

## **O Uso do *Software* Iramuteq: Fundamentos de Lexicometria para Pesquisas Qualitativas**

**Yuri Sá Oliveira Sousa\***

Universidade Federal da Bahia, Salvador - UFBA, BA, Brasil  
ORCID: <http://orcid.org/0000-0002-8713-5543>

### **RESUMO**

A análise de dados qualitativos tem sido auxiliada por diferentes programas computacionais (e.g., Alceste, Iramuteq, TXM, Lexico). O Iramuteq é um software gratuito que auxilia o tratamento de dados textuais e oferece diferentes possibilidades de análise baseadas na estatística de texto, ou lexicometria. O artigo buscou apresentar essa abordagem e oferecer fundamentos teórico-metodológicos para o uso do Iramuteq em pesquisas qualitativas. O texto discorre sobre os princípios da lexicometria e apresenta as características, potencialidades e limites das diferentes técnicas de tratamento lexical viabilizadas pelo Iramuteq: Estatísticas Textuais Clássicas, Análise de Especificidades, Análise de Similitude, Análise Fatorial por Correspondência, Classificação Hierárquica Descendente (método Reinert - Alceste), Nuvem de Palavras e Análise Prototípica de evocações livres. Espera-se que o trabalho possa estimular a utilização crítica e contextualizada do Iramuteq em pesquisas qualitativas, assim como contribuir para a formação de estudantes de pós-graduação e pesquisadores da psicologia e áreas afins que possam se beneficiar do uso da ferramenta em seus estudos.

**Palavras-chave:** metodologia, pesquisa qualitativa, análise lexical, iramuteq.

## **The Use of the Iramuteq Software: Fundamentals of Lexicometry for Qualitative Research**

### **ABSTRACT**

Qualitative data analysis has been aided by various computer programs (e.g., Alceste Iramuteq, TXM, Lexicon). Iramuteq is a free software that helps the treatment of textual data and offers different possibilities of analysis based on text statistics, or lexicometry. This paper aims to present this approach and provide theoretical and methodological foundations for the use of Iramuteq in qualitative research. The text discusses the principles of lexicometry and presents the characteristics, potentialities and limits of the different lexical treatment techniques made possible by Iramuteq: Classic Textual Statistics, Specificity Analysis, Similitude Analysis, Factorial Analysis by Correspondence, Descending Hierarchical Classification (Reinert's method - Alceste), Word Cloud and Prototypical Analysis of free evocations. It is expected that the work can stimulate the critical and

contextualized use of Iramuteq in qualitative research, as well as contribute to the training of graduate students and researchers in psychology and related fields who can benefit from the use of the tool in their studies.

**Keywords:** methodology, qualitative research, lexical analysis, iramuteq.

## **El uso del *Software* Iramuteq: Fundamentos de Lexicometría para la Investigación Cualitativa**

### **RESUMEN**

El análisis de datos cualitativos ha sido ayudado por diferentes programas de computadora (e.g., Alceste, Iramuteq, TXM, Lexico). Iramuteq es un *software* gratuito que ayuda al procesamiento de datos textuales y ofrece diferentes posibilidades de análisis basadas en estadísticas de texto o lexicometría. El artículo buscó presentar este enfoque y ofrecer fundamentos teóricos y metodológicos para el uso de Iramuteq en la investigación cualitativa. El texto analiza los principios de la lexicometría y presenta las características, potencialidades y límites de las diferentes técnicas de tratamiento léxico de Iramuteq: Estadística Textual Clásica, Análisis de Especificidad, Análisis de Similitud, Análisis Factorial por Correspondencia, Clasificación Jerárquica Descendente (método Reinert - Alceste), Nube de Palabras y Análisis Prototípico de evocaciones libres. Se espera que el trabajo pueda estimular el uso crítico y contextualizado de Iramuteq en la investigación cualitativa, así como contribuir a la capacitación de estudiantes graduados e investigadores en psicología y campos relacionados que puedan beneficiarse del uso de la herramienta en sus estudios.

**Palabras clave:** metodología, investigación cualitativa, análisis léxico, iramuteq.

Após algumas décadas de desenvolvimento em diferentes contextos disciplinares, o campo da pesquisa qualitativa é atualmente caracterizado pela diversidade de métodos e abordagens. A despeito dessa heterogeneidade, os estudos qualitativos possuem como interesse geral a classificação e interpretação de dados linguísticos ou visuais com o objetivo de realizar inferências sobre dimensões e estruturas da produção de significados (Flick, 2014). A utilização do computador como ferramenta de auxílio a análises de dados qualitativos ganhou popularidade com o desenvolvimento de dois conjuntos de recursos. De um lado, encontram-se os *softwares* do tipo CAQDAS – *Computer Assisted Qualitative Data Analysis Software*, cujo objetivo é facilitar o processo de codificação e organização do material de análise em categorias (e.g., Atlas.ti., NVivo, MaxQDA, QDA Miner). De outro, identificam-se ferramentas baseadas na lexicometria, que oferecem opções de tratamento automático do

texto a partir de cálculos estatísticos sobre o vocabulário do *corpus* (e.g., Alceste, Lexico, TXM, Iramuteq) (Dany, 2016). É sobre o segundo conjunto de recursos que este trabalho se debruça. Mais especificamente, o presente artigo tem o objetivo de oferecer fundamentos teórico-metodológicos para a utilização do *software* Iramuteq na análise de dados qualitativos.

A proposta pode ser justificada pelo fato de haver poucos trabalhos teóricos sobre estatística textual e análise lexical publicados em língua portuguesa (e.g., Justo & Camargo, 2014; Nascimento & Menandro, 2006), o que pode limitar ou dificultar as primeiras aproximações com o uso de *softwares* baseados na lexicometria. Ainda que seja possível encontrar manuais e textos técnicos sobre o Iramuteq (e.g., Camargo & Justo, 2013, 2018; Salviati, 2017), pouca atenção tem sido dada aos fundamentos lexicométricos de suas diferentes possibilidades de aplicação. Nesse sentido, acredita-se que o artigo pode contribuir de modo significativo para a formação de estudantes de pós-graduação e pesquisadores que desenvolvem estudos qualitativos na psicologia e áreas afins.

A estrutura do artigo está organizada em duas seções. Inicialmente, a lexicometria será apresentada e contextualizada em relação a outras abordagens metodológicas que também são utilizadas em pesquisas qualitativas. Adicionalmente, busca-se caracterizar o Iramuteq como um programa informatizado, voltado à análise de dados textuais, cujos procedimentos fundamentam-se na estatística textual. A segunda seção visa descrever os principais procedimentos lexicométricos viabilizados pelo Iramuteq em função dos seus pressupostos teóricos e operacionais.

