

TEXTOS COMO DADOS: INTRODUÇÃO À ANÁLISE AUTOMATIZADA DE CONTEÚDO EM PESQUISAS QUALITATIVAS

TEXTS AS DATA: INTRODUCTION TO AUTOMATED CONTENT ANALYSIS IN QUALITATIVE RESEARCH

Murilo Mesquita¹

Programa de Pós-Graduação em Ciência Política
Universidade Federal de Pernambuco
Recife – Pernambuco – Brasil

Resumo: Como utilizar ferramentas computacionais, como um *software*, na pesquisa qualitativa? A busca por responder essa pergunta se dá em meio a noção, por vezes equivocada, de que pesquisas qualitativas não reivindicam um maior refinamento metodológico. Sendo assim, a partir dessa pergunta, a pesquisa busca auxiliar no avanço metodológico de desenhos de pesquisas qualitativas para o campo de estudo de Relações Internacionais. Destarte, tem como objetivo geral apresentar uma ferramenta de pesquisa pautada em técnicas de Análise Automatizada de Conteúdo (AAC), em que pese desenvolver os seguintes objetivos específicos: (i) explicar o que é a Análise de Conteúdo; (ii) identificar as características pertinentes à AAC; e (iii) introduzir o software IRAMUTEQ junto à pesquisa qualitativa do campo de Relações Internacionais. O trabalho possui uma característica didático-pedagógica porque introduz um *software* para a produção de pesquisa qualitativa. Ao fim estima contribuir para introdução de uma ferramenta que possa auxiliar o refinamento metodológico dentro do mundo da pesquisa qualitativa em Relações Internacionais.

Palavras-chave: Pesquisa Qualitativa. Análise de Conteúdo. IRAMUTEQ. Metodologia de Pesquisa.

Abstract: How to use computational tools, such as software, in qualitative research? The search for answering this question takes place amidst the notion, sometimes mistaken, that qualitative research does not demand greater methodological refinement. Therefore, based on this question, the research seeks to assist in the methodological advance of qualitative research designs for the field of study of International Relations. Hence, its general aim is to present a research tool based on Automated Content Analysis (AAC) techniques, in spite of developing the following specific objectives: (i) explain what Content Analysis is; (ii) identify the characteristics relevant to the AAC; and (iii) introduce the IRAMUTEQ software together with qualitative research in the field of International Relations. The work has a didactic-pedagogical characteristic because it introduces software for the production of qualitative research. In the end, it hopes to contribute to the introduction of a tool that can help the methodological refinement within the world of qualitative research in International Relations.

Key-words: Qualitative research. Content analysis. IRAMUTEQ. Research Methodology.

Recebido: 20/01/2022.

Aprovado: 06/02/2022.

DOI:

¹ murilo_mesquita@hotmail.com

Orcid: <https://orcid.org/0000-0002-7313-2974>

Introdução

A pesquisa qualitativa, de um modo geral, possui um histórico que é cercado por alguns mitos. Um desses mitos é de que a metodologia qualitativa não possui o refinamento adequado para produzir inferências causais ou descritivas. Com essa visão, julga-se que as complexidades sociais não permitem observar padrões, estimar comportamentos, mensurar dados qualitativo ou, tampouco, produzir inferências válidas. Com essas perspectivas, as pesquisas científicas operadas a partir de uma cultura metodológica qualitativa são, corriqueiramente, observadas, não apenas como avessas à métodos e técnicas quantitativas, mas também aos próprios métodos e técnicas qualitativas que permitam o refinamento metodológico.

Diante desse contexto, essa investigação tem como problema de pesquisa a seguinte questão: Como utilizar ferramentas computacionais, como um *software*, na pesquisa qualitativa? O desenvolvimento de uma resposta para essa pergunta assume a justificativa científica de auxiliar no avanço metodológico de desenhos de pesquisas qualitativos para o campo de estudo de Relações Internacionais. Para tanto, tem como objetivo geral apresentar uma ferramenta de pesquisa pautada em técnicas de Análise Automatizada de Conteúdo (AAC), em que pese desenvolver os seguintes objetivos específicos: (i) explicar o que é a Análise de Conteúdo; (ii) identificar as características pertinentes à AAC; e (iii) introduzir o software IRAMUTEQ junto à pesquisa qualitativa do campo de Relações Internacionais.

Dado esse desenho de pesquisa, a investigação possui uma característica didático-pedagógica, porque introduz um *software* para a produção de pesquisa qualitativa, em que pese apresentar sua potencialidade ao tratar com grandes quantidades de dados. Dessa forma, a pesquisa é segmentada da seguinte maneira: Na primeira seção, é feita uma apresentação histórica da técnica de análise de conteúdo, quando são identificadas suas particularidades e fragilidades. A segunda seção versa sobre a AAC e suas potencialidades para realizar pesquisas com grande quantidade de dados textuais. A terceira seção apresenta o software IRAMUTEQ, suas vantagens e desvantagens e as técnicas que podem ser operacionalizadas para análise quantitativa de dados qualitativos. Por fim, com as considerações finais, esse trabalho estima ter feito uma contribuição pedagógica para introduzir uma ferramenta que possa auxiliar o refinamento metodológico dentro do mundo da pesquisa qualitativa em Relações Internacionais.

1. Textos como Dados: a história da Análise de Conteúdo

O método de Análise de Conteúdo (AC), como uma estratégia de investigação na pesquisa qualitativa possibilita, aumentar a capacidade de inferência a partir da utilização de “textos como dados”. Nesse sentido, importa um breve apanhado histórico acerca do que é a AC.

