



***Software* para análise de dados na pesquisa qualitativa: relato de uso do Iramuteq em estudo baseado em teoria fundamentada em dados**

Software for data analysis in qualitative research: a report on the use of Iramuteq in a grounded theory-based study

Ana Flavia Lucas de Faria Kama

Doutora em Ciência da Informação pela Universidade de Brasília (UnB). Bibliotecária da Universidade de Brasília (UnB).
E-mail: anakama@unb.br

Fernando César Lima Leite

Doutor em Ciência da Informação pela Universidade de Brasília (UnB). Professor da Faculdade de Ciência da Informação da Universidade de Brasília (UnB).
E-mail: fernandoc@unb.br

RESUMO

Este artigo possui como objetivo descrever o uso do *software* Iramuteq na pesquisa qualitativa conduzida por meio da Teoria Fundamentada em Dados, com foco na produção, distribuição e uso do livro acadêmico nas Ciências Sociais e Humanas. Adotou-se uma abordagem qualitativa, com base na Teoria Fundamentada em Dados e na técnica estatística da Classificação Hierárquica Descendente (CHD), aplicada ao *corpus* de entrevistas com pesquisadores bolsistas de produtividade do CNPq nas áreas do Direito e da Sociologia. Os achados evidenciam a eficácia do Iramuteq na análise de dados qualitativos e no suporte à aplicação da Teoria Fundamentada em Dados, especialmente por meio da CHD. As análises revelaram a centralidade persistente do livro na comunicação científica nas CS&H e evidenciaram distinções entre as áreas investigadas quanto às práticas editoriais, como a relevância atribuída a capítulos de livro, livros organizados e coletâneas. O uso do Iramuteq potencializou a análise qualitativa, mas requer domínio técnico e teórico por parte do pesquisador. O estudo reforça a importância do uso crítico e fundamentado de ferramentas computacionais em pesquisas qualitativas na área da Ciência da Informação.

Palavras-chave: Iramuteq, teoria fundamentada em dados, livro, Ciências Sociais e Humanas, comunicação científica.

ABSTRACT

To describe the use of the Iramuteq software in qualitative research conducted through Grounded Theory, focusing on the production, distribution, and use of academic books in the Social Sciences and Humanities (SSH). A qualitative approach was adopted, based on Grounded Theory and the statistical technique of Descending Hierarchical Classification (DHC), applied to a corpus of interviews with CNPq productivity scholarship researchers in the fields of Law and Sociology. The findings highlight the effectiveness of Iramuteq in qualitative data analysis and its support for the application of Grounded Theory, especially through DHC. The analyses showed the continued centrality of the book in scholarly communication within the SSH and revealed differences between the fields studied regarding editorial practices, such as the relevance attributed to book chapters, edited volumes, and collections. The use of Iramuteq enhanced the qualitative analysis but requires both technical and theoretical expertise from the researcher. The study reinforces the importance of a critical and well-grounded use of computational tools in qualitative research in the field of Information Science..

Keywords: Iramuteq, grounded theory, book, Social Sciences and Humanities, scholarly communication.

1 INTRODUÇÃO

O avanço das ferramentas computacionais tem desempenhado um papel fundamental na pesquisa qualitativa, especialmente na análise de dados textuais. No âmbito das Ciências Sociais e Humanas (CS&H), a utilização de *softwares* de apoio à análise qualitativa de dados, conhecidos como CAQDAS (*Computer Aided Qualitative Data Analysis Software*), tem crescido significativamente (Camargo; Justo, 2016; Canuto et al., 2020). Essas ferramentas possibilitam a organização, recuperação e análise sistemática de grandes volumes de dados qualitativos, tornando o processo de investigação mais estruturado e eficiente.

Dentre os diversos *softwares* disponíveis, destaca-se o Iramuteq (*Interface de R pour les Analyses Multidimensionnelles de Textes et de Questionnaires*), um *software* livre baseado no ambiente estatístico R, amplamente utilizado para análise lexicográfica e classificação de dados textuais. Neste artigo, é explorado a aplicação do Iramuteq em uma pesquisa qualitativa na área da Ciência da Informação, especialmente no contexto do emprego do método da Teoria Fundamentada em Dados (TFD), o qual permite o desenvolvimento de teorias emergentes a partir da análise indutiva de dados empíricos.