### **A Abordagem da Lexicometria e o Uso de *Softwares* na Pesquisa Qualitativa**

Ainda que existam nuances de significado (cf. Leblanc, 2015), termos como estatística textual, textometria, logometria, análise do discurso assistida por computador e tratamento automático do discurso guardam relação entre si pelo tipo de abordagem em que se apoiam: a lexicometria (Marchand, 2013). Por sua vez, a lexicometria pode ser definida como um conjunto de procedimentos baseados em critérios formais que permitem reorganizar a estrutura de um texto ou conjunto de textos, assim como realizar cálculos estatísticos a partir do seu vocabulário (Salem, 1986). Assim, a abordagem pode ser caracterizada como uma estratégia que aplica métodos quantitativos (estatística descritiva e inferencial) a dados qualitativos (textos) com o objetivo de realizar observações sobre as características de um conjunto de comunicações (e.g., aspectos semântico-lexicais e pragmáticos). Diante da diversidade metodológica do campo, Leblanc (2015) distingue as ferramentas lexicométricas

tradicionais, baseadas na análise das características linguísticas de um texto, como frequência de palavras, hápax, formas gramaticais, vocabulário utilizado etc., das ferramentas lexicométricas estruturantes, cujo objetivo é reorganizar o *corpus* em estruturas significantes salientes. Estas últimas buscam identificar relações entre as palavras e o seu grau de “vizinhança” com diferentes textos ou partes de um texto. A análise dirige-se, portanto, ao modo como o vocabulário se distribui no *corpus*, com a possibilidade de realizar observações baseadas em probabilidades (Leblanc, 2015). Para Marchand (2013), análises lexicométricas conduzem à pergunta sobre “por que, no universo de palavras – e categorias de palavras – possíveis, aquelas foram escolhidas e quais relações elas estabelecem entre si e com as suas condições de produção” (Marchand, 2013, p. 39, tradução nossa). Nessa direção, Reinert (2009) salienta que a análise lexical é particularmente proveitosa quando utiliza variáveis categóricas de contexto (características do enunciador, da situação comunicativa e/ou da audiência) no tratamento dos dados, permitindo considerar as relações entre os textos e suas condições de produção e recepção. Por exemplo, a análise lexical permite perscrutar a existência de diferenças significativas no conteúdo produzido por participantes de uma pesquisa em função das suas características e grupos de pertença.

Inicialmente, essa caracterização permitiria aproximar a lexicometria da análise de conteúdo que, em sentido amplo, designa um conjunto de técnicas de análise das comunicações que possuem o objetivo de realizar inferências acerca das condições de produção de uma mensagem e seus efeitos em dada realidade (Bardin, 1977). Não por acaso, é possível encontrar comparações explícitas entre os resultados de uma análise de conteúdo clássica e uma análise lexicométrica (Oliveira, Teixeira, Fischer, & Amaral, 2003; Reinert, 1983). Apesar das aproximações, é preciso ressaltar que, enquanto a lexicometria sempre faz uso da quantificação, o uso de técnicas quantitativas na análise de conteúdo não é uma condição *sine qua non*, mas uma possibilidade (Bardin, 1977). De toda maneira, tanto a análise de conteúdo quanto a lexicometria exigem a adoção de procedimentos sistemáticos e objetivos, no sentido de garantir que as etapas envolvidas no processo de análise possam ser explicitadas e replicadas (Lebart & Salem, 1994). Particularmente, a lexicometria pode ser útil nos seguintes casos: quando a unidade de registro é a palavra e o indicador principal é a sua frequência; quando a análise é complexa e multivariada, o que exige um tratamento simultâneo de categorias e unidades de registro; quando se deseja analisar a coocorrência de palavras em unidades de contexto; quando a investigação possui etapas sucessivas e a análise demanda tratamentos estatísticos complexos (Bardin, 1977).

Além da aproximação com a análise de conteúdo, as técnicas lexicométricas são algumas vezes caracterizadas como ferramentas de análise de discurso (Leblanc, 2015; Marchand, 2013; Reinert, 2001, 2002; Salem, 1986), o que pode ser interpretado como uma forma de demarcação implícita de premissas teóricas relacionadas ao campo da linguagem e comunicação. Nessa direção, o texto analisado não é entendido apenas como suporte formal da comunicação, mas sim como produto de interações sócio-discursivas entre o locutor, a audiência, suas respectivas pertencas sociais e o contexto comunicativo (Marchand, 2013). Ademais, um texto é produto das escolhas realizadas pelo locutor, mas também decorre da sua relação com outros enunciados disponíveis em dado espaço e tempo (Salem, 1986). Assim, o uso da lexicometria na análise do discurso visa identificar regularidades, tendências e estilos discursivos em um conjunto de textos, podendo ser orientada tanto às dimensões formais do material (e.g., estrutura e características linguísticas), como aos conteúdos semânticos que podem ser identificados nos eventos comunicativos analisados (Leblanc, 2015).

As primeiras aplicações da estatística textual surgiram na França por volta da década de 1970. O seu desenvolvimento nas ciências humanas e sociais esteve particularmente atrelado à análise de textos políticos, como os panfletos do Maio de 1968, onde se pretendia mensurar as características do seu vocabulário e conteúdo (Marchand, 2013). Ainda na década de 1970, o estatístico Jean-Paul Benzécri (1977) havia lançado as bases da análise de correspondência, um método exploratório multidimensional que podia ser aplicado a dados textuais (Lebart & Salem, 1994). A inovação do método consistia na possibilidade de considerar, simultaneamente, as múltiplas relações estabelecidas entre as variáveis observadas (Pereira, 1987).

No final dos anos 1980, o *software* Alceste – *Analyse des Lexèmes Cooccurrents dans les Enoncés Simplifiés d'un Texte* – foi desenvolvido por Max Reinert (1986) e trouxe inovações metodológicas que foram incorporadas ao funcionamento de outros programas, como é o caso do próprio Iramuteq. De acordo com Leblanc (2015), antes do lançamento do Alceste, os métodos tradicionais consistiam prioritariamente em decompor os textos em unidades mínimas (ocorrências e formas lexicais) para, em seguida, descrever os resultados em termos de frequências e analisar a distribuição do vocabulário em diferentes partições, divisões ou categorias constitutivas do *corpus*. Ao seu turno, o algoritmo que fundamenta o Alceste (Reinert, 1983, 1987) permitia trabalhar com um *corpus* não particionado, reestruturando-o em classes lexicais em função da coocorrência de formas em segmentos de texto (Leblanc, 2015). De todo modo, a operacionalização de variáveis categóricas permanece sendo uma das principais potencialidades da abordagem lexicométrica, uma vez que permite

analisar comparativamente o vocabulário de diferentes divisões de um conjunto de textos, sejam elas naturais ou experimentais (Leblanc, 2015). O Alceste tem sido largamente utilizado em pesquisas qualitativas de diferentes áreas e, na psicologia, é possível destacar aquelas desenvolvidas no campo teórico das Representações Sociais (Kalampalikis, 2003). Essas pesquisas utilizam a ferramenta com o objetivo principal de identificar diferentes formas de discurso relacionadas a um fenômeno ou tópico de interesse (Nascimento & Menandro, 2006).

Com o passar dos anos, outros programas voltados à análise estatística de texto foram desenvolvidos (e.g., Calliope, Prospéro, Trideux, Tropes), mas os *softwares* de livre acesso surgem apenas no final dos anos 2000, como é o caso do TXM, em 2007, e do Iramuteq, em 2009 (Marchand, 2013). O fato de esses *softwares* serem gratuitos pode fazer com que pesquisadores passem a utilizá-los como uma alternativa às ferramentas lexicométricas comercializadas. Por exemplo, o Iramuteq possui um procedimento chamado “método Reinert”, que utiliza o mesmo algoritmo do Alceste (Reinert, 1983, 1986), além de oferecer outros recursos de tratamento textual. De todo modo, um estudo de revisão que analisou o uso de *softwares* em pesquisas qualitativas da área da saúde (Santos et al., 2017) identificou que tanto o Alceste quanto o Iramuteq têm sido empregados com a mesma frequência. Além disso, ressalta-se que o principal recurso do Iramuteq utilizado nos estudos foi o método Reinert (CHD), empregado em aproximadamente 85% das pesquisas. Esses resultados parecem indicar que o Iramuteq tem sido procurado como uma alternativa gratuita ao Alceste, mas também convidam a refletir sobre o pouco aproveitamento do *software* em relação às suas possibilidades de aplicação.

Conforme é possível depreender até aqui, o Iramuteq – *Interface de R pour les Analyses Multidimensionnelles de Textes et de Questionnaires* (Camargo & Justo, 2013; Loubère & Ratinaud, 2014; Ratinaud, 2014) – é um *software* gratuito destinado à realização de análises estatísticas de textos e questionários. Desenvolvido por Pierre Ratinaud (2014) no *Laboratoire d'Études et de Recherches Appliquées en Sciences Sociales* (LERASS), a ferramenta funciona como uma interface de R ([www.r-project.org](http://www.r-project.org)) que auxilia a organização e tratamento estatístico de dados textuais. Criado inicialmente em língua francesa, o *software* começou a ser adaptado para o português em 2011 (Camargo & Justo, 2013) e, atualmente, oferece suporte para tratamento e análise de dados nessa língua. O Iramuteq reúne um conjunto de procedimentos lexicométricos distintos, tais como: descrições estatísticas clássicas; Análise de Especificidades linguísticas; Classificação Hierárquica Descendente (CHD), ou método Reinert; Análise Fatorial por Correspondência (AFC); Análise de

Similitude e Análise Prototípica de evocações. Os fundamentos lexicométricos de cada um desses recursos serão apresentados ao longo do artigo, mas, para uma descrição operacional, recomenda-se consultar manuais e textos técnicos publicados com esse fim (Camargo & Justo, 2013, 2018; Loubère & Ratinaud, 2014; Péliissier, 2017; Ratinaud, 2014).

Importa enfatizar que a escolha por utilizar um ou outro recurso de tratamento dos dados deve estar fundamentada em uma reflexão teórico-metodológica quanto ao problema e aos objetivos da pesquisa (Leblanc, 2015). De modo geral, conforme discutem Lebart e Salem (1994), o uso da estatística na análise de dados textuais pode ser decomposto em quatro etapas: (1) construção de um **problema** de pesquisa que viabilize o emprego da estatística textual como abordagem; (2) produção de **dados** pertinentes ao problema; (3) **tratamento** estatístico com finalidades exploratórias ou baseadas em testes de hipóteses; e (4) **interpretação** dos resultados com base em reflexões críticas sobre a validade e significado das observações. Cada uma dessas etapas mobiliza questões que podem influenciar as escolhas do pesquisador, de modo que o potencial heurístico das ferramentas de análise utilizadas depende do planejamento da pesquisa como um todo.

Além disso, é importante conhecer os fundamentos operacionais que caracterizam cada procedimento para que os resultados possam ser interpretados de maneira crítica e parcimoniosa. A aparente facilidade de uso e a beleza das representações gráficas obtidas nos resultados não podem prescindir de uma reflexão contextualizada (Péliissier, 2017), pois “em toda análise que mobiliza uma representação é importante, antes de tudo, saber quais são os dados que são representados, como eles foram coletados, mensurados, e o que representa o gráfico proposto” (Leblanc, 2015, p. 51, tradução nossa). Diante disso, a seção seguinte pretende explicitar os fundamentos operacionais das principais técnicas lexicométricas viabilizadas pelo Iramuteq, indicar contextos de aplicação e fornecer subsídios teóricos para sua utilização em pesquisas qualitativas.

### **As Técnicas Lexicométricas do Iramuteq: Fundamentos Teóricos e Operacionais**

Conforme mencionado, o Iramuteq reúne um conjunto de possibilidades distintas de tratamento de dados qualitativos fundamentadas na lexicometria, oferecendo recursos tanto tradicionais como estruturantes na análise de textos e questionários. As diferentes técnicas lexicométricas são iniciadas por um processo de segmentação automática, etapa em que são reconhecidas as ocorrências que compõem um texto, de modo que dois segmentos idênticos são contabilizados como duas ocorrências da mesma forma linguística. Esse processo é

responsável por construir o vocabulário de um *corpus*, isto é, o conjunto de formas distintas que são utilizadas em todo material (Salem, 1986). Tal como realizado em outras ferramentas, o Iramuteq permite lematizar as ocorrências do *corpus* com o objetivo de reduzir a variabilidade do vocabulário, o que permite conferir maior homogeneidade ao material que será submetido às diferentes técnicas de tratamento lexicométrico. Isso significa que a qualidade dos resultados também depende da qualidade do material analisado (Pélissier, 2017). Por conta disso, alguns autores chegam a recomendar modificações no *corpus* para melhorar a legibilidade do *software* nos seus procedimentos automáticos (e.g., Aubert-Lotarski & Capdevielle-Mougnibas, 2002). Por sua vez, a etapa da lematização permite reagrupar os verbos ao infinitivo, os substantivos ao singular e os adjetivos ao singular masculino (Salem, 1986). Tanto a segmentação automática como a lematização são processos que permitem transformar o conteúdo linguístico em variáveis binárias, ou dicotômicas, cujos indicadores são utilizados para quantificar e representar as características do conteúdo analisado. Por exemplo, isso permite saber quantas vezes cada palavra lematizada foi utilizada no conjunto de textos ou em subdivisões específicas.

As informações gerais sobre o conjunto de textos podem ser exploradas com o auxílio de **Estatísticas Textuais Clássicas**, que são aquelas que permitem verificar o número de textos incluídos no *corpus* (e.g., respostas a questionários abertos, entrevistas, matérias de jornal); saber a frequência das ocorrências e das formas linguísticas lematizadas; detectar os hápax (i.e. formas lexicais que ocorrem apenas uma vez); identificar o número médio de ocorrências por texto; e sumarizar as características do vocabulário que constitui o *corpus* (e.g., frequência e classe gramatical de cada forma). No âmbito das pesquisas qualitativas, as estatísticas textuais clássicas permitem explorar aspectos formais de um texto ou conjunto de textos, facilitando a identificação de estilos discursivos que caracterizam a produção linguística de determinado autor, grupo, fonte documental ou contexto de coleta. Por exemplo, Lebart e Salem (1994) mencionam o uso da lexicometria na análise de textos literários com o objetivo de mensurar a extensão do vocabulário de um autor em diferentes obras, bem como avaliar a sua evolução ao longo do tempo. Caso seja necessário ir além das informações descritivas e realizar uma análise comparativa baseada na estatística inferencial, o método da Análise de Especificidades pode se revelar heurísticamente útil.

A **Análise de Especificidades**, ou de contraste, é um método que visa comparar a distribuição de formas linguísticas em diferentes partições de um texto com a finalidade de realizar observações baseadas em probabilidades (Leblanc, 2015). As partições são automaticamente construídas com base nas variáveis categóricas que forem associadas a cada



texto do *corpus* na etapa de formatação e preparação do material (cf. Camargo & Justo, 2018, p. 7-10; Salviati, 2017, p. 16-20). Por exemplo, a análise de dados provenientes de entrevistas ou de questionários abertos pode considerar as características dos indivíduos na construção de partições com finalidades comparativas (e.g., sexo, raça/cor, faixa etária, grau de escolaridade, faixa de renda, grupos de pertença, etc.). O mais importante nesse tipo de análise não é avaliar a frequência de cada forma, mas verificar o quão improvável é a sua distribuição nas diferentes partições. A técnica pode ser ilustrada por um dos estudos realizados por Justo (2016), onde foram analisadas entrevistas semidirigidas sobre corpo, obesidade e sobrepeso. A autora considerou a existência de especificidades lexicais no discurso de participantes com e sem excesso de peso, evidenciando a relação entre os conteúdos de representação compartilhados e os grupos de pertença dos indivíduos entrevistados.

Tal abordagem permite ultrapassar as limitações de uma análise descritiva focada exclusivamente nas frequências observadas, que são bastante sensíveis ao tamanho da partição considerada. De outro modo, busca-se produzir indicadores probabilísticos sobre a banalidade ou a especificidade do uso de cada forma linguística nas diferentes subdivisões do *corpus* (Lafon, 1980). A configuração padrão da Análise de Especificidades realizadas pelo Iramuteq parte do modelo de distribuição hipergeométrica e considera quatro indicadores no cálculo das probabilidades (Leblanc, 2015): o número de ocorrências no *corpus*, o número de ocorrências na partição, a frequência de cada forma no *corpus* e a frequência de cada forma na partição. Assim, o procedimento contrasta o conteúdo das partes e do todo para, então, realizar julgamentos de probabilidade que indicam a sub-representação ou sobre-representação de cada forma nas divisões do *corpus* (Leblanc, 2015). Também é possível realizar a Análise de Especificidades com base em outros testes estatísticos, como é o caso do qui-quadrado ( $\chi^2$ ). No entanto, o modelo de distribuição hipergeométrica aparenta ser mais vantajoso, uma vez que o seu resultado se mantém preciso em diferentes condições de tamanho amostral (Lafon, 1980).

Com base no modelo de distribuição hipergeométrica, o Iramuteq fornece um indicador de especificidade para as diferentes formas lexicais em cada partição analisada. Quanto mais próximo de zero for o indicador, mais banal (genérica) é a sua distribuição, evidenciando que a frequência observada da forma se aproxima da frequência esperada (Leblanc, 2015). Tecnicamente, o indicador de especificidade do Iramuteq representa o logaritmo decimal de uma probabilidade. Por exemplo, um resultado maior do que 3 indica que há menos de uma chance em mil de obter uma frequência tão elevada da palavra na

partição quanto aquela que foi encontrada (Pateyron, Weber, & Germain, 2015). Para efeitos práticos, isso significa que qualquer resultado maior do que 2 possui um valor-p menor ou igual a 0,01, o que favorece a hipótese de que há tendências na distribuição da forma lexical em função da variável categórica considerada. Além disso, indicadores positivos denotam a sobre-representação da forma, ao passo que valores negativos expressam a sua sub-representação na partição.

Deve-se destacar que Análise de Especificidades também pode ser realizada após uma categorização temática dos dados. Tal procedimento é viabilizado pelo recurso TGen (*Types Généralisés*), que permite reagrupar um léxico específico para tratamento posterior (Aubert-Lotarski & Capdevielle-Mougnibas, 2002). Nesse contexto, trata-se de investigar se há tendências na distribuição de um conjunto de formas que são analisadas simultaneamente em função das variáveis categóricas de interesse. Apesar das suas potencialidades, a literatura carece de exemplos de utilização do recurso em análises de especificidades com o Iramuteq.

Quando a Análise de Especificidades é realizada a partir de uma variável com pelo menos três modalidades (Camargo & Justo, 2018), também é possível operacionalizar uma **Análise Fatorial por Correspondência (AFC)**. As técnicas baseadas na análise de correspondência permitem transformar dados categóricos em medidas quantitativas, por meio de tabelas de contingência, para, em seguida, realizar uma redução dimensional. Frequentemente, essas análises produzem representações gráficas sobre as relações de interdependência existentes entre as variáveis consideradas (Hair, Black, Babin, Anderson, & Tatham, 2009). No caso da AFC, entende-se, por princípio, que as relações entre as variáveis de uma matriz de contingência particionada apresentam regularidades que permitem efetuar a redução dos dados em alguns fatores (Marchand, 2013). O algoritmo da AFC considera o cruzamento entre as formas linguísticas de um *corpus* e as diferentes partições que o compõem para, em seguida, apresentar os resultados em um plano fatorial. O plano fatorial pode ser entendido como um mapa perceptual, que exhibe a posição relativa de todas as variáveis analisadas (Hair et al., 2009).

De acordo com Pélissier (2017), a representação gráfica de uma AFC pode ser interpretada a partir das oposições encontradas entre os elementos que são distribuídos nos dois eixos (horizontal e vertical) (Leblanc, 2015). Conforme explicam Lebart e Salem (1994), essas representações são fundamentadas na álgebra linear e traduzem as associações estatísticas entre as linhas de uma tabela de continência (e.g., indivíduos, textos) e as suas colunas (e.g., variáveis categóricas e formas lexicais). Assim, além de oposições estruturantes, a representação gráfica permite visualizar relações de “vizinhança” entre

variáveis e formas que compõem um *corpus* textual. Cumpre observar que, embora a AFC possa ser executada de modo conjugado com a Análise de Especificidades, o método tem sido predominantemente utilizado em análises baseadas na Classificação Hierárquica Descendente.

Conforme mencionado anteriormente, a **Classificação Hierárquica Descendente (CHD)**, ou método Reinert, fundamenta-se no algoritmo base do *software* Alceste (cf. Reinert, 1983, 1987, 2001, 2002). Essa forma de tratamento foi elaborada com o propósito de ser aplicada a dados textuais provenientes de entrevistas e questionários abertos (Reinert, 1983), mas passou a ser utilizada na análise de um conjunto bastante diverso de materiais, incluindo documentos literários (e. g., Reinert, 1987, 2001), políticos (e.g., Ratinaud & Marchand, 2016) e midiáticos (e.g., Marty, Marchand, & Ratinaud, 2013). A CHD pode ser descrita como uma análise de agrupamentos (*clusters*) em que os segmentos de texto de um *corpus* são sucessivamente particionados em função da coocorrência de formas lexicais. Conforme explicam Hair et al. (2009), as análises de agrupamentos têm o objetivo de classificar uma amostra de entidades (e.g., indivíduos, objetos, textos) em um número menor de grupos mutuamente excludentes. Esses grupos são definidos após o tratamento dos dados e consideram similaridades encontradas entre as entidades analisadas. No contexto da CHD, a análise de agrupamentos é utilizada para identificar grupos de segmentos de texto que possuem vocabulários semelhantes entre si.

Do ponto de vista operacional, esse procedimento ocorre em quatro etapas (cf. Nascimento & Menandro, 2006; Oliveira et al., 2003). Na primeira, o *software* realiza uma leitura automática do *corpus* textual, divide os textos em segmentos de texto com base no tamanho do enunciado e pontuação, lematiza as ocorrências e calcula os dicionários que serão utilizados nas etapas subsequentes. Aqui ocorre a distinção entre formas ativas (e.g., verbos, substantivos e adjetivos) e formas suplementares (e.g., pronomes, artigos e advérbios). Essa diferenciação é fundamental, pois a maior parte dos cálculos é realizada com as formas ativas, as quais indicam características semânticas dos enunciados (Nascimento & Menandro, 2006). Na segunda etapa, são construídas as matrizes de contingência com os segmentos de texto e seus respectivos vocabulários para, então, realizar a Classificação Hierárquica Descendente (CHD) propriamente dita. Inicialmente, o conjunto de segmentos de texto é dividido em duas classes a partir do contraste entre os seus vocabulários. O procedimento é continuamente repetido nos agrupamentos resultantes até que não sejam produzidas novas classes estáveis (Nascimento & Menandro, 2006; Oliveira et al., 2003). Em seguida, são realizados testes de qui-quadrado ( $\chi^2$ ) que verificam o grau de associação entre as formas lexicais e as classes, apresentando um dendrograma com as formas mais específicas de cada partição. A terceira

etapa produz os perfis lexicais de cada classe, reunindo informações detalhadas sobre a distribuição de formas ativas e suplementares, bem como os resultados detalhados dos testes de qui-quadrado ( $\chi^2$ ). Nessa fase também é realizada uma Análise Fatorial por Correspondência (AFC), que cruza as formas reduzidas e variáveis categóricas com as classes resultantes (Nascimento & Menandro, 2006). Por fim, a quarta etapa realiza cálculos complementares e identifica os segmentos de texto mais característicos de cada agrupamento, o que permite resgatar os contextos de enunciação das formas lexicais.

Ao final desse procedimento, nem todos os segmentos de texto serão classificados e o percentual do material aproveitado será exibido nos resultados. De modo geral, considera-se pouco proveitoso trabalhar com uma classificação que retenha menos de 75% do material (Camargo & Justo, 2018). Por outro lado, Pélissier (2017) argumenta que discursos orais, por exemplo, tendem a ser menos homogêneos e podem apresentar níveis de retenção mais baixos. De toda maneira, classificações com menos de 60% de retenção provavelmente indicam problemas na homogeneidade do *corpus* (Pélissier, 2017). Além disso, os resultados de uma CHD são bastante sensíveis tanto ao tipo e tamanho do *corpus*, como aos parâmetros de tratamento utilizados (Reinert, 2009). Por essa razão, nem sempre a utilização das configurações padronizadas oferece a melhor solução para a apresentação dos resultados. Quando se tem um *corpus* volumoso, por exemplo, pode ser útil aumentar o número de classes terminais a fim de alcançar resultados mais detalhados, dada a possibilidade de identificar campos lexicais menores, porém mais homogêneos do que aqueles obtidos em um tratamento padrão (Pélissier, 2017).

Os resultados do método Reinert podem ser apresentados de diferentes maneiras. Uma representação bastante comum é o dendrograma, que reúne as formas linguísticas mais características (específicas) de cada classe. Também é possível apresentar segmentos de texto típicos de cada agrupamento, que são aqueles que reúnem diferentes formas linguísticas características da classe. Isso permite, por exemplo, resgatar os contextos de enunciação em que o vocabulário do agrupamento é empregado. Além das maneiras citadas, os resultados podem ser sumarizados a partir de informações detalhadas sobre o perfil de cada classe lexical, da exportação do *corpus* em cores, assim como da apresentação de um plano fatorial construído pelo método da Análise Fatorial por Correspondência (AFC). Nesse contexto específico, a AFC considera as classes lexicais como partições do texto (Pélissier, 2017), cruzando, assim, as formas reduzidas e variáveis de contexto dos enunciados com os agrupamentos resultantes da classificação (Nascimento & Menandro, 2006).

Qualquer que seja a forma de apresentação escolhida, os resultados devem ser entendidos como um auxílio à interpretação da realidade e não como a própria realidade (Reinert, 2009). Ademais, também pode ser útil combinar a análise lexical com outros métodos, como a própria análise de conteúdo tradicional (Nascimento & Menandro, 2006; Oliveira et al., 2003). Nesse ponto, guardadas as especificidades de cada abordagem, tanto o método Reinert (CHD) como a análise de conteúdo partem de princípios semelhantes: a construção de categorias com base no princípio de exclusão mútua; a garantia de homogeneidade categorial e temática; e a valorização da objetividade na descrição dos resultados (Oliveira et al., 2003). De modo geral, a interpretação sobre o significado das classes lexicais se sustenta na premissa de que a coocorrência regular de formas linguísticas em segmentos de texto remete a representações, noções ou conceitos comuns (Reinert, 1987). Por essa razão, o método é frequentemente utilizado com o objetivo de identificar as temáticas que compõem um conjunto de textos. No campo da psicologia social, por exemplo, estudos sobre representações sociais têm interpretado as classes lexicais como indicadores da existência de discursos compartilhados sobre determinado objeto (Oliveira et al., 2003). Além do conteúdo, Reinert (1987) destaca que a CHD pode facilitar a identificação de processos subjacentes à produção dos enunciados, os quais guardam relação com a estratificação da sociedade e as respectivas posições sociais que ocupam os enunciadores. Nessa direção, reitera-se a importância da operacionalização de variáveis categóricas com a finalidade de explorar associações entre as condições de produção do texto e o conteúdo das classes lexicais.

Outro procedimento que pode ser realizado com o auxílio do Iramuteq é a **Análise de Similitude**, que normalmente utiliza a coocorrência de formas lexicais em textos ou segmentos de texto para construir representações gráficas sobre a estrutura do conteúdo de um *corpus*. Os resultados são apresentados na forma de grafos, onde as palavras constituem os vértices e as arestas representam a relação entre elas (Marchand & Ratinaud, 2012). Esse tratamento é particularmente útil na identificação da conectividade entre as formas linguísticas de um conjunto de textos, o que remete ao modo como o conteúdo se estrutura (Camargo & Justo, 2018). Ressalta-se que também é possível combinar a Análise de Similitude com uma análise de agrupamentos (*clusters*) baseada na coocorrência de palavras presentes nos textos. O resultado desse tratamento pode ser apresentado na forma de comunidades linguísticas identificadas por cores distintas no grafo gerado. Adicionalmente, o Iramuteq permite combinar, em uma mesma representação gráfica, os resultados de uma análise de similitude com aqueles obtidos em uma Análise de Especificidades. Nesse caso, a Análise de Similitude

considera alguma variável de contexto no tratamento e o grafo apresentado destaca as palavras sobrerepresentadas em cada partição ou modalidade da variável.

O último recurso do *software* aplicado a documentos textuais é a **Nuvem de Palavras**, que oferece uma representação gráfica das ocorrências do *corpus* onde o tamanho de cada forma é proporcional a sua frequência. Apesar do efeito estético, a Nuvem de Palavras é um recurso que pouco acrescenta na interpretação dos dados em contextos de pesquisa. Em primeiro lugar, a representação é construída unicamente com base na frequência de cada forma sem, entretanto, apresentar qualquer indicador suplementar. Nesse caso, a Nuvem de Palavras pode ser substituída por uma tabela de frequências, que, apesar de apresentar perdas estéticas, informa melhor sobre as características do conteúdo. Em segundo lugar, a organização das palavras na imagem não segue critérios objetivos que permitam estabelecer relações de proximidade, distanciamento ou conexidade entre os elementos do vocabulário. Em vez disso, a imagem pode levar a interpretações equivocadas sobre a existência de relações entre os elementos que são apresentados em uma mesma região da imagem. Caso a pesquisa busque realizar inferências nessa direção, outros métodos apresentados revelam-se bem mais informativos (e.g., AFC e Análise de Similitude). Apesar dessas ressalvas, a Nuvem de Palavras pode ser útil na apresentação dos resultados obtidos em um tratamento lexicométrico anterior. Por exemplo, as formas características que compõem as classes lexicais geradas em uma CHD podem ser representadas em uma Nuvem de Palavras (e.g., Calassa, Penso, & Freitas, 2015).

Além das possibilidades apresentadas, o Iramuteq permite tratar dados de matrizes do tipo indivíduo/palavras, como aquelas geradas em coletas baseadas na Técnica de Associação Livre de Palavras (TALP). Muito utilizada no campo teórico das representações sociais, essa técnica consiste em solicitar aos participantes que escrevam palavras ou expressões que lhes venham espontaneamente à cabeça a partir de um termo indutor, o que permite identificar rapidamente os principais elementos simbólicos associados ao objeto (Abric, 1994). Os termos evocados podem ser analisados pela sua frequência e, de modo complementar, por um indicador de centralidade ou saliência, como a ordem média da evocação no conjunto de dados ou a ordem média de importância atribuída pelos sujeitos (cf. Dany, Urdapilleta, & Monaco, 2015). A **Análise Prototípica** realizada no Iramuteq é baseada em um método elaborado por Vergès (1992), posteriormente incorporado ao *software* EVOC (Vergès, Scano, & Junique, 2002). O tratamento consiste em distribuir os elementos evocados em uma representação gráfica de quatro quadrantes, os quais são construídos em função dos dois indicadores citados: (1) +frequentes/+salientes; (2) +frequentes/-salientes; (3) –

frequentes/+salientes; (4) –frequentes/–salientes. Notadamente, a Análise Prototípica auxilia a interpretação dos resultados de pesquisas fundamentadas na teoria do núcleo central, característica da abordagem estrutural das representações sociais (Dany et al., 2015). Além da Análise Prototípica, também é possível realizar outros tipos de tratamento utilizando matrizes do tipo indivíduo/palavras, como os testes de qui-quadrado que permitem comparar a distribuição do vocabulário em função de variáveis categóricas, o método Reinert e a Análise de Similitude.

A despeito da pluralidade de opções oferecidas pelo Iramuteq, o procedimento mais utilizado na literatura tem sido o método Reinert (cf. Santos et al., 2017). Essa tendência pode ser explicada tanto pelo fato de o Iramuteq representar uma alternativa gratuita ao Alceste, como pelas suas próprias potencialidades. Notadamente, o método permite reorganizar o conjunto de textos em classes lexicais, frequentemente tratadas como categorias analíticas endógenas, a partir das quais os processos inferenciais podem ser desenvolvidos (Landivar, Mathonnat, & Tichit, 2015). Trata-se, portanto, de um processo de categorização empiricamente orientado, onde as categorias resultantes não dependem, inicialmente, da interpretação do pesquisador (Scelles, 1997). Apesar das potencialidades e da popularidade do método Reinert (CHD), reitera-se que a escolha por uma técnica de tratamento de dados precisa ser compatível com os objetivos e alcance da pesquisa.

Por fim, cumpre ressaltar que é possível utilizar diferentes técnicas lexicométricas com o objetivo de integrar dimensões complementares do fenômeno analisado. Por exemplo, enquanto a CHD evidencia regularidades e aspectos comuns que são identificados nos textos, a AFC permite destacar relações de oposição que estruturam o conteúdo (Aubert-Lotarski & Capdevielle-Mougnibas, 2002). A Análise de Especificidades, por sua vez, auxilia a identificação de tendências na distribuição do vocabulário em função de variáveis categóricas de contexto, o que possibilita desenvolver hipóteses a respeito da relação entre as condições de produção dos textos e o seu conteúdo. Ao seu turno, os resultados de uma Análise de Similitude representam a estrutura de relações estabelecidas entre as formas lexicais, ao mesmo tempo em que também podem evidenciar especificidades relacionadas às modalidades de uma variável de contexto (Marchand & Ratinaud, 2012). Em suma, entende-se que a utilização conjugada de diferentes procedimentos pode contribuir significativamente para ampliar o alcance das observações desenvolvidas em estudos qualitativos que empregam o Iramuteq.

## Considerações Finais

Conforme apresentado nas seções anteriores, o Iramuteq é um *software* gratuito que oferece diferentes possibilidades de análise baseadas na lexicometria. Essa abordagem encontra utilidade quando a exploração de um texto ou conjunto de textos é realizada a partir de hipóteses a respeito do uso da linguagem na construção de significados e do modo como o vocabulário se distribui em diferentes partições de *corpus*. Embora a utilização do Iramuteq tenha ganhado popularidade nos últimos anos, deve-se atentar para os riscos de que a ferramenta seja utilizada de maneira acrítica e indiscriminada. Por essa razão, argumenta-se em favor da necessidade de conhecer os fundamentos teóricos e operacionais de cada um de seus procedimentos para, assim, realizar escolhas metodológicas compatíveis com a natureza dos dados, o problema e os objetivos da pesquisa. Ainda que o artigo não seja capaz de aprofundar todas as questões suscitadas, acredita-se que o texto oferece contribuições suficientes a respeito das potencialidades e dos limites das diferentes técnicas de tratamento lexical viabilizadas pelo Iramuteq. Nesse sentido, espera-se que o presente trabalho permita ampliar as formas de utilização da ferramenta, incentivando o uso conjugado de diferentes recursos lexicométricos com o objetivo de acessar dimensões complementares do fenômeno analisado. Por fim, acredita-se que o texto também pode ser útil para respaldar reflexões em pesquisas qualitativas sobre o funcionamento de *softwares* como o Iramuteq e sobre a validade e o significado dos seus resultados.

## Referências

- Abric, J.-C. (1994). Méthodologie de recueil des représentations sociales. In J.-C. Abric (Org.), *Pratiques sociales et représentations* (p. 59-82). Paris: PUF.
- Aubert-Lotarski, A., & Capdevielle-Mougribas, V. (2002). Dialogue méthodologique autour de l'utilisation du logiciel Alceste en sciences humaines et sociales: "Lisibilité" du *corpus* et interprétation des résultats. *Actes des 6ème Journées internationales d'Analyse statistique des Données Textuelles* (pp. 45-56), Paris, France, 6. Recuperado de [http://lexicometrica.univ-paris3.fr/jadt/jadt2002/PDF-2002/aubert\\_capdevielle.pdf](http://lexicometrica.univ-paris3.fr/jadt/jadt2002/PDF-2002/aubert_capdevielle.pdf)
- Bardin, L. (1977). *Análise de conteúdo*. Lisboa: Edições 70.
- Benzecri, J.-P. (1977). Histoire et préhistoire de l'analyse des données: Partie V L'analyse des correspondances. *Les cahiers de l'analyse des données*, 2(1), 9-40.



- Calassa, G. D. B., Penso, M. A., & Freitas, L. G. (2015). Redução de danos na visão dos profissionais que atuam no CAPS AD II do Distrito Federal. *Psicologia em Pesquisa*, 9(2), 177–187. doi: 10.5327/Z1982-1247201500020008
- Camargo, B. V., & Justo, A. M. (2013). IRAMUTEQ: Um software gratuito para análise de dados textuais. *Temas em Psicologia*, 21(2), 513-518. doi: 10.9788/TP2013.2-16
- Camargo, B. V., & Justo, A. M. (2018). *Tutorial para uso do software IRaMuTeQ (Interface de R pour les Analyses Multidimensionnelles de Textes et de Questionnaires)*. Santa Florianópolis: Laboratório de Psicologia Social da Comunicação e Cognição - UFSC. Recuperado de <http://iramuteq.org/documentation/fichiers/tutoriel-portugais-22-11-2018>
- Dany, L. (2016). Analyse qualitative du contenu des représentations sociales. In G. Lo Monaco, S. Delouée, & P. Rateau (Orgs.), *Les représentations sociales: Théories, méthodes et applications* (pp. 85p102). Paris: De Boeck Supérieur.
- Dany, L., Urdapilleta, I., & Monaco, G. L. (2015). Free associations and social representations: Some reflections on rank-frequency and importance-frequency methods. *Quality & Quantity*, 49(2), 489-507. doi: 10.1007/s11135-014-0005-z
- Flick, U. (2014). Mapping the Field. In U. Flick (Org.), *The SAGE handbook of qualitative data analysis* (pp. 3-18). Los Angeles: Sage.
- Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2009). *Análise multivariada de dados* (6a ed). Recuperado de <http://site.ebrary.com/id/10795050>
- Justo, A. M. (2016). *Corpo e representações sociais: Sobrepeso, obesidade e práticas de controle de peso* (Tese de Doutorado). Universidade Federal de Santa Catarina, Centro de Filosofia e Ciências Humanas, Programa de Pós-Graduação em Psicologia, Florianópolis. Recuperado de <https://1library.org/document/ye81md7y-corpo-representacoes-sociais-sobrepeso-obesidade-praticas-controle-peso.html>
- Justo, A. M., & Camargo, B. V. (2014). Estudos qualitativos e o uso de softwares para análises lexicais. In C. Novikoff, S. R. M. Santos, & O. B. Mithidieri (Orgs.), *Caderno de artigos: X SIAT & II Serpro (2014: Duque de Caxias, RJ)* (pp. 37–54). Duque de Caxias, RJ: UNIGRANRIO. Recuperado de [http://www.academia.edu/download/37180083/Justo\\_\\_\\_Camargo\\_2014.pdf](http://www.academia.edu/download/37180083/Justo___Camargo_2014.pdf)
- Kalampalikis, N. (2003). L’apport de la méthode Alceste dans l’analyse des représentations sociales. In J.-C. Abric (Org.), *Méthodes d’étude des représentations sociales* (p. 147-163). Ramonville-Saint-Agne: Érès.

- Lafon, P. (1980). Sur la variabilité de la fréquence des formes dans un corpus. *Mots - Les langages du politique*, 1(1), 127-165. doi: 10.3406/mots.1980.1008
- Landivar, D. S., Mathonnat, C., & Tichit, A. (2015). Classification des systèmes de monnaies non-bancaires: Ce que disent les données du Web. *Etudes & documents*, (25), 1-25. Recuperado de <https://hal.uca.fr/halshs-01100849v2>
- Lebart, L., & Salem, A. (1994). *Statistique textuelle*. Paris: Dunod.
- Leblanc, J.-M. (2015). Proposition de protocole pour l'analyse des données textuelles: Pour une démarche expérimentale en lexicométrie. *Nouvelles perspectives en sciences sociales (NPSS)*, 11(1), 25-63. doi: 10.7202/1035932ar
- Loubère, L., & Ratinaud, P. (2014). *Documentation IraMuTeQ - 0.6 alpha 3 version 0.1*. Recuperado de [http://www.iramuteq.org/documentation/fichiers/documentation\\_19\\_02\\_2014.pdf](http://www.iramuteq.org/documentation/fichiers/documentation_19_02_2014.pdf)
- Marchand, P. (2013). Quelques traces chronologiques de l'exploration textométrique. *Bulletin de Méthodologie Sociologique*, 120(1), 38-46. doi: 10.1177/0759106313497856
- 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 11ème Journées internationales d'Analyse statistique des Données Textuelles* (pp. 687-699). Liège, Belgique, 11. Recuperado de <http://lexicometrica.univ-paris3.fr/jadt/jadt2012/Communications/Marchand,%20Pascal%20et%20al.%20-%20L%27analyse%20de%20similitude%20appliquee%20aux%20corpus%20textuels.pdf>
- 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. *Bulletin de Méthodologie Sociologique*, 117(1), 46-60. doi: 10.1177/0759106312465550
- Nascimento, A. R. A., & Menandro, P. R. M. (2006). Análise lexical e análise de conteúdo: Uma proposta de utilização conjugada. *Estudos e Pesquisas em Psicologia*, 6(2), 72-88. Recuperado de <http://www.revispsi.uerj.br/v6n2/artigos/pdf/v6n2a07.pdf>
- Oliveira, D. C., Teixeira, M. C. T. V., Fischer, F. M., & Amaral, M. A. (2003). Estudo das representações sociais através de duas metodologias de análise de dados. *Revista de Enfermagem da UERJ*, 11(3), 317-327. Recuperado de <https://pesquisa.bvsalud.org/portal/resource/pt/lil-364798>
- Pateyron, B., Weber, M., & Germain, P. (2015). Essai d'analyse lexicale et stemma codicum de quatre-vingt-trois rituels de Chevaliers Kadosh de la collation du fonds de l'atelier

- de recherches Sources. *Nouvelles perspectives en sciences sociales*, 11(1), 93-144. doi: 10.7202/1035934ar
- Pélessier, D. (2017). *Initiation à la lexicométrie: Approche pédagogique à partir de l'étude d'un corpus avec le logiciel Iramuteq*. Toulouse: IDETCOM - Université Toulouse.
- Pereira, H. G. (1987). Tratamento informático de questionários: O ponto de vista da análise factorial das correspondências. *Análise Social*, 23(98), 733-746.
- Ratinaud, P. (2014). *IRAMUTEQ: Interface de R pour les Analyses Multidimensionnelles de Textes et de Questionnaires-0.7 alpha 2*. Recuperado de <http://www.iramuteq.org>
- Ratinaud, P., & Marchand, P. (2016). Quelques méthodes pour l'étude des relations entre classifications lexicales de corpus hétérogènes: Application aux débats à l'Assemblée Nationale et aux sites web de partis politiques. *Journées internationales d'Analyse statistique des Données Textuelles*. Université Nice Sophia Antipolis, Nice, France, 13. Recuperado de <http://lexicometrica.univ-paris3.fr/jadt/jadt2016/01-ACTES/83670/83670.pdf>
- Reinert, M. (1983). Une méthode de classification descendante hiérarchique: Application à l'analyse lexicale par contexte. *Cahiers de l'analyse des données*, 8(2), 187-198.
- Reinert, M. (1986). Un logiciel d'analyse lexicale: ALCESTE. *Cahiers de l'analyse des données*, 11(4), 471-484.
- Reinert, M. (1987). Classification Descendante Hierarchique et Analyse Lexicale par Contexte: Application au Corpus des Poesies D'A. Rimbaud. *Bulletin de Méthodologie Sociologique*, 13(1), 53-90.
- Reinert, M. (2001). Alceste, une méthode statistique et sémiotique d'analyse de discours. Application aux Rêveries du promeneur solitaire. *Revue française de psychiatrie et de psychologie médicale*, 5(49), 32-36.
- Reinert, M. (2002). *Alceste: Un logiciel d'aide pour l'analyse de discours: Notice simplifiée (de la version de base commune aux versions 4x)*. Paris: Université de Saint-Quentin-en-Yvelines. Recuperado de <http://www.ling.uqam.ca/forum/satoman/images/AlcesteNoticeSimplifiee.pdf>
- Reinert, M. (2009). Journée d'étude du 21 août 2009 sur la méthodologie «Alceste» Arguments des Interventions. *Bulletin de Méthodologie Sociologique*, 104(1), 39-46. doi: 10.1177/075910630910400105
- Salem, A. (1986). Segments répétés et analyse statistique des données textuelles. *Histoire & Mesure*, 1(2), 5-28.

- Salviati, M. E. (2017). *Manual do Aplicativo Iramuteq (versão 0.7 Alpha 2 e R Versão 3.2.3): Compilação, organização e notas de Maria Elisabeth Salviati*. Planaltina: Autor. Recuperado de <http://www.iramuteq.org/documentation/fichiers/manual-do-aplicativo-iramuteq-par-maria-elisabeth-salviati>
- Santos, V., Salvador, P., Gomes, A., Rodrigues, C., Tavares, F., Alves, K., & Bezerril, M. (2017). IRAMUTEQ nas pesquisas qualitativas brasileiras da área da saúde: Scoping review. *Atas CIAIQ 2017*, 2, 392-401. Recuperado de <https://proceedings.ciaiq.org/index.php/ciaiq2017/article/view/1230>
- Scelles, R. (1997). Apports de logiciels d'analyse de données textuelles (ADT) dans les procédures d'analyse de contenu d'entretiens semi-directifs de recherche: Alceste et hyperbase. *Bulletin de Méthodologie Sociologique*, 57(1), 25-48. doi: 10.1177/075910639705700102
- Vergès, P. (1992). L'évocation de l'argent: Une méthode pour la définition du noyau central d'une représentation. *Bulletin de Psychologie*, (45), 203-209.
- Vergès, P., Scano, S., & Junique, C. (2002). *Ensembles de programmes permettant l'analyse des evocations (Manuel version 2)*. Provence-Alpes-Côte d'Azur: Université d'Aix-en-Provence.

#### **Endereço para correspondência**

**Yuri Sá Oliveira Sousa**

Instituto de Psicologia - IPS/UFBA

Rua Prof. Aristίδes Novis, 197, Salvador - BA, Brasil. CEP 40210-909

Endereço eletrônico: [yurisousas@gmail.com](mailto:yurisousas@gmail.com)

Recebido em: 14/02/2020

Reformulado em: 23/07/2020

Aceito em: 05/08/2020

#### **Notas**

\* Professor Adjunto do Instituto de Psicologia (IPS) da Universidade Federal da Bahia (UFBA).

Este artigo de revista **Estudos e Pesquisas em Psicologia** é licenciado sob uma *Licença Creative Commons Atribuição-Não Comercial 3.0 Não Adaptada*.