



O SOFTWARE IRAMUTEQ COMO FERRAMENTA METODOLÓGICA: CONTRIBUIÇÕES PARA AS PESQUISAS EM EDUCAÇÃO MATEMÁTICA

Yohana Taise Hoffmann¹

David Antonio da Costa²

Resumo: O objetivo dessa comunicação é apresentar o software IRaMuTeQ (Interface de R pour les Analyses Multidimensionnelles de Textes et de Questionnaires) como uma nova perspectiva metodológica, seus recursos para a realização das análises textuais, visando contribuir na elaboração dos trabalhos desenvolvidos pelos diversos projetos de pesquisa (iniciação científica, mestrado, doutorado e pós-doutorado). A escolha do IRaMuTeQ deve-se ao fato de se tratar de um software gratuito. Este software permite a realização de análises lexicográficas, que identificam a quantidade de palavras, a frequência média e o número de hapax (palavras com frequência um); as análises de especificidades e a análise fatorial de correspondência (AFC); a classificação hierárquica descendente (CHD), que classifica os segmentos de texto em função dos seus respectivos vocabulários; a análise de similitude, baseada na teoria dos grafos, permitindo identificar as co-ocorrências entre as palavras; a construção de nuvens de palavras, por meio do agrupamento e da organização das palavras em função da sua frequência.

Palavras-chave: Análise Textual. Ferramenta Metodológica. IRaMuTeQ. Humanidades Digitais.

1. APRESENTAÇÃO

O grupo de pesquisa GHEMAT-SC (Grupo de Pesquisa de História da Educação Matemática de Santa Catarina) tem se aproximado da área da Humanidades Digitais, ou seja, uma intersecção entre as perspectivas analíticas tradicionalmente conhecidas do campo das humanidades, com as novas possibilidades metodológicas e reflexivas advindas das possibilidades mediadas pelo digital, configurando-se como um campo interdisciplinar nas reflexões e práticas científicas. Atualmente existem diversos softwares que permitem a realização de análises textuais, contribuindo para a visualização dos dados, e de análises de conteúdo, transformando dados qualitativos em quantitativos. A opção por uma ou outra ferramenta de tratamento de dados depende dos seus princípios de funcionamento e, sobretudo, do delineamento teórico-metodológico utilizado na pesquisa.

¹ Mestre em Educação Científica e Tecnológica, Universidade Federal de Santa Catarina (UFSC). Doutoranda no Programa de Pós-Graduação em Educação Científica e Tecnológica (PPGECT – UFSC). E-mail: yohana.thc@gmail.com.

² Doutor em Educação Matemática, Pontifícia Universidade Católica (PUC/SP). Docente da Universidade Federal de Santa Catarina (UFSC), Departamento de Metodologia de Ensino. E-mail: david.costa@ufsc.br.



O objetivo dessa comunicação é apresentar o software IRaMuTeQ (Interface de R pour les Analyses Multidimensionnelles de Textes et de Questionnaires) como uma nova perspectiva metodológica, seus recursos para a realização das análises textuais, visando contribuir na elaboração dos trabalhos desenvolvidos pelos diversos projetos de pesquisa (iniciação científica, mestrado, doutorado e pós-doutorado).

Dessa forma, o público-alvo para a utilização do software são pesquisadores de diferentes áreas e níveis. Para este evento, a comunicação encontra-se no Eixo C: Pesquisa em Educação Matemática, e como solicitado as discussões serão organizadas pelo subeixo 10 – Tendências em Educação Matemática, tais como: Modelagem Matemática, Resolução de Problemas, Tecnologias Digitais, Etnomatemática, Investigação Matemática, História da Matemática, Histórico-Cultural.

2. ASPECTOS GERAIS DO SOFTWARE

O IRaMuTeQ (<http://www.iramuteq.org/>) é um software livre desenvolvido por Pierre Ratinaud, junto ao Laboratoire d'Études et de Recherches Appliquées en Sciences Sociales (LERASS) da Universidade de Toulouse, e lançado ano de 2008. Tem como base o pacote estatístico software R (www.r-project.org) e a linguagem de programação Python (www.python.org). Permite fazer análises estatísticas sobre *corpus* textuais e sobre tabelas indivíduos/palavras. Ou seja, ele dá um viés quantitativo para dados qualitativos.