Diante do desafio do entendimento e descrição do uso dessa ferramenta e da escassez de materiais publicados sobre CAQDAS em pesquisas na área, este artigo teve como objetivo descrever o uso do *software* Iramuteq na pesquisa qualitativa com emprego da TFD. O estudo em cuja análise de dados foi utilizado o Iramuteq intitulou-se "O lugar do livro na produção, distribuição e uso do conhecimento científico nas Ciências Sociais e Humanas", que buscou identificar recomendações, práticas e percepções sobre a produção, distribuição e uso de livros no processo de comunicação científica em disciplinas das Ciências Sociais e Humanas (Kama, 2025).

2 REVISÃO DE LITERATURA

A produção de conhecimento nas CS&H tem sido objeto de discussão no que tange às especificidades epistemológicas e metodológicas. A TFD (Charmaz, 2009) se destaca por seu potencial de gerar teorias a partir da análise sistemática de dados empíricos, enquanto o Iramuteq tem se consolidado como uma ferramenta relevante na classificação lexical e análise textual (Ratinaud, 2025). Estudos recentes apontam que a combinação do método da TFD e o uso de ferramentas de apoio à análise qualitativa de dados potencializa

a identificação de padrões e tendências nos discursos, proporcionando interpretações mais consistentes sobre os fenômenos investigados (Gama; Zaninelli; Santos Neto, 2023; Souza *et al.*, 2018).

A Teoria Fundamentada em Dados (TFD) surgiu na década de 1960 com os trabalhos de Glaser e Strauss (1967), que propuseram uma abordagem qualitativa indutiva para a formulação de teorias a partir da análise sistemática de dados empíricos. A metodologia envolve a codificação aberta, axial e seletiva, permitindo a identificação de categorias emergentes e o desenvolvimento de relações teóricas. Um dos passos para a consecução da TFD é a definição da amostragem teórica. Ela possibilita que o pesquisador colete seus dados levando em consideração as experiências de indivíduos que demonstram conhecimento da situação estudada (Dantas *et al.*, 2009; Gama; Zaninelli; Santos Neto, 2023). Dessa maneira, torna-se possível o desenvolvimento da teoria, que emergirá a partir dos dados obtidos. A amostragem teórica constitui a ideia de que não há uma amostragem pré-determinada, mas sim seu desenvolvimento durante o processo de pesquisa, conforme versam Strauss e Corbin (2008).

Nesse contexto, um dos passos essenciais na aplicação da TFD é a amostragem teórica, que se caracteriza pela flexibilidade e adaptabilidade no processo de coleta de dados. Como apontam Dantas *et al.* (2009) e Gama, Zaninelli e Santos Neto (2023), a amostragem teórica permite que o pesquisador selecione os participantes ou fontes de dados com base em sua relevância para o desenvolvimento da teoria, em vez de seguir um plano amostral rígido e pré-definido. À medida que a análise dos dados avança, o pesquisador ajusta a coleta, direcionando-a para áreas que necessitam de maior exploração teórica.

Essa flexibilidade possibilita o refinamento contínuo da teoria que se constrói a partir dos dados emergentes. A coleta de dados prossegue até que se alcance a saturação teórica, momento em que novos dados deixam de trazer descobertas significativas para o progresso das categorias teóricas, indicando que os dados são suficientes para o desenvolvimento da teoria. Como explicam Strauss e Corbin (2008), a amostragem teórica se desenvolve interativamente ao longo da pesquisa, preenchendo lacunas e ampliando a compreensão dos conceitos emergentes, o que permite que a teoria se forme diretamente a partir dos dados de maneira robusta e dinâmica.