O texto é, de longe, o artefato mais difuso do comportamento político, remontando a tempos mais remotos da convivência humana, notadamente, com marco para a invenção da escrita. A possibilidade de analisar textos e providenciar *insights* sobre o processo político também consta de longa data (Krippendorff, 2004; Cervi, 2018). Dado esse imenso aspecto longitudinal, identificam-se quatro fases de desenvolvimento da moderna forma de análise de conteúdo.

A primeira fase, por volta de 1940-1950, chamada de Científica, ocorre nos EUA, quando pesquisadores buscam um maior rigor científico à interpretação de texto. É nesse momento que as técnicas de Lasswell ganham ressonância e transbordam dos textos jornalísticos para as peças publicitárias, quando passam a ser utilizadas para identificar a existência de propaganda subversiva em jornais estadunidenses (Lasswell & Kaplan, 1979; Krippendorff, 2004; Drisko & Maschi, 2016; Cervi, 2018). No Projeto Comunicação em Tempo de Guerra (*Wartimes Communication Project*), a AC se torna a técnica padrão para análise de comunicações inimigas (Monroe & Schrodt, 2008) e para prever ou inferir fenômenos que não podem ser observados diretamente (Krippendorff, 2004).

A segunda fase ocorre na década de 1960. Aqui, reconhece-se as limitações do modelo quando se passa a afirmar que não é possível extrair do texto mais do que ele apresenta explicitamente. Qualquer informação a mais seria fruto da subjetividade interpretativa do analista (Krippendorff, 2004; Drisko & Maschi, 2016; Cervi, 2018). Sendo assim, passa-se a conceber a AC como “uma técnica de investigação que tem por finalidade a descrição objetiva, sistemática e quantitativa do conteúdo manifesto da comunicação” (Bardin, 1977, p. 19).

A terceira fase ocorre por volta do final da década de 1970, até os anos 2000. Nesse momento, as limitações teórico-metodológicas observadas no período anterior são reconhecidas. Porém, o desenvolvimento tecnológico passa a permitir a utilização de programas de computadores que auxiliam a AC. Com o processamento de dados, a partir de *softwares* específicos, aumenta-se o rigor analítico e a amplitude de análise, o que permite que a AC ultrapasse a descrição dos dados para buscar a inferências analíticas (Lasswell & Kaplan, 1979; Cervi, 2018).

Do final dos anos 2000 até o presente, deduz-se que é possível observar a existência de uma quarta fase de desenvolvimento da AC. No contexto de uma realidade com *big data* e com uma

quantidade imensurável de dados que circula na *internet*, reivindica-se uma renovação técnica que permita a operação com velocidade e capacidade adequada para o volume de dados que existem. Sendo assim, é a partir do desenvolvimento tecnológico, com *software* mais robustos e sofisticados, que se marca a quarta fase do processo de automação da AC.

2. Por dentro da Análise de Conteúdo (AC)

Após essa breve apresentação histórica, um conceito para a técnica de AC tende a ser feita a partir de concepções berelsianas (Stemler, 2001). Isso quer dizer que a AC se configura como “um conjunto de técnicas de comunicações que tem como objetivo a descrição sistemática de conteúdo das mensagens a partir de indicadores quantitativos que permitem fazer inferências sobre as condições de produção” (Cervi, 2018, p. 7).

Esse conceito está atrelado ao que Drisko e Maschi (2016) chamam de Análise de Conteúdo Básica, caracterizada pela pouca interferência humana, pelo uso de *software* para sistematização dos dados e, geralmente, atrelada às técnicas quantitativas da estatística descritiva. Esse modelo de AC costuma ser elaborado por métodos que permitem a dedução frequencial ou a análise por categorias temáticas (Caregnato & Mutti, 2006).

Importante destacar que há variações epistemológicas nas técnicas de AC. A AC padrão trabalha com uma lógica frequentista, que reivindica métodos quantitativos de contagem, para auferir validade e replicação em função da frequência e da relação de coocorrência das palavras. A AC interpretativista e qualitativa, como é o caso, por exemplo, da Análise do Discurso, trabalha com uma lógica reflexivista que reivindica técnicas qualitativas de análise. Por isso, esta última versa sobre procedimentos subjetivos para compreensão do discurso e dos seus sistemas de significados resultantes do contexto das relações de poder.

Apesar da diferença acima citada, ambos os modelos de análise comungam da necessidade de rigor metodológico para a sistematização e a análise dos dados, bem como reiteram a importância e indispensabilidade da subjetividade do analista para compreender o dado em um contexto específico. Nesses termos, apesar dessas variações, essas técnicas não são mutuamente excludentes, ao fim e ao cabo são modelos complementares (Drisko & Maschi, 2016). Para uma abordagem quantitativista de AC, cita-se Neuendorf (2002) e Krippendorff (2004) e, para uma abordagem reflexivista, há Campbell (1992).