Os recursos do IRaMuTeQ estão relacionados às análises de *corpus* textual, que corresponde ao material escrito e transcrito, sendo possível analisar textos, entrevistas, documentos, redações etc. De acordo com Camargo e Justo (2013, p. 514), a análise textual ou lexical contribui para a superação da “dicotomia clássica entre quantitativo e qualitativo na análise de dados, na medida em que possibilita que se quantifique e empregue cálculos estatísticos sobre variáveis essencialmente qualitativas - os textos”. Segundo Camargo e Justo (2018), o software permite realizar:

- Análises lexicográficas, que identifica a quantidade de palavras, a frequência média e o número de *hapax* (palavras com frequência um);
- Análises de especificidades e análise fatorial de correspondência (AFC), são especialmente indicadas para descrever um grande volume de dados, associando textos com modalidades de uma única variável de caracterização, ou seja, possibilita a comparação (contraste) da produção textual destas modalidades;



- A classificação hierárquica descendente (CHD), que classifica os segmentos de texto em função dos seus respectivos vocabulários, formando um esquema hierárquico de classes, e o conjunto deles é repartido em função da presença ou ausência das formas reduzidas;
- Análise de similitude, baseada na teoria dos grafos, permitindo identificar as co-ocorrências entre as palavras e seu resultado traz indicações da conexidade entre as palavras, auxiliando na identificação da estrutura do conteúdo de um *corpus* textual;
- Nuvem de palavras, um recurso contribui para a visualização de um *corpus* textual, agrupa e organiza as palavras em função da sua frequência. As palavras com maior frequência são as maiores e são colocadas ao centro do gráfico, as menores representam frequências inferiores.

Cabe mencionar que o *corpus* textual que será analisado passa por uma preparação, de acordo com Moraes e Ambrósio (2007, p. 9), que a “preparação dos dados é a primeira etapa do processo de descoberta de conhecimento em textos. Envolve a seleção dos dados que constituem a base de textos de interesse e o trabalho inicial para tentar selecionar o núcleo que melhor expressa o conteúdo destes textos”.

Para a inclusão dos dados no software é necessário a separação dos textos com linhas de comando, da seguinte maneira: digitar quatro asteriscos (sem espaço em branco antes deles), um espaço branco subsequente, um asterisco e o nome da variável (sem espaço em branco entre eles). No entanto, recomenda-se que não se justifique o texto, não use negrito, nem itálico ou outro recurso semelhante. As palavras compostas hifenizadas quando digitadas com hífen são entendidas como duas palavras (o hífen vira espaço em branco), ou as palavras que dão uma ideia de ligação ou são nomes próprios devem ser substituídos com um traço underline, por exemplo, história da educação matemática é alterado para história_educação_matemática. Os números devem ser mantidos na forma de algarismos, não usar caracteres especiais como: aspas, apóstrofo, hífen, cifrão, porcentagem, reticências, e nem asterisco, apenas como linha de comando (HOFFMANN, BISSET, MARTI-LAHERA, 2020).

Como podemos verificar no seguinte exemplo:

****** *Resumo2017_01**

A intenção desta pesquisa foi compreender e elaborar versões históricas, a partir de diversas perspectivas documentais e testemunhais, de como se deu a formação dos professores que ensinavam matemática no estado do RN, nas séries que correspondem aos atuais ensinos fundamental e médio, incluindo possíveis aspectos relacionados ao desenvolvimento desse processo de formação, antes e durante o período de criação dos 4 mais antigos cursos superiores que habilitavam professores em Matemática neste estado. Utilizamos como metodologia de pesquisa a história_oral, produzindo e



analisando 20 narrativas e documentos acessados em arquivos públicos e privados. Seguindo uma perspectiva transdisciplinar, elaboramos compreensões sobre o processo, formação de professores de matemática, e constituímos narrativas segundo 3 disposições distintas mas interconectadas, a análise narrativa de narrativas, a análise de singularidades e a análise de convergências. O texto final desse movimento de pesquisa tem um estilo alternativo que pretende dialogar com as teorizações feitas (HOFFMANN, BISSET, MARTI-LAHERA, 2020, p. 114).