O delineamento da amostra para a pesquisa analisada nesse estudo passou pela definição das áreas de conhecimento representativas das Ciências Sociais e das

Humanidades, ou Ciências Humanas, culminando no Direito e na Sociologia, respectivamente. Para refletir a prática da pesquisa no cotidiano dos participantes, definiu-se como necessário estabelecer critério de seleção desses indivíduos. Para tanto, foi proposto que apenas aqueles com Bolsa Produtividade em Pesquisa (PQ) do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) nas áreas de Sociologia e Direito seriam incluídos na amostra. Além disso, somente pesquisadores credenciados em programas de pós-graduação de universidades federais brasileiras comporiam a amostra. Essa decisão baseia-se no fato de que essas instituições representam a maior parte da produção científica no país (Moura, 2019), sendo seguidas pelas Instituições de Ensino Superior (IES) públicas estaduais, institutos de pesquisa e agências de fomento.

A escolha do Direito e da Sociologia entre as diversas disciplinas que compõem as Ciências Sociais e Humanas (CS&H), justifica-se tanto por critérios epistemológicos e pragmáticos, vinculados à natureza da pesquisa, quanto ao objetivo de compreender como se dá a produção, distribuição e uso do conhecimento científico em formato de livro. Para tanto, optou-se por um recorte didático que contemplasse a diversidade interna das CS&H, elegendo o Direito como representante das Ciências Sociais Aplicadas e a Sociologia como representante das Ciências Humanas. O Direito, além de constituir uma área numericamente expressiva em termos de pesquisadores com produtividade reconhecida (como bolsistas de produtividade do CNPq), é também uma das disciplinas que mais recorre ao livro como principal meio de comunicação científica, devido à complexidade normativa e à necessidade de sistematização doutrinária que caracterizam seu campo (Araújo; Miguel, 2017; Jokić; Mervar; Mateljan, 2019). A Sociologia, por sua vez, possui uma tradição consolidada de valorização do livro como suporte privilegiado de disseminação do conhecimento, especialmente por seu caráter teórico-reflexivo e por operar em um campo multiparadigmático e crítico, típico das Ciências Humanas (Carvalho et al., 2013; Lindholm-Romantschuk; Warner, 1996). Assim, as duas disciplinas foram escolhidas não por exclusão das demais, mas por reunirem características complementares que permitem analisar, de forma contrastiva e representativa, o papel do livro na comunicação científica dentro da pluralidade que constitui as CS&H.

Dessa forma, foram coletados dados a partir de um grupo de 17 pesquisadores de CS&H, das áreas de Sociologia e Direito, que possuíam Bolsa de Produtividade em Pesquisa (PQ) descrita e publicizada no site do CNPq e que também tinham vínculo com alguma universidade federal. Ou seja, cientistas sociais e humanistas de Institutos

Federais e de pesquisa, de IES particulares, de agências de fomento e demais instituições que não sejam universidades federais, não foram considerados. A escolha das disciplinas representativas da amostra trouxe foco para os campos da Sociologia e do Direito, com o objetivo de analisar as percepções de representantes das Ciências Humanas e das Ciências Sociais Aplicadas, respectivamente. De acordo com o parecer consubstanciado nº 7.108.697 do Comitê de Ética em Pesquisa em Ciências Humanas e Sociais da Universidade de Brasília (CEP/CHS/UnB), que autorizou a coleta dos dados da pesquisa mediante solicitação de Aceites Institucionais das universidades envolvidas na amostra, foi necessário solicitar autorização dos gestores responsáveis por essas instituições antes do contato e marcação de entrevista com os pesquisadores com bolsas PQ.

O uso de CAQDAS na pesquisa qualitativa tem sido amplamente debatido na literatura, destacando-se vantagens como a melhor organização dos dados, rapidez na codificação e maior transparência no processo analítico. O Iramuteq desenvolvido por Pierre Ratinaud em 2009, é um *software* livre que opera com base na linguagem R e permite desde análises lexicais básicas até procedimentos estatísticos mais complexos, como classificação hierárquica descendente (CHD), Análise de Similitude e Análise Fatorial de Correspondências (Camargo; Justo, 2016; Canuto et al., 2020). Ao permitir a visualização de ocorrências e associações entre termos, o Iramuteq contribui significativamente para a compreensão dos núcleos de sentido presentes nos discursos analisados, oferecendo subsídios valiosos à codificação e à formação de categorias teóricas.

A integração entre a TFD e o Iramuteq, nesse sentido, não se deu de forma meramente instrumental, mas como estratégia analítica complementar, capaz de potencializar o processo de codificação e de extração de conceitos a partir do material empírico. A utilização do *software* auxiliou na identificação de regularidades linguísticas e padrões discursivos que, posteriormente, foram refinados por meio da codificação axial e seletiva, conforme preconizado pela TFD. Essa articulação metodológica favoreceu a triangulação analítica, ampliando a consistência dos achados e fortalecendo o processo de construção teórica. Dessa forma, o uso do Iramuteq nesta pesquisa não apenas agregou rigor à análise textual, como também reafirmou o compromisso com a sistematização transparente e fundamentada dos dados qualitativos, em consonância com os princípios da TFD.

3 METODOLOGIA

Entre as pretensões do fazer científico, busca-se alcançar teorias, propor ou descobrir explicações viáveis sobre fenômenos naturais e sociais. Portanto, a fim de compreender os modos como cientistas sociais e humanistas produzem, distribuem e usam informação científica e/ou resultantes da construção de conhecimentos, optou-se por um método mais apropriado possível para a análise centrada em indivíduos, elaborada a partir de dados remetidos de uma amostra desse grupo de pessoas e que pudesse, portanto, explicar tal fenômeno particular. A TFD é uma proposta metodológica qualitativa que preconiza o uso de procedimentos para desenvolver de forma indutiva uma teoria derivada dos dados obtidos (Cassiani; Caliri; Pelá, 1996).

Para tanto, a identificação e construção cíclica e lógica de categorias e conceitos foi necessária para a busca de uma resposta indutiva de uma teoria de comunicação científica, por meio da análise qualitativa de dados levantados em entrevistas semiestruturadas e dinâmicas ao longo da coleta com o grupo amostral. Conforme demonstra Strauss e Corbin (2008), todo o percurso da TFD possui como meta a identificação, desenvolvimento e relacionamento de conceitos. Podemos afirmar, portanto, que a teoria que se buscou fundar no estudo em tela não advém de um corpo propriamente pré-existente, embora possa englobar aspectos de outras teorias anteriores, mas sim de uma construção de *corpus* e conceitos que pretendeu agregar aos marcos científicos e teóricos já existentes na Ciência da Informação.

O método de coleta escolhido, a entrevista semiestruturada, busca levantar ciclicamente dados empíricos da amostra descrita, com as vantagens de possibilitar maneiras de motivar e esclarecer o respondente, além de imprimir certa flexibilidade e melhor manuseio sobre a situação e validação dos dados, via análise comportamental não verbal dos participantes (Cassiani; Caliri; Pelá, 1996; Kvale; Brinkmann, 2009). Segundo Gama *et al.* (2023), o uso da entrevista como um instrumento de coleta tem sido o mais evidente dentre as pesquisas na área da Ciência da Informação que utilizam a TFD como abordagem metodológica, ao refletir o pressuposto de que esse método foi criado substancialmente como uma metodologia que visa levantar dados em campo.

Um dos passos para a consecução da TFD é a definição e execução da amostragem teórica. Ela possibilita que o pesquisador colete seus dados levando em consideração as experiências de indivíduos que demonstram conhecimento da situação estudada (Dantas *et al.*, 2009; Gama; Zaninelli; Santos Neto, 2023). Desta maneira, é possível o

desenvolvimento da teoria que poderá elevar-se a partir dos dados obtidos. A amostragem teórica imbui a ideia de que não há uma amostragem pré-determinada e sim seu desenvolvimento durante o processo de pesquisa, conforme versam Strauss e Corbin (2008). Além disso, ela é cumulativa no sentido de trazer o acréscimo das variações de cada evento à coleta e às análises anteriores. Contudo, a fim de definir parâmetros metodológicos e didáticos, além de desenhar uma metodologia para apreciação do CEP/CHS/UnB durante o ano de 2023, foi delineado o seguinte universo e população da pesquisa, explicitado no Quadro 1.

Quadro 1 – Objetivos, universo e população da pesquisa

Objetivo	Universo / população Pesquisadores as áreas das CS&H	
	Humanidades Sociologia	Ciências Sociais Aplicadas Direito
Identificar práticas e percepções de produção, distribuição e uso de livros por cientistas sociais e humanistas ligados a instituições federais de ensino superior públicas brasileiras	Pesquisadores vinculados à programas de pós-graduação universidades federais, que possuam bolsa produtividade (PQ) do CNPq na área da Sociologia	Pesquisadores vinculados à programas de pós-graduação universidades federais, que possuam bolsa produtividade (PQ) do CNPq na área do Direito

Fonte: Elaborado pelos autores (2025)

Após revisão bibliográfica sobre o tema, foi definido um roteiro para a pesquisa semiestruturada com três blocos didaticamente divididos em produção, distribuição e uso do conhecimento científico, resultando em 9 perguntas, as quais foram cuidadosamente elaboradas de forma a permitir um desenvolvimento livre de seus respondentes.

Em seguida, os procedimentos metodológicos e coleta de dados para a consecução deste estudo foram submetidos e aprovados pelo CEP/CHS/UnB, em 2023, por meio do Certificado de Apresentação de Apreciação Ética (CAAE) nº 67697323.8.0000.5540 e parecer nº 7.108.697, em consonância com os normativos do Conselho Nacional de Saúde (CNS): Resolução nº 466, de 12 de dezembro de 2012, que aprova diretrizes e normas regulamentadoras de pesquisas envolvendo seres humanos (CNS, 2013), e Resolução nº 510, de 07 de abril de 2016, que dispõe sobre normas aplicáveis a pesquisas em Ciências Humanas e Sociais (CNS, 2016). Após aprovação do parecer substanciado do CEP/CHS/UnB, o levantamento ocorreu no segundo semestre de 2023, de forma online e com cerca de uma hora cada, com pesquisadores/as de 7 Universidades Federais distribuídas entre as 5 regiões do país.

Tendo em vista os fundamentos descritos por Strauss e Corbin (2008) para a execução da TFD, foram definidas as seguintes fases de codificação dos dados qualitativos levantados, dispostas no Quadro 2.

Quadro 2 - Objetivos universo e população da pesquisa

Codificação	Descrição
Aberta	Processo analítico por meio do qual os conceitos são identificados e suas propriedades e suas dimensões são descobertas nos dados
Axial	Processo que relaciona categorias às suas subcategorias, chamado de “axial” já que ocorre em torno do eixo de uma categoria, ao associar uma categoria ao nível de propriedades e dimensões, destrinchando conceitos
Seletiva	O processo em si de integrar e de refinar a teoria a partir das codificações aberta e axial serem integradas e contextualizadas

Fonte: Strauss & Corbin, 2008 - com adaptações

Importante ressaltar que já em 1967, Glaser e Strauss, em seus primeiros estudos sobre a TFD, definiram desenho semelhante ao descrito no Quadro 2, onde havia uma ideia de teoria, seguida de modelo teórico com descrição detalhada, culminando em uma elaboração teórica nova e baseada em conceitos identificados naquela que se convencionou chamar de amostragem teórica. Toda a base indutiva da TFD baseia-se em uma coleta cíclica de dados, que são identificados a cada nova categorização da amostra (Cassiani; Almeida, 1999; Glaser; Strauss, 1967; Strauss; Corbin, 2008). Desta maneira, para este estudo, percebeu-se a adequação de algumas perguntas do roteiro da entrevista no decorrer das análises e categorizações feitas nos primeiros levantamentos, sempre respeitando o desenho metodológico aprovado pelo CEP/CHS/UnB, tendo em vista que a pesquisa foi realizada com pessoas e que se mostra necessário o adequado rigor ético com a coleta.

A análise dos dados das entrevistas semiestruturadas, foi conduzida com o apoio do *software* Iramuteq, que permitiu a categorização e comparação das respostas. Esse processo de tabulação e codificação dos dados fornece ferramentas analíticas que facilitam o desenvolvimento teórico da coleta (Quadro 2). Um dos principais diferenciais do Iramuteq é sua capacidade de lidar com grandes volumes de dados qualitativos, apresentando-os de forma visual e organizada, como dendogramas que representam classes e subcategorias. Popularizado no Brasil a partir de 2013, o *software* tem sido usado em diversas áreas por sua versatilidade e acesso gratuito, facilitando o uso em pesquisas variadas (Ramos; Lima; Amaral-Rosa, 2019; Reinert, 1990; Souza et al., 2018). O Iramuteq oferece cinco tipos principais de análise: estatísticas textuais clássicas, especificação de

grupos, classificação hierárquica descendente, análises de similitude e nuvens de palavras, tornando-o uma ferramenta valiosa para explorar dados qualitativos. Para este estudo, a principal análise utilizada foi a Classificação Hierárquica Descendente (CHD).

O Iramuteq auxiliou na análise dos dados obtidos com as entrevistas semiestruturadas. O uso da ferramenta para esta pesquisa contou com duas de suas principais análises de segmentos textuais (relações lexicais): a Classificação Hierárquica Descendente (CHD), que permitiu a identificação de categorias e subcategorias como unidades de análise, e a Análise de Similitude, que destacou as palavras mais relevantes do *corpus* textual e suas conexões (Camargo; Justo, 2016; Souza et al., 2018). Dessa forma, o Iramuteq processou os dados qualitativos na fase preliminar e aberta da codificação para que fosse possível analisá-los com maior rigor estatístico e conceitual nas fases axial e seletiva (Camargo; Justo, 2016; Glaser; Strauss, 1967). Portanto, ao considerar que a técnica de análise dos dados escolhida foi a TFD, utilizou-se a CHD no tratamento das entrevistas, com o objetivo de identificar categorizações de primeiro e segundo níveis (categorias e subcategorias).

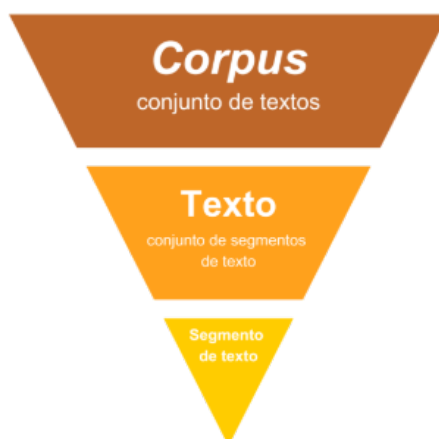
Logo, a CHD organizou os discursos com base em suas semelhanças e diferenças, formando classes, considerando que seu vocabulário se distingue dos segmentos textuais de outras, por meio de vários testes qui-quadrado (χ^2) (Camargo; Justo, 2016). A análise lexicográfica possibilitou a contextualização do vocabulário típico de cada classe, agrupando as percepções dos participantes. A partir desse vocabulário, foi possível descrever cada classe, levando em conta as categorias preliminares acerca da comunicação científica nas CS&H, identificada em Revisão Sistemática da Literatura (RSL) e no referencial teórico da pesquisa.

De forma a fazer um processo comparativo e cumulativo de análises, os dados coletados nas entrevistas foram separados em dois *corpus*: um para os participantes do Direito e outro para os participantes da Sociologia. Um conjunto de unidades de textos compõe um *corpus* de análise e para que ele seja adequado à análise do tipo CHD, deve constituir-se de um conjunto textual centrado em um tema. O material textual deve ser monotemático, pois a análise de textos sobre vários itens previamente estruturados ou diversos temas resulta na reprodução da estruturação prévia dos mesmos (Camargo; Justo, 2016).