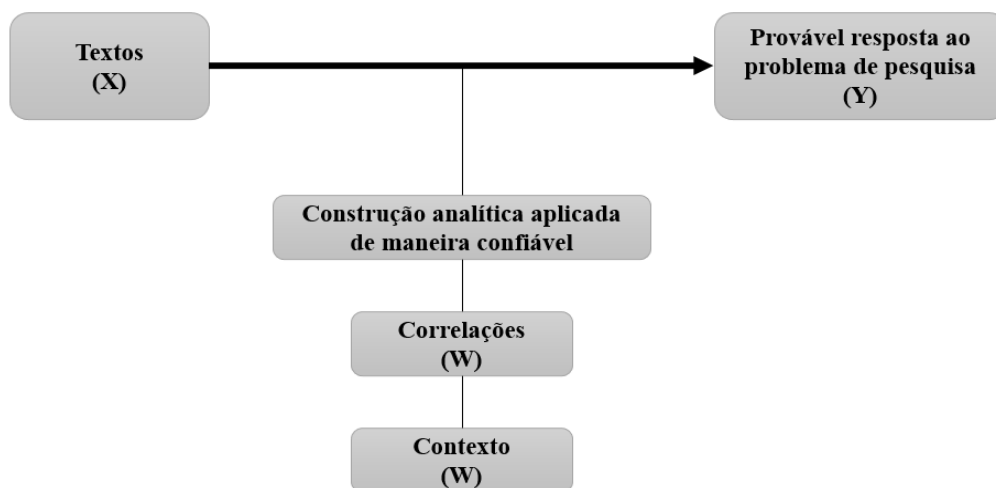
A partir da percepção supra demonstrada, a AC é “*a research technique for making replicable and valid inferences from texts (or other meaningful matter) to the contexts of their use*”.

(Krippendorff, 2004, p. 18). Por isso, uma análise de conteúdo rigorosa envolve procedimentos técnicos que providenciam novos *insights* e aumentam o conhecimento sobre o fenômeno estudado dentro de um contexto, especialmente quando trabalha com uma noção de sistematização do que é descrito, para que o modelo de análise seja passível de replicação (Drisko & Maschi, 2016).

Portanto, é a compreensão clínica da AC, porque objetiva e sistemática, que ao requerer a replicação e a validação dos dados, torna a inferência uma questão fundamental ao método (Bardin, 1977; Janis, 1979; Krippendorff, 2004). Essa inferência pode ser alcançada pelos indicadores de frequências e/ou de coocorrências (contingência), quando, a partir deles, a análise pode regressar às causas ou chegar aos resultados das características das comunicações.

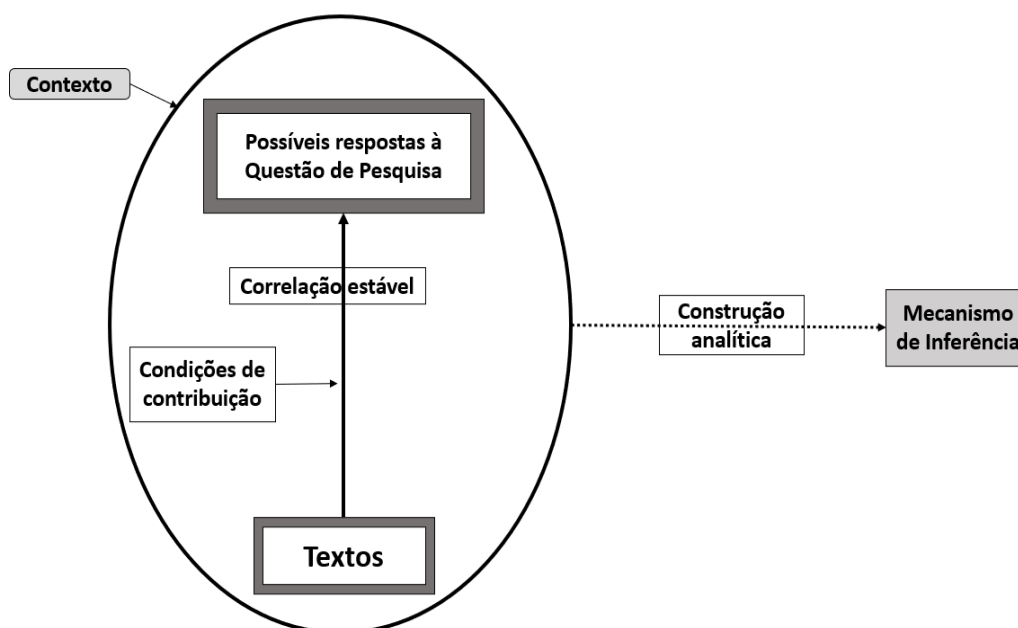
Para Krippendorff (2004), essa lógica inferencial trabalha a partir da concepção de que um Dado “X” gera um resultado “Y” se ele é sustentado pela(s) condição (ões) “W”. Na AC, esses dados são os textos e o resultado é a resposta à pergunta-problema da pesquisa, enquanto W são as correlações oriundas do modelo procedimental que é contextualizado e teoricamente orientado. As duas figuras (1 e 2) abaixo buscam ilustrar didática e respectivamente a relação da inferência e como ocorre sua operacionalização:

Figura 1: Modelo inferencial para Análise de Conteúdo



Fonte: Krippendorff (2004), p. 38, tradução do autor.

Figura 2: Operacionalização do Modelo inferencial para Análise de Conteúdo



Fonte: Krippendorff (2004), p. 90, tradução do autor.

Com a interpretação dessas figuras, é possível observar que a AC possui algumas distinções para a produção de inferência, quando comparada a outras técnicas. A primeira distinção é a inferência do tipo abdução. Essa inferência se baseia “*on analytical constructs or models of the chosen context as warrants*” (Krippendorff, 2004, p. 83). Com ela, o contexto se configura como uma variável interveniente na investigação, porque “*it directs the analysis of a text, and the results of the analysis contribute to a (re)conceptualization of the context*” (Krippendorff, 2004, p. 86).