O resultado da análise de estatísticas textuais traz cinco informações que resumem o *corpus* textual: a) Número de textos - é o número de textos (registros) contidos no *corpus*; b) Número de ocorrências - é o número total de palavras contidas no *corpus*; c) Número de formas - são as palavras consideradas ativas (adjetivos, nomes, verbos) e suplementares (artigos e pronomes); d) Número de hápax - são palavras que aparecem apenas uma vez em todo o *corpus*; e) Média de ocorrências por texto - é o número de ocorrências dividido pelo número de textos (SALVIATI, 2017, p. 33-34).

Os autores Hoffmann, Bisset, Marti-Lahera (2020) apresentam uma análise comparativa entre teses e dissertações do campo da História da Educação Matemática (Hem) no Brasil, realizadas nos anos de 2012 e 2017, correspondendo a 18 e 33 trabalhos respectivamente. Em relação a análise estatística textuais os autores apresentam dois quadros (Quadro 1 e Quadro 2) que ilustram como o software apresenta os dados:

Quadro 1: Resumo *corpus* textual 2012

Resumo	
Número de textos	18
Número de ocorrências	5373
Número de formas	1246
Número de hápax	688 12.80% de ocorrências 55.22% de formas
Média de ocorrências por texto	298.50

Fonte: Hoffmann, Bisset, Marti-Lahera (2020, p. 116).

Quadro 2: Resumo *corpus* textual 2017

Resumo	
Número de textos	33
Número de ocorrências	8153
Número de formas	1487
Número de hápax	738 9.05% de ocorrências 49.63% de formas
Média de ocorrências por texto	247.06

Fonte: Hoffmann, Bisset, Marti-Lahera (2020, p. 116).



Cabe destacar que os autores Hoffmann, Bisset, Marti-Lahera (2020) utilizaram o IRaMuTeQ como um recurso metodológico, que contribui para a pré análise das teses e dissertações defendidas em 2012 e 2017, utilizando alguns recursos do software como: a nuvem de palavras e a AFC. Dessa forma, os autores conseguiram identificar novas temáticas e referências metodológicas, principalmente as pesquisas vinculadas aos grupos: Grupo de Pesquisa de História da Educação Matemática no Brasil (GHEMAT) e Grupo de História Oral e Educação Matemática (GHOEM). Assim, “se evidencia a influência de grupos de pesquisa nas agendas de pesquisa das universidades” (HOFFMANN, BISSET, MARTI-LAHERA, 2020, p. 124).

Outro exemplo da utilização do software como um recurso metodológico pode ser observado no trabalho de Hoffmann, Costa e Nakamura (2020). Os autores analisaram 17 trabalhos apresentados no GT 15: História da Educação Matemática do VII Seminário Internacional de Pesquisa em Educação Matemática (SIPEM). Em relação a análise de estatísticas textuais os autores mencionam:

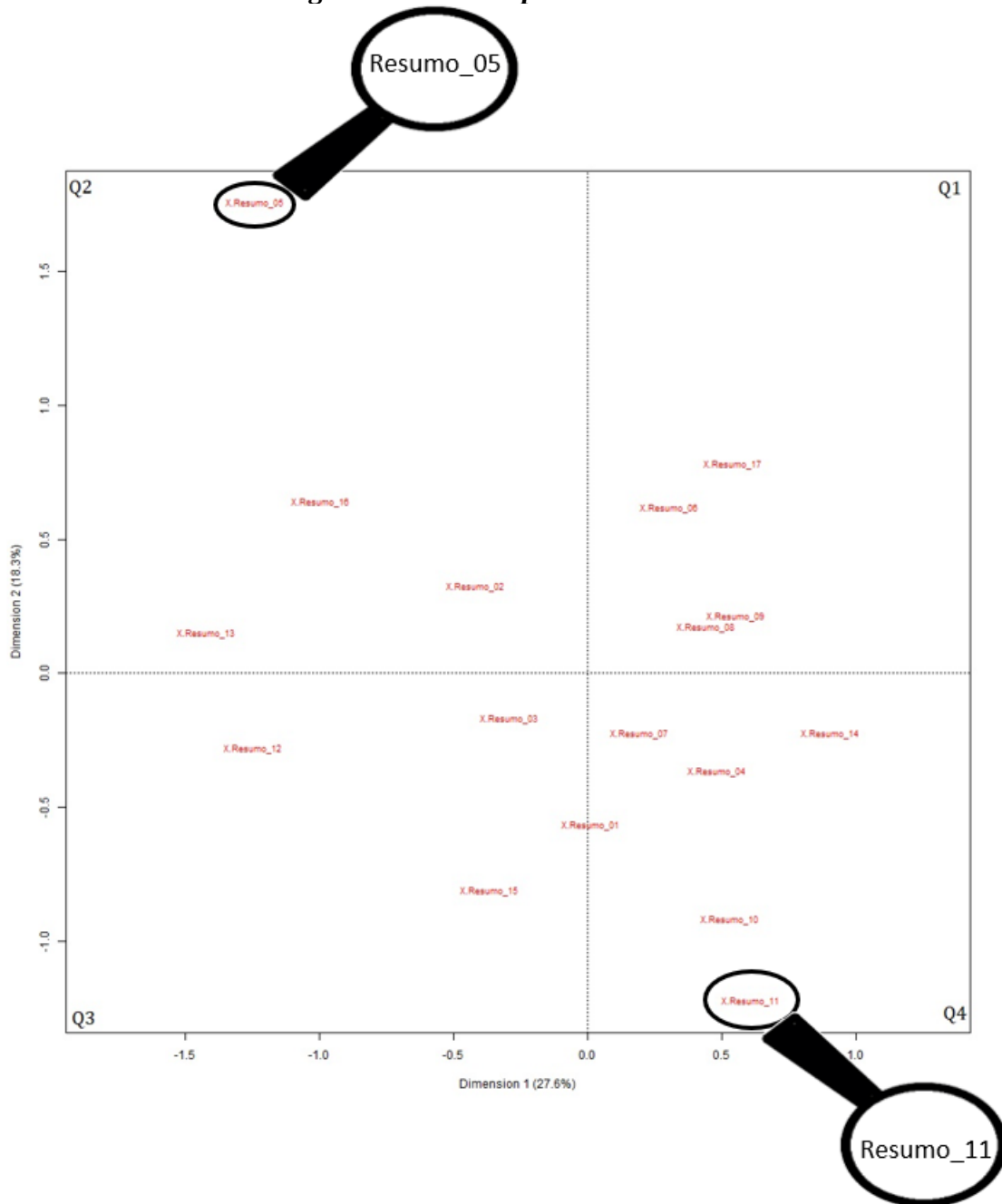
- a) Número de textos - são os textos (registros) contidos no corpus, em nosso caso, por exemplo, o corpus corresponde aos trabalhos do GT15 compostos pelos 17 resumos;
- b) Número de ocorrências - é o número total de palavras contidas no corpus, correspondente a 2.164 palavras;
- c) Número de formas - são as palavras consideradas ativas (adjetivos, nomes, verbos) e suplementares (artigos e pronomes). Para a análise foram eliminados os advérbios, as conjunções, onomatopeias e as preposições, que totalizam 665 palavras (HOFFMANN, COSTA, NAKAMURA, 2020, p. 925).

Para visualizar o *corpus* textual, neste caso os resumos dos 17 trabalhos, os autores utilizaram a nuvem de palavras (Figura 1), a palavra “pesquisa” é central (citada 28 vezes), algumas palavras indicam temáticas e linhas de pesquisas como “ensinar (frequência 14); matemática (frequência 13); ensino (frequência 12); história da educação matemática, formação de professores e estudo (frequência 10); professor e educação matemática (frequência 9)” (HOFFMANN, COSTA, NAKAMURA, 2020, p. 926).

[illegible]

Outro trabalho que se distancia dos demais é o T11 do Grupo de Estudos Contemporâneos em Educação Matemática (GECEM), localizado no Q4. De acordo com Hoffmann, Costa e Nakamura (2020, p. 929), é possível “identificar que o T11 é o único dos trabalhos que trata da arqueogenealogia do saber segundo Michel Foucault, razão de seu distanciamento dos demais”.

Figura 2 – AFC *corpus* textual GT15.



Fonte: Hoffmann, Costa, Nakamura (2020, p. 928).

Os autores Hoffmann, Costa e Nakamura (2020) a partir da AFC conseguiram identificar as temáticas que mais se aproximam e se distanciam entre si. Outra questão que podemos observar é a participação dos grupos de pesquisas, os autores identificam que:

dentro deste universo, são o GHEMAT, com seis trabalhos apresentados e o GH OEM, com cinco trabalhos. A identificação dos GP's pode ser um fator relevante para caracterizar as diferentes temáticas dos trabalhos, assim como, a participação de um pesquisador em mais de um grupo pode apresentar uma circulação de ideias, práticas e teorias (HOFFMANN, COSTA, NAKAMURA, 2020, p. 934).



Assim, apresentamos brevemente a utilização dos recursos e análises do software. A respeito da instalação do IRaMuTeQ há diversos tutoriais disponíveis, assim como, artigos que apresentam as especificidades do software e como as análises podem contribuir nas diferentes pesquisas que se encontram em andamento nas mais diversas áreas e níveis.

3. CONSIDERAÇÕES FINAIS

Para concluir, é importante destacar que o IRaMuTeQ é utilizado, nestes exemplos, como um recurso metodológico, que contribui para uma pré análise dos documentos a serem investigados, como as teses, dissertações e trabalhos completos apresentados em eventos.

De acordo com Camargo e Justo (2018, p. 517) “os relatórios gerados pelo software não são, em si, a análise dos dados” cabe ao pesquisador fazê-las. O IRaMuTeQ manipulou os dados qualitativos, o *corpus* textual, e transformou em dados quantitativos, contribuindo com recursos que auxiliam o pesquisador na descoberta de novos conhecimentos.

REFERÊNCIAS

CAMARGO, B. V.; JUSTO, A. M. **IRAMUTEQ**: Tutorial para uso do software de análise textual IRAMUTEQ. Florianópolis: Laboratório de Psicologia Social da Comunicação e Cognição- LACCOS, Universidade Federal de Santa Catarina, Brasil, 2018.

CAMARGO, B. V.; JUSTO, A. M. IRAMUTEQ: um software gratuito para análise de dados textuais. **Temas psicol.**, Ribeirão Preto, v. 21, n. 2, p. 513-518, dez. 2013.

HOFFMANN, Y. T.; BISSET, E.; MARTI-LAHERA, Y. Análise textual com IRaMuTeQ de pesquisas recentes em História da educação matemática no Brasil: um exemplo de Humanidades Digitais. **Revista: Investigación Bibliotecológica: archivonomía, bibliotecología e información.** v. 34, p. 103-133, 2020. DOI: <http://dx.doi.org/10.22201/iibi.24488321xe.2020.84.58097>

HOFFMANN, Y. T.; COSTA, D. A.; NAKAMURA, L. R. Una mirada a las producciones en Historia de la educación matemática en el VII Seminario Internacional de Investigación en Educación Matemática. **PARADIGMA (MARACAY)**, p. 912-937, 2020. DOI: <https://doi.org/10.37618/PARADIGMA.1011-2251.2020.p912-937.id775>

MORAIS, E. A. M.; AMBRÓSIO, A. P. L. **Mineração de textos**. Goiás: Instituto de Informática da Universidade Federal de Goiás, 2007. 30p. (Relatório Técnico, RT- INF 005-07).

SALVIATI, M. E. Manual do Aplicativo Iramuteq (versão 0.7 Alpha 2 e R Versão 3.2.3). Compilação, organização e notas de Maria Elisabeth Salviati. 2017. Disponível



em: <<http://iramuteq.org/documentation/fichiers/manual-do-aplicativo-iramuteq-par-maria-elisabeth-salviati>> . Acesso em: 10 nov. 2018.

FONTE(S) FINANCIADORA(S)

Os autores agradecem o apoio desta pesquisa dado pelas seguintes agências de fomento:

1. Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES)
2. Fundação de Amparo à Pesquisa e Inovação do Estado de Santa Catarina (FAPESC)