Aliada a essa abordagem de uso da ferramenta, optou-se pela divisão da análise em dois *corpus*, seguindo as estratégias de codificação sugeridas por Corbin e Strauss (2008).

Dessa forma, por meio das codificações aberta, axial e seletiva para cada conjunto de pesquisadores, Direito e Sociologia, buscou-se desenvolver uma teoria substantivada mais contextualizada e interpretada. A Figura 1 ilustra os conceitos de *Corpus* (conjunto de textos), *Texto* (conjunto de segmentos de textos) e *Segmento de Texto* utilizados para este estudo, seguindo as definições apresentadas por Camargo e Justo (2016).

Figura 1 - Noções de *Corpus*, *texto* e *segmento de texto*



Fonte: Camargo e Justo, 2016

Descrição da imagem: Infográfico em formato de cone invertido, com três cortes horizontais, tendo fundos em tons laranja e amarelo. No primeiro e maior corte: “*Corpus*: conjunto de textos”; no corte do meio: “*Texto*: conjunto de segmentos de texto”; no último e menor corte: “*Segmento de texto*”.

Logo, foram elaborados dois dendrogramas (Figura 2 e 3), um para cada *corpus*, ilustrando as classes (subcategorias), as divisões que as originaram (categorias) e as relações entre elas (Camargo; Justo, 2016). Dessa forma, foi possível realizar a primeira fase da análise das entrevistas semiestruturadas, por meio da segmentação de texto, suas similitudes entre si e diferenciação em classes (CDH). Após identificação das classes de vocábulos de cada *corpus*, foram demonstradas análises lexicográficas que especificaram o χ^2 (x^2), ou seja, o ponto de corte para que uma palavra fosse significativa para uma classe e sua associação com a classe em si, assim como sua porcentagem de ocorrência nos segmentos de texto de cada classe, em comparação com sua ocorrência no *corpus* geral (Camargo; Justo, 2016; Moura, 2024).

4 RESULTADOS

Os resultados indicam que o Iramuteq se mostrou uma ferramenta eficaz na análise de dados qualitativos, facilitando a identificação de padrões e relações entre conceitos, em especial no método proposto pela TFD. A CHD permitiu a segmentação dos

dados em categorias temáticas, enquanto as análises de similitude evidenciaram conexões relevantes entre os termos utilizados pelos entrevistados. Além disso, a geração de nuvens de palavras contribuiu para a visualização da recorrência de conceitos-chave na comunicação científica na área da Ciência da Informação.

A TFD possui uma fase preliminar de identificação de categorias e dimensões do conhecimento na literatura selecionada sobre o assunto estudado, de forma a possibilitar a caracterizar as categorias preliminares relacionadas à pesquisa. Assim, a primeira parte da pesquisa foi realizada por meio de um Revisão Sistematizada da Literatura (RSL) (Kama; Leite, 2023), a qual levou em consideração como o livro se insere na comunicação científica nas áreas das CS&H, tanto a partir da literatura existente quanto da percepção dos próprios pesquisadores identificados. Para delinear as categorias e dimensões de conhecimento que orientam as codificações da coleta de dados voltada ao objetivo deste estudo, a RSL citada e sua análise gerou um esquema de categorias preliminares acerca da comunicação científica nas CS&H, conforme Quadro 3.

Quadro 3 - Categorias preliminares acerca da comunicação científica nas CS&H

Categorias preliminares
Produção - como acontece
Formas de produção e validação científica: O modelo de produção nas Ciências Sociais e Humanas (CS&H) difere significativamente do das ciências exatas. O livro é o principal veículo de comunicação nessas áreas, pois proporciona o espaço necessário para a profundidade das discussões
Procedimentos epistemológicos: A produção do conhecimento nas CS&H envolve debates e interpretações mais longas, com foco em questões teóricas e metodológicas, o que pode tornar o livro o meio mais adequado
Qualidade e preservação: A revisão por pares e a preservação da integridade do conteúdo são fundamentais para garantir a validade científica das publicações, principalmente em um contexto em que os livros desempenham papel central na disseminação do conhