Reinert, M	1987	Classification Descendante Hierarchique et Analyse Lexicale par Contexte- Application au Corpus des Poesies D'A. Rihbaud
Prather, R.E.	1988	Comparison and extension of theories of zipf and halstead.
Cros, M	1993	Les apports de la linguistique: langage des jeunes et sida. In ANRS (Agence Nationale de Recherche sur le sida). Les jeunes face au sida: de la recherche à l'action
Cross, M C., & Hoehnerberg, P. C.	1993	Pattern formation outside of equilibrium
Lahlou, S.	1994	L'analyse lexicale. Variances,
Noel-Jorand, MC., Reinert, M, Bonnor	1995	Discourse analysis and psychological adaptation to high altitude hypoxia Stress Medicine
Achard, P., et all,	1996	Rapport Slade, Sociologie du Langage, Analyse de Discours, Énonciation, présenté au CNRS pour la constitution
Lapalut, S.	1996	Text clustering to help knowledge acquisition from documents Lecture Notes in Computer Science
Groussier, M L.	2000	On Antoine Culioli's theory of enunciative operations Lingua
Bregonje, M	2005	Patents: A unique source for scientific technical information in chemistry related industry? World Patent Information.
Derycke, M	2007	En chemin avec Pierre Achard: Perspectives sur réévaluation et sur les Fondements de sémantique discursive
Leimdorfer, P.F.	2007	Actualité de la sociologie du langage de Pierre Achard
Normand, C.	2011	langue, parole, sujet chez saussure et berger
Marchand, P., & Ratinaud, P.	2012	L'analyse de similitude appliquée aux corpus textuels: les primaires socialistes pour l'élection présidentielle française.
Ratinaud, P., Marchand, P.	2012	Recherche improbable d'une homogène diversité: Le débat sur l'identité nationale.
Camargo, B. V., & Justo, A. M	2013	IRAMUTEQ: um software gratuito para análise de dados textuais
Marty, E., Marchand, P., Ratinaud, P.	2013	Les médias et l'opinion- Eléments théoriques et méthodologiques pour une analyse du débat sur l'identité nationale
Lelu, A	2014	Jean-Baptiste Estoup and the origins of Zipf's law: A stenographer with a scientific mind (1868-1950)
Yin, R. K.	2015	Estudo de Caso-: Planejamento e Métodos
Benjafield, J.G.	2016	The digital history of the anglophone vocabulary of psychology an exploration using Zipfian methods History of Psychology
Lerat, P.	2016	Technique, langue, traduction. Il y a sertissage et sertissage Hermes

Quadro 1 - Relação de artigos sobre lexicometria analisados

Fonte: Autores, 2017

Pode-se verificar que a definição da lei de Zipf em 1945 favoreceu o desenvolvimento dos estudos de lexicometria. Especialmente em 1973, com os estudos de Benzécri, sobre a análise fatorial de correspondência e em 1991 com os estudos de Reinert sobre a classificação hierárquica descendente foram definidas as bases matemáticas complementares. Neste relato técnico, serão usadas as duas visões (análise fatorial de correspondência e a classificação hierárquica descendente para analisar os resumos dos artigos, subdivididos em quatro classes (classes 2 e 3 - autores e descrição do fenômeno; classes 1 e 4 - processos de análises e bases cognitivas, respectivamente), da figura 1.



Resultados

A análise de conteúdo dos resumos dos artigos mais citados sobre análise de conteúdo e lexicometria, mostrados no quadro 1, resultou na descoberta de aspectos explícitos e implícitos sobre o tema. Foram levantadas quatro classes semânticas, de acordo com a classificação hierárquica descendente (Reinert, 1991), apresentadas na Figura 1. A Figura 2, apresenta as mesmas classes, no entanto, na visão da análise fatorial de correspondência (Benzécri, 1973). Os aspectos encontrados serão detalhados a seguir.

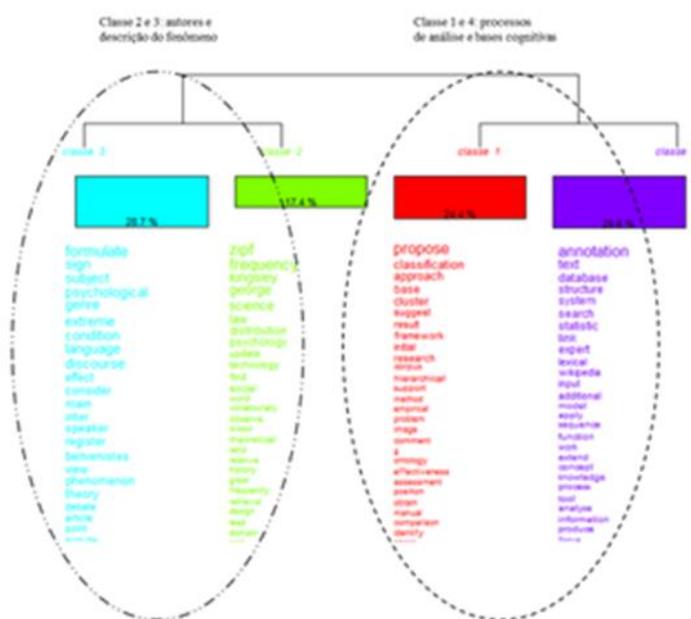


Figura 1 - Classificação hierárquica descendente dos resumos dos artigos do Quadro 1.
Fonte - elaborada pelos autores

Bases Cognitivas (anotações, textos e bancos de dados): O conteúdo textual é o elemento essencial para as análises lexicográficas clássicas, constituindo-se como o banco de dados para o tratamento do que se chama *corpus* de texto. *Corpus* de texto formam os bancos de dados submetidos à análise do software (Lahlou, 1994). O conteúdo apresentado no resultado desta análise sugere que a base do conteúdo e forma dos *corpus* textuais analisados constituem-se de formas expressivas do pensamento humano, associadas à natureza semântica e os meios utilizados para difusão da produção escrita (Bardin, 1977).

Trajectoria Científica (psicologia, tecnologia e linguística): Na medida em que as pesquisas desenvolvem as bases teóricas para os campos da psicologia, linguística e tecnologia, desenvolve-se a sustentação da análise de conteúdo, em termos relacionados à linguística e psicologia, em um primeiro momento (Zipf, 1945; Harris, et al., 1969), e posteriormente se confirma com a tecnologia que dá origem ao desenvolvimento de algoritmos para análise fatorial de correspondência, incluída nos softwares de análise (Benzécri, 1973).



Autores (George Kingsley Zipf): As análises lexicográficas clássicas são objeto de estudos realizados por Jean Baptiste Estoup em 1916, e se constitui no fundamento para a pesquisa e desenvolvimento dos estudos de frequência de palavras e demais atributos presentes na palavra. É neste contexto que se criam as condições definidoras de algumas leis aplicadas à linguística (Zipf, 1945).

Método e Desempenho (proposição, classificação e Interpretação): A análise dos *corpus* textuais se consolida nos relatórios de saída dos softwares de análise de conteúdo. Tal relatório apresenta as possíveis visões das análises de conteúdo como a análise de similitude, baseada na teoria dos grafos (utilizada com frequência por pesquisadores no campo de pesquisa de ciências sociais). Por esta análise é possível identificar a co-ocorrência existente entre as palavras, e assim permitir a análise de vínculo entre as palavras existentes nos *corpus* textuais e as variáveis categoriais que as descrevem (Marchand & Ratinaud, 2012). No método da classificação hierárquica descendente (CHD), a análise visa obter as classes de segmento de textos que atendam ao mesmo tempo a apresentação de vocabulário semelhante entre si, e vocabulário diferente dos segmentos das outras classes. A partir deste ponto o software realiza análise dos dados apresentado no formato de um dendograma, que são representações das relações entre as classes, fornecendo resultados que permitem uma descrição de cada uma das classes, que são fundamentadas no vocabulário léxico e suas variáveis (Reinert, M. 1987). O software de análise de conteúdo fornece também a análise fatorial de correspondência, realizada a partir da CHD. Baseando-se nas classificações escolhidas, o software processa o cálculo, fornecendo em seguida os segmentos de textos mais presentes em cada uma das classes, possibilitando desta forma uma análise de contextualização do vocabulário típico de cada classe, apresentados sobre um eixo bidimensional, cujas distâncias e posicionamentos entre as palavras, representam os conceitos correlatos (mesmo quadrante) ou opostos (quadrantes diferentes). Na análise Fatorial de Correspondência a associação dos textos com suas variáveis, possibilita a análise da produção textual em razão das variáveis que os caracterizam. Nesta situação, fornece uma análise fatorial de correspondência, para as variáveis com no mínimo de três modalidades. Finalmente, o software de análise de conteúdo oferece a visão da nuvem de palavras. A nuvem de palavras é o agrupamento de palavras apresentado em função da frequência de ocorrência, em uma formatação gráfica. Trata-se de uma análise mais simples em termos lexicais, porém, tem sua apresentação gráfica funcional e representativa do *corpus* textual analisado (Lapalut, 1996).

Elementos de análise (distribuição, frequência e vocabulário): Na análise dos *corpus* textuais, situação em que a unidade de texto é identificada transformando-se o texto completo em segmento de texto, ocorre a identificação da quantidade de palavras, estabelecendo se a frequência média, o hápax (palavras que tem frequência de apresentação igual a 1), e demais decomposições léxicas. Ao decompor o texto original em segmentos e os segmentos em palavras devidamente classificadas (adjetivos, substantivos, verbos, não exaustivamente) evidencia-se o vocabulário que compõe o texto original e conseqüentemente as suas métricas. As métricas são originadas da redução das palavras para suas estruturas raízes (radicais),



criando o dicionário de formas reduzidas e identificando as formas ativas e suplementares (Zipf, 1945).

Processos de análise (agrupamento e classificação): Os segmentos de textos e as classes de palavras são tratadas em razão de uma classificação realizada pelo software de análise de conteúdo quanto a distribuição do vocabulário. Em pesquisas realizadas no campo da linguística foram indicados que estas classes são interpretadas como campos lexicais (Cross, M. C., & Hohenberg, P. C.1993);

Fundamentação Teórica (discurso, fala: linguagem natural): A linguagem natural é elemento presente em todas as manifestações presentes no meio difusivo da comunicação, utilizada como portadora de conteúdo, a linguagem natural evoca o pensamento, sensações, valores e ideias da sociedade, e nesta trajetória tem se elementos presentes do conhecimento e padrões culturais, que são registros manifestos na produção textual como produto para análise e são definidas por agrupamentos baseados em frequência (clusters) e classificações baseadas em algoritmos (classes) (Lahlou, 1994).

Explicações do Fenômeno (linguagem e aspectos psicológicos): A linguagem se manifesta nas expressões do cotidiano em uma sociedade multifacetada, deste contexto de relações onde a linguagem se afirmar como portadora de significados, emergem aspectos da natureza psicológica da sociedade, e são caracterizadas em atitudes, comportamentos e simbologias sociais (Cross, M. C., & Hohenberg, P. C. (1993).

Na figura 2, são apresentados os aspectos da classificação hierárquica descendente numa visão bidimensional, por meio da análise fatorial de correspondência.

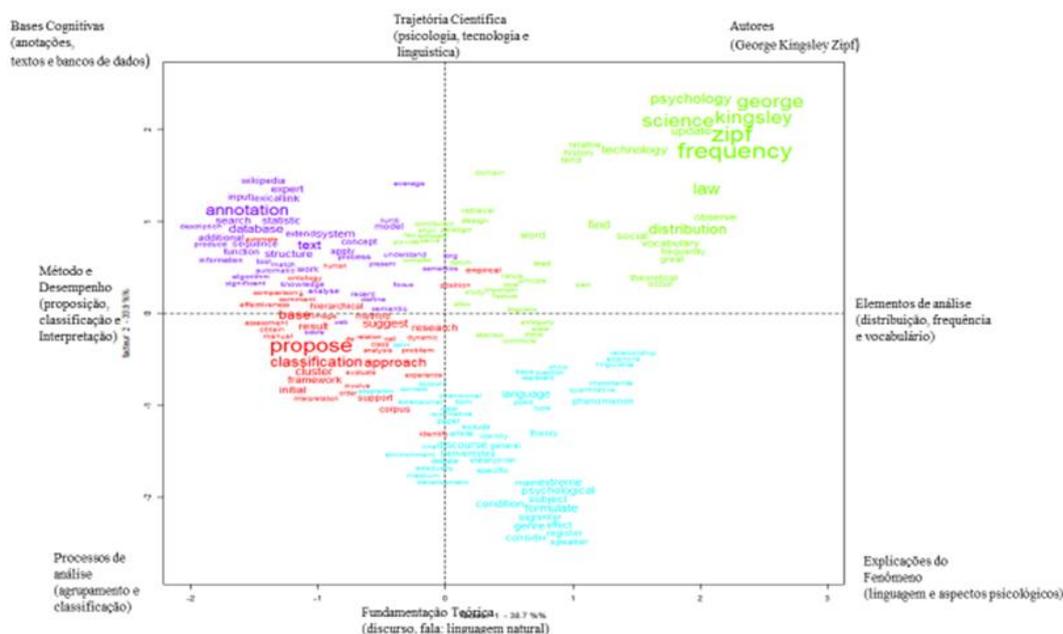


Figura 2-Análise fatorial de correspondência dos resumos dos artigos indicados no Quadro 1.
Fonte: autores, 2017



CONCLUSÕES/CONSIDERAÇÕES FINAIS:

A atividade de pesquisa se conforma no ato de conhecer as capacidades e limites do pensamento, que são as bases de fundamento da epistemologia, a atividade de pesquisa é geradora de conhecimento que se consolida no curso do tempo. Boa parte da pesquisa científica é levada a efeito quando comunicada na forma escrita. Parece relevante compreender que as Bases Cognitivas, representam um dos oito aspectos levantadas na análise dos resumos dos artigos que tratam do tema. As bases cognitivas em questão são os próprios registros da pesquisa quando divulgada na forma escrita. A relevância de George Zipf para o campo da análise lexicográfica, semântica e de conteúdo, merece destaque. A lei de Zipf demonstrou ser o impulsionador dos estudos do campo da análise de conteúdo. Ainda que as contribuições de Reinert (1991) ou de Benzécri (1973) sejam matematicamente fundamentais, observa-se nos resultados que foi George Zipf (1945) o autor que recebeu maior relevo, em termos de citações.

Do ponto de vista gerencial e prático, ao destacar os oito aspectos encontrados (base cognitiva, trajetória científica, autores, elementos de análise, explicações do fenômeno, fundamentação teórica, processo de análise e método e desempenho) é possível argumentar que mesmo os bancos de dados não estruturados ou que recebem *inputs* em forma de *posts* em redes sociais por exemplo, podem ser considerados bases cognitivas e propensas a análise.

Ao elencar os aspectos envolvidos nos resumos dos artigos sobre lexicometria e análise de conteúdo, percebem-se as partes implícitas aos usuários dos softwares de análise de conteúdo. Os pesquisadores da área de administração de empresas tendem a ser mais assertivos em sua análise qualitativas, uma vez identificados e compreendidos os oito aspectos indicados neste relato técnico. O aprofundamento das análises também parece ser possível, uma vez que, ao analisar determinados *corpus* de texto, tem-se alguma sistematização (não linear, é verdade) na forma de pensar a análise de conteúdo aplicadas as respectivas pesquisas.



Referências

- Achard, P. (1997). L'engagement de l'analyste à l'épreuve d'un événement. *Langage et société*, 79(1), 5-38.
- Bardin, L., & Bardin, L. (1977). *L'analyse de contenu* (No. Sirsi) a456144).
- Benzécri, J. P. (1973). *L'analyse des données* (Vol. 2, p. 1). Paris: Dunod.
- Bregonje, M. (2005). Patents: A unique source for scientific technical information in chemistry related industry? *World Patent Information*, 27(4), 309-315.
- Camargo, B. V., & Justo, A. M. (2013). IRAMUTEQ: um software gratuito para análise de dados textuais. *Temas em Psicologia*, 21(2), 513-518.
- Cross, M. C., & Hohenberg, P. C. (1993). Pattern formation outside of equilibrium. *Reviews of modern physics*, 65(3), 851.
- Harris, Z. S., & Dubois-Charlier, F. (1969). Analyse du discours. *Langages*, 8-45.
- Lahlou, S. (1994). L'analyse lexicale. *Variances*, (3), 13-24.
- Lapalut, S. (1996). Text clustering to help knowledge acquisition from documents. *Advances in Knowledge Acquisition*, 115-130.
- Lelu, A. (2014). Jean-Baptiste Estoup and the origins of Zipf's law: a stenographer with a scientific mind (1868-1950). *Boletín de Estadística e Investigación Operativa*, 30(1), 66-77.
- Marchand, P., & Ratinaud, P. (2012). L'analyse de similitude appliquée aux corpus textuels: les primaires socialistes pour l'élection présidentielle française (septembre-octobre 2011). *Actes des 11eme Journées internationales d'Analyse statistique des Données Textuelles. JADT, 2012*, 687-699.
- Peirce, C. S. (1978). Notes on the doctrine of chances. In *Dispositions* (pp. 237-245). Springer Netherlands.
- Ratinaud, P., & Marchand, P. (2012). Recherche improbable d'une homogène diversité: le débat sur l'identité nationale. *Langages*, (3), 93-107.
- Reinert, M. (1987). Classification Descendante Hiérarchique et Analyse Lexicale par Contexte-Application Au Corpus des Poésies D'A. Rimbaud. *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique*, 13(1), 53-90.
- Toledo, J. A. D., & Rodrigues, M. C. (2017). Teoria da mente em adultos: uma revisão narrativa da literatura. *Boletim-Academia Paulista de Psicologia*, 37(92), 139-156.
- Yin, R. K. (2015). *Estudo de Caso-: Planejamento e Métodos*. Bookman editora.
- Zipf, G. K. (1945). The meaning-frequency relationship of words. *The Journal of general psychology*, 33(2), 251-256.

Links: Recuperados em: <http://www.iramuteq.org/news>, citado em 14 de agosto de 2017.

Tutorial IRaMuTeq versão em português:

http://www.iramuteq.org/documentation/fichiers/Tutorial%20IRaMuTeQ%20em%20portugues_17.03.2016.pdf/view

Instalação do software R – No Brasil: <http://nbcgib.uesc.br/mirrors/cran/>

Instalação do software IRaMuTeq : <http://sourceforge.net/projects/iramuteq/>



VI SINGEP

Simposio Internacional de Gestao de Projetos, Inovacao e Sustentabilidade
International Symposium on Project Management, Innovation and Sustainability

ISSN: 2317-8302

V ELBE

Encontro Luso-Brasileiro de Estrategia
Iberoamerican Meeting on Strategic Management