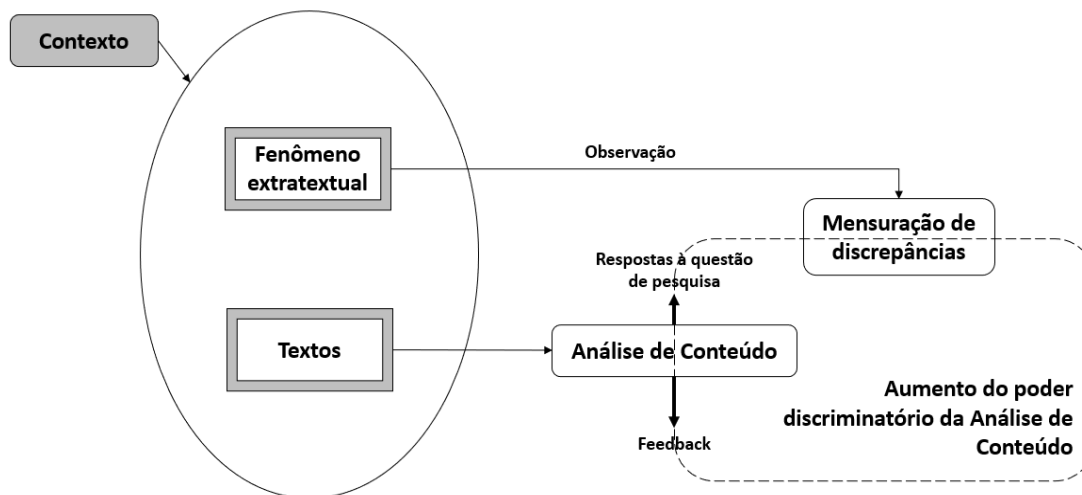
A segunda distinção é a importância que a correlação possui entre os dados e as condições atreladas ao contexto. Essa correlação é atribuída deterministicamente e com um direcionamento teórico que permita sua validação. Essa característica faz com que seja ressaltada a noção de correlação estável, porque opera em um contexto específico, mas, a depender da sua força, tem a capacidade de generalização de médio alcance.

Segundo Krippendorff (2004, p. 91-92), “[*t*]he analyst compares inferences from a content analysis of text with relevant observations of the context and uses any discrepancies found to alter incrementally the relevant parts of the analysis”. Com esse procedimento, constrói-se um desenho que pode ser paulatinamente ajustado no intuito de ultrapassar a razoável dúvida sobre o modelo.

A terceira distinção ocorre porque esse modelo é interativo. Isso quer dizer que a medida em que os dados são obtidos e analisados, eles devem ser observados em função do contexto e da teoria. Dessa maneira o procedimento analítico da AC permite avaliar a relação dos dados em função do contexto e, assim, aumentar a discriminação dessa função para chegar na inferência e responder o problema

de pesquisa. Para Krippendorff (2004, p. 92), essa distinção pode ser didaticamente expressada conforme figura 3 abaixo.

Figura 3: Representação do aspecto interativo do modelo de AC



Fonte: Krippendorff (2004), p. 92, tradução do autor

Diante disso, o rigor metodológico da AC passa a ter como objetivos dirimir as incertezas referentes ao que se julga observar nas mensagens e, conseqüentemente, enriquecer a compreensão do fenômeno estudado. Nesses termos, o caráter científico da análise ocorre quando o analista, por meio de descrições precisas e da transparência procedimental para obtenção e análise dos dados, consegue estabelecer uma relação entre as características dos dados, o comunicador e a audiência (Drisko & Maschi, 2016).

Por isso, esse tipo de análise tem duas funções que se complementam: (i) heurística, de enriquecimento do procedimento exploratório; e (ii) de administração da prova, quanto testa hipóteses (Bardin, 1977; Janis, 1979; Krippendorff, 2004; Drisko & Maschi, 2016). Com essas funções, a AC é desenvolvida, via de regra, em três fases (Bardin, 1977, p. 95-120):

- (i) Pré-análise: Nessa fase, a intuição ainda é muito forte. A partir dela, sistematizam-se as ideias iniciais de maneira a estabelecer um caminho de desenvolvimento para as operações de tratamento que ocorre em três momentos não necessariamente lineares:
 - a. Escolha dos documentos que constituem o *corpus* de pesquisa. A sua constituição implica seguir três regras:
 - i. Não-seletividade: o procedimento de seleção deve ser claro, objetivo e justificável;

- ii. Homogeneidade: a seleção deve obedecer a critérios precisos; e
 - iii. Pertinência: os documentos devem ser adequados ao problema da pesquisa e atrelados à hipótese e objetivos;
- b. Formulação de hipóteses e de objetivos; e
 - c. Formulação dos indicadores que fundamentem a interpretação final (à guisa de fundamentação teórica)
- (ii) Exploração material: Nessa fase, os resultados são levados a “falarem”, ou seja, a possuir significados válidos. Nesse momento, operações estatísticas permitem a produção de dados com informações relevantes do conteúdo extraído dos documentos analisados. Por isso, é pertinente o processo de codificação. Esse é o momento em que o “como” (a técnica) pode precisar o “porquê” (a teoria). Assim,
- a codificação corresponde a uma transformação - efetivada segundo regras precisas – dos dados brutos do texto, transformação esta que, por recorte, agregação e enumeração, permite atingir uma representação do conteúdo, ou da expressão, susceptível de esclarecer o analista acerca das características do texto (Bardin, 1977, p. 103).
- (iii) Tratamento dos resultados: Essa fase é caracterizada pelo processo de categorização do conteúdo. A categorização “é uma operação de classificação de elementos constitutivos de um conjunto, por diferenciação e, seguidamente, por reagrupamento segundo o gênero (analogia), com os critérios previamente definidos” (Bardin, 1977, p. 117). Esses critérios podem ser: Semânticos (por temas), sintáticos (verbos ou adjetivos) ou lexicais e/ou expressivos.

Ainda, seguem algumas regras (Bardin, 1977, p. 120; Carlomagno & Rocha, 2016, p. 178):

1. Exclusão mútua: Um elemento não pode estar contido em duas categorias;
2. Homogeneidade: As categorias são parcimoniosas a ponto de não conceber duplo sentido;
3. Pertinência: As categorias e os elementos alocados são convergentes ao quadro teórico definido;
4. Objetividade e fidelidade: O processo de alocação dos elementos às categorias é feito de forma simples e elegante, ou seja, de forma objetiva e sem viés; e

5. Produtividade: O quadro categórico elaborado passa pelo escrutínio científico de forma a ter seus resultados validados.

Por tanto, a AC é uma técnica especificamente atrelada a um processo de identificação, descrição e predição de elementos textuais que, ao mesmo tempo em que assume uma lógica metodológica frequentista, (Camargo & Justo, 2013; Carlomagno & Rocha, 2016; Cervi, 2018), requisita a interferência e a subjetividade do pesquisador, com a diferença de que essa interferência, por mais que seja indispensável, deve ser a menor possível (Krippendorff, 2004; Drisko & Maschi, 2016).

Por isso, um dos destaques dessa técnica de análise é a confiabilidade e a validade dos resultados obtidos decorrentes da interpretação dos dados. Em que pese essa característica, importa entender que, na AC, o texto não possui um significado único. Ele é refém do contexto, do processo de tratamento e, especialmente, da interpretação e robustez teórica dos analistas (Krippendorff, 2004; Drisko; Maschi, 2016; Riffe et al., 2014 *Apud* Cervi, 2018).

Por se configurar assim, a AC pode ser compreendida como uma abordagem metodológica híbrida, que comunga tanto das epistemologias quanto das metodologias quantitativas e qualitativas (Drisko & Maschi, 2016). Devido a isso, opta-se, aqui, por otimizar a referência quantitativa da AC, a partir da automatização de parte dessa análise. Este intento permite minimizar a interferência de possíveis subjetividades humanas no processo de tratamento dos dados.

3. A Análise Automatizada de Conteúdo

A análise de dados se configura hoje como um dos imperativos à pesquisa qualitativa. Em função da subjetividade humana e do volume de dados do atual momento histórico, cada vez mais requisita-se um maior rigor metodológico para o tratamento desses dados. Em face disso, a Análise Automatizada de Conteúdo (AAC) se configura como um novo campo de estudo e uma nova ferramenta para a pesquisa científica que trabalha desde uma cultura qualitativa.

Para tanto, *softwares* são desenvolvidos e tornam-se cada vez mais presentes nos estudos das áreas das Ciências Humanas, Sociais e da Saúde, notadamente naquelas áreas do conhecimento que têm seus estudos baseados em um volumoso *corpus* para análise (Salvador et al., 2018).

Nesse sentido, a análise automatizada permite a realização de uma tarefa sistemática, longa e exaustiva em larga escala de dados textuais, sem que seja necessário o emprego da subjetividade e da interferência humana. Nessa análise, métodos automatizados permitem que a complexidade da linguagem seja trabalhada como suporte à leitura detalhada e cuidadosa dos textos (Izumi & Moreira, 2018).

Por isso, as técnicas de análise automatizadas de conteúdos são identificadas como amplificadoras e ampliadoras e ocorrem em função da revolução causada pela Internet, que torna possível aos pesquisadores das ciências sociais mobilizar uma quantidade de dados nunca antes disponível (Monroe & Schrodtt, 2008; Grimmer & Stewart, 2013).

Nesse processo de ampliação e amplificação, duas tarefas são pertinentes: A classificação e a categorização. A classificação organiza textos dentro de categorias. Para essa tarefa dois modelos são úteis. O primeiro modelo de análise recorre a um dicionário, que usa a frequência de palavras-chave para determinar sua categoria (Grimmer & Stewart, 2013, p. 268). Dentro desse método há um caminho supervisionado que ocorre quando o pesquisador inicia a codificação dentro de um conjunto predeterminado de categorias, para, então, introduzir modelos estatísticos que classificam o restante dos documentos. Esse modelo é regido por quatro princípios (Grimmer & Stewart, 2013, p. 269-271; Izumi & Moreira, 2018, p. 140):

- (i) Todo modelo quantitativo de linguagem está errado, mas alguns são úteis. A partir desse princípio, tem-se que o processo de geração de dados a partir de qualquer texto é um mistério. A dependência estrutural de qualquer sentença ao todo do texto torna qualquer documento complexo, o que implica dizer que todo método necessariamente falha ao precisar uma conta a partir do processo de geração de dados desde textos.
- (ii) Métodos quantitativos aumentam a capacidade humana de análise, mas não a substitui. Nesse princípio, tem-se que os métodos quantitativos auxiliam a análise do pesquisador. Saliente-se que eles não podem ser tomados por si só, mas como ferramentas que ampliam e amplificam as habilidades do pesquisador.
- (iii) Não há um melhor método para análise automatizada de texto. Esse princípio regula a noção de que diferentes conjuntos de dados e diferentes questões de pesquisa levam a diferentes interesses, diferentes modelos, diferentes métodos de validação.
- (iv) É imperativo validar a pesquisa de análise de conteúdo. Esse princípio se refere exatamente ao fato de que diferentes métodos para diferentes desenhos de pesquisa requisitam a validação da pesquisa.

Com esses princípios, resta ao pesquisador estruturar sua pesquisa a partir de um *corpus* de análise que possa ser operacionalizado por programas estatísticos. Ao ser configurado dessa forma, o *corpus*, como unidade de análise, passa a ser visto como uma “*bag of words*” (Grimmer & Stewart,

2013, p. 272), dentro da qual não importa como estão distribuídas, para a análise ser instrumentalizada.

Na operacionalização desse modelo, descarta-se o ordenamento das palavras e o vocabulário é simplificado com o *stemming*, que “removes the ends of words to reduce the total number of unique words in the data set, or reduce the dimensionality of text”. (Grimmer & Stewart, 2013, p. 272). Ou seja, esse processo reduz e deflexiona as palavras ao mapeá-las em função do seu conceito original – o radical das palavras. Ainda, são excluídas as *stop words*: pontuações, palavras muito incomuns e conectivos. De regra, removem-se palavras que aparecem menos de 1% e mais que 99% nas unidades de análise do *corpus* (Grimmer & Stewart, 2013, p. 273; Izumi & Moreira, 2018, p. 142).

Esse modelo de análise de conteúdo tem experimentado um crescimento considerável em anos recentes e a aplicação pivô é aquela desenvolvida por Benoit e Laver com seu modelo *WordScores* de 2003. Marco do desenvolvimento desse modelo de análise ocorre, em 2006, com um workshop na *American Political Science Association*, intitulado *Automated Content Analysis and Computer Annotation* e organizado por Stephen Purpura (Monroe & Schrodt, 2008).

Importa registrar que a utilização de textos como dados perpassa observar que o processo de geração, produção e seleção de dados resultantes da comunicação humana ainda se configura como demasiado complexo. “Esta complexidade faz com que modelos estatísticos desenvolvidos falhem na tarefa de prover um relato preciso do processo de geração de dados” (Izumi & Moreira, 2018, p. 139). Exatamente por isso, a análise automatizada de conteúdo não deve ser mobilizada como único instrumento de análise, mas como ferramenta analítica suplementar ao exercício de interpretação teoricamente orientada do analista que, ao fim e ao cabo, continua indispensável.

3.1. O IRAMUTEQ

Como salientado, o volume de dados com que o analista se depara nos dias atuais faz com que recursos tecnológicos se configurem como ferramentas importantes para o desenvolvimento de pesquisas que trabalham com um volume excessivo de dados. Além disso, recursos tecnológicos, como *software* de análise textual, permitem ultrapassar a clássica e já anacrônica dicotomia entre pesquisas qualitativas e quantitativas, porque possibilitam a mensuração e o emprego de estratégias estatísticas de análise em cima de variáveis essencialmente qualitativas, como um texto, sem dispensar a subjetividade do pesquisador, necessária para compreender o que é produzido pela AAC (Carmargo & Justo, 2013; Carmargo, 2013; Salvador et al., 2018; Cervi, 2018).

Dada essa importância, a técnica de AAC empregada nessa pesquisa busca verificar a relação entre os dados produzidos por diferentes atores em um contexto específico. Por meio dessa relação,

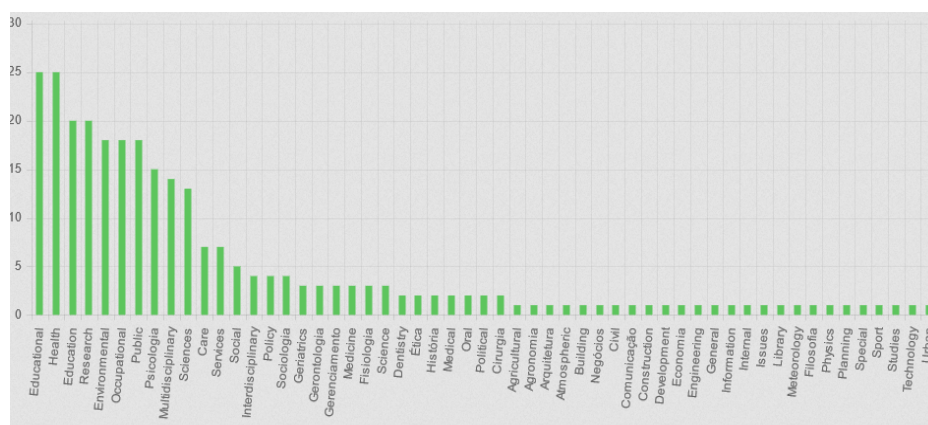
o modelo analítico concebe que a especificidade conjuntural permite a existência de um mundo lexical em que as palavras podem ser relacionadas umas às outras dentro de uma relação de coocorrência e de similaridade entre aqueles que produzem as palavras (Carmargo & Justo, 2013; Carmargo, 2013; Salvador et al., 2018; Cervi, 2018).

Destarte, para operar esse modelo de análise, a pesquisa faz uso de um *software* livre, o IRAMUTEQ, uma sigla usual para *Interface de R pour les Analyses Multidimensionnelles de Textes et de Questionnaires*, versão 0.7 alpha 2. No Brasil, ele é muito utilizado nas áreas de Psicologia, Enfermagem e Educação, quando busca identificar as percepções dos indivíduos em função de algum tratamento que estão a receber (Carmargo & Justo, 2013; Carmargo, 2013).

Sua utilização data de 2013, no Laboratório de Psicologia Social da Comunicação e Cognição da Universidade Federal de Santa Catarina (LACCOS/UFSC), em parceria com o Centro Internacional de Estudos em Representações Sociais e Subjetividade - Educação, da Fundação Carlos Chagas (CIERS-ed/FCC) e com o grupo de pesquisa Valores, Educação e Formação de Professores da Universidade Estadual Paulista Júlio de Mesquita Filho (UNESP) (Salvador et al., 2018).

Na plataforma *Scientific Electronic Library Online* (SciELO) do Brasil, ao utilizar o restritor IRAMUTEQ são encontrados 82 trabalhos. Porém, nenhum na área de RI, conforme é possível observar no gráfico 1 abaixo.

Gráfico 1: Distribuição de trabalhos com IRAMUTEQ no BR (2015-2021)



Fonte: Elaboração do autor, a partir da base SCIELO, em 2021.

Na área de Relações Internacionais, a pesquisa não encontra referências de seu uso no Banco de Teses do Instituto de Pesquisa em Relações Internacionais (IPRI), o que evidencia o caráter inovador desse recurso no campo de estudo de RI, no Brasil.

Esse programa possui uma base algorítmica ancorada em outro *software*, o *R*, que permite a utilização de pacotes úteis à análise de dados textuais. Nessa pesquisa, a versão do *R* que permite o

bom funcionamento do IRAMUTEQ é a 3.5.2, com a qual é possível realizar análises estatísticas sobre variáveis qualitativas (Camargo & Justo, 2013; Camargo, 2013).

As análises produzidas pelo IRAMUTEQ não usam um livro de códigos ou um dicionário, como ocorre no procedimento padrão de AC, que exige uma maior intervenção do pesquisador. Aqui, o *software* faz uso do próprio *corpus* empírico para identificar os termos que mais aparecerem, se aproximam e se distanciam entre si (Cervi, 2018, p. 9). Nesses termos, as análises podem ser operadas a partir de técnicas, desde as mais simples até as mais complexas, conforme o Quadro 1 abaixo.

Quadro 1: Técnicas analíticas do IRAMUTEQ

Técnica	Objetivo	Meio
N u v e m d e Palavras	Identificar a frequências das palavras;	Nuvem de Palavras Apresenta as palavras com maior frequência no texto;
C l a s s i f i c a ç ã o H i e r á r q u i c a D e s c e n d e n t e (CHD)	Obter categorias de palavras a partir de segmentos de textos (st) semelhantes entre e si e diferentes de outras categorias;	Dendograma. Por meio da lematização e teste qui-quadrado, é apresentada a relação entre as classes de palavras de um corpus, dado o seu (i) percentual de composição e sua (ii) a significância estatística (χ^2) no <i>corpus</i> ; Sua leitura é feita da esquerda para a direita;
Análise Fatorial de Correspondência (AFC).	Verificar a relação entre categorias de palavras de um mesmo ou de diferentes conjuntos de dados;	Plano cartesiano São evidenciados a relação de co-ocorrência entre os dados e entre os atores; Demonstra a aproximação ou distanciamento entre os dados;
A n á l i s e d e Similitude	Estabelecer a relação de co-ocorrências entre cluster de palavras estatisticamente significantes.	Arvore de similitude As ocorrências de palavras são apresentados em função da disposição das suas categorias dentro de um ordenamento específico de importância do centro para a periferia.

Fonte: Elaboração do autor, a partir de Camargo & Justo (2013), Camargo (2013), Cervi (2018) e Salvador et al. (2018).

Com essas técnicas, o próprio *software* gera *clusters* de palavras com homogeneidade interna e heterogeneidade externa, medidos a partir do χ^2 (chi-quadrado) e do p-valor <0,0001 das palavras dentro e entre os *clusters*. Essas duas medições estão na base do Método de Reinert, que servem como ferramentas lógicas para conceber a correlação entre os termos que compõem os segmentos de texto, porque conseguem observar o limite do *corpus* e a intensidade de presença dos termos nos diferentes *clusters* de um mesmo *corpus*.

Isso torna possível ultrapassar a simples verificação da frequência e quantidade dos termos lexicais, para observar sua associação por proximidade e intensidade. Por isso, AAC com o

IRAMUTEQ não faz uso de um dicionário, ou seja, um livro de códigos com variáveis pré-estabelecidas que servem de guia para análise dos dados (Cervi, 2018).

Assim, com a operacionalização das técnicas descritas acima, os discursos decodificados, a partir do Método de Reinert, são quebrados em função da significância estatística das palavras dentro e entre as classes, para, então, formar uma variedade de segmentos de texto pertinentes à classe e ao ator que produziu o ato de fala.

Assim, quando a CHD é operacionalizada, a técnica produz uma variedade de segmentos de textos que impõem a necessidade de seguir critérios para selecionar as peças de evidência que devem compor o dendrograma. Esses critérios são dois: (i) um objetivo, referente ao p-valor $<0,0001$, que elimina palavras que não estão diretamente associadas à respectiva classe; e (ii) um subjetivo, vinculado à interpretação da palavra com o segmento de texto correspondente e à sua classe.

Por sua vez, ao operacionalizar a AFC, o plano cartesiano produzido deixa evidente, porque visualmente mais didático, como as delegações se aproximam ou se distanciam umas das outras em torno do que está a ser discutido durante o processo de aprovação das Resoluções. Com esse dado, a associação por proximidade e intensidade entre os termos produzidos pelas delegações torna possível verificar a correlação entre as ideias e os interesses dos atores envolvidos no debate.

Disso posto, importa reiterar que, independentemente do nível de complexidade da técnica operada com o IRAMUTEQ, os resultados somente podem ser analisados a partir de uma interpretação teoricamente orientada e em função do contexto em que os atores estão inseridos. Somente assim é possível produzir inferências a partir das evidências encontradas (Camargo & Justo, 2013; Camargo, 2013).

Conclusão

Essa pesquisa se desenvolveu a partir de uma premissa didático-pedagógica de apresentar uma ferramenta computacional que possa suplementar a pesquisa elaborada a partir de uma cultura metodológica qualitativa. Para tanto buscou responder a seguinte questão: Como utilizar ferramentas computacionais, como um *software*, na pesquisa qualitativa?

No processo de responder esse problema, argumentou-se a necessidade de um maior refinamento metodológico a partir de uma estratégia de pesquisa que passa a utilizar ferramentas computacionais, como um *software*, para otimizar a análise de dados qualitativos, como textos, e quando sua disponibilidade se ocorre em demasia. Nesse sentido, o recurso ao *software* para análise

quantitativa de dados qualitativos tendo a complementar a pesquisa qualitativa, desde que operacionalizado de maneira adequada e a partir de um substrato teórico que permita a compreensão dos dados.

Dessa forma, depois de uma breve apresentação histórica do que é AC e de como essa técnica é reforçada, com o advento de uma máquina, para auxiliar a análise dos dados, a AAC, quando operada a partir de *software*, como o IRAMUTEQ, tem a potencialidade de diminuir a interferência da subjetividade do pesquisador sobre os dados, dada a característica desse *software* produzir o próprio *corpus* de análise e segmentos de texto para a produção de *clusters* de palavras.

Embora dirima a interferência da subjetividade do pesquisador, já que uma pesquisa qualitativa tem como pressuposto básico um caráter reflexivista, a atuação do pesquisador sobre os dados se faz imprescindível, desde que teoricamente orientada. É a teoria, na AAC, que permite a compreensão do que é gerado com a operacionalização das técnicas acima especificadas.

Com isso, esse trabalho encontra potencial em complementar pesquisas qualitativas em Relações Internacionais, notadamente no subcampo de Análise de Política Externa. Com o IRAMUTEQ, a análise de dados, quando produzidos por líderes quando se expressam, seja em entrevistas orais ou escritas ao serem indagados sobre objetivos, interesses e diretrizes de política externa de seus respectivos países, torna-se mais prática e objetiva.

Referências

- BARDIN, Laurence. (1977). **Análise de conteúdo**. Lisboa: Edições 70.
- CAMARGO, Brígido. (2013). IRAMUTEQ: tutorial para uso do software de análise textual. Disponível em: [www.iramuteq.org/documentation/fichiers/tutoriel-en-portugais]. Acesso: 15/07/2021.
- CAMARGO, Brígido; JUSTO, Ana Maria. (2013). IRAMUTEQ: um software gratuito para análise de dados textuais. **Temas em Psicologia**, 21 (2): 513-518. <http://dx.doi.org/10.9788/TP2013.2-16>.
- CAMPBELL, David. (1992). **Writing Security: United States Foreign Policy and the Politics of Identity**. Minneapolis: University of Minnesota Press.
- CAREGNATO, Rita Catalina; MUTTI, Regina. (2006). Pesquisa qualitativa: análise de discurso *Versus* análise de conteúdo. **Texto Contexto Enfermagem**, 15 (4): 679-684. <http://dx.doi.org/10.1590/S0104-07072006000400017>.
- CARLOMAGNO, Márcio C.; ROCHA, Leonardo C. da. (2016). Como criar e classificar categorias para fazer análise de conteúdo: uma questão metodológica. **Revista Eletrônica de Ciência Política**, 7 (1): 173-188. <http://dx.doi.org/10.5380/recp.v7i1.45771>.
- CERVI, Emerson U. (2018). Análise de Conteúdo Automatizada para Conversações em Redes Sociais Online: uma proposta metodológica. In **42º Encontro Anual da ANPOCS**. São Paulo: ANPOCS. 22 a 26 de outubro de 2018; Trabalho. Disponível em: [https://www.anpocs.com/

index.php/papers-40-encontro-3/gt-31/gt17-22/11253-analise-de-conteudo-automatizada-para-conversacoes-em-redes-sociais-online-uma-proposta-metodologica/file]. Acesso: 15/07/2021.

DRISKO, James; MASCHI, Tina. (2016). **Analysis Content**. Nova York: Oxford University Press.

GRIMMER, Justin; STEWART, Brandon. (2013). Text as Data: the promise and pitfalls of automatic content analysis methods for political texts. **Political Analysis**, 21: 267-297. <http://dx.doi.org/10.1093/pan/mps028>.

IZUMI, Maurício; MOREIRA, Davi. (2018). O texto como dado: desafios e oportunidades para as ciências sociais. **BIB**, 86: 138-174. Disponível em: [https://www.anpocs.com/index.php/bib-pt/bib-86/11215-o-texto-como-dado-desafios-e-oportunidades-para-as-ciencias-sociais/file]. Acesso: 15/07/2021.

JANIS, Irving. (1979). O problema da validação da análise de conteúdo. In LASSWELL, Harold; KAPLAN, Abraham. **A Linguagem da Política**. Brasília: UnB.

KRIPPENDORFF, Klaus. (2004). **Content Analysis: An Introduction to Its Methodology**. London: Sage Publications.

LASSWELL, Harold; KAPLAN, Abraham. (1979). **A Linguagem da Política**. Brasília: UnB.

MONROE, Burt; SCHRODT, Philip. (2008). Introduction to the Special Issue: the Statistical Analysis of Political Text. **Political Analysis**, 16: 351-355. <http://dx.doi.org/10.1093/pan/mpn017>.

NEUENDORF, Kimbely. (2002). **The Content Analysis Guidebook**. California: SAGE Publications.

SALVADOR, Pétala T. C. de O., et al. (2018). Uso do *software Iramuteq* nas pesquisas brasileiras da área da saúde: uma *Scoping Review*. **Revista Brasileira em Promoção da Saúde**, 31: 1-9.

STEMLER, S. (2001). An overview of content analysis. **Practical Assessment, Research, and Evaluation**, 7: 17. <https://doi.org/10.7275/z6fm-2e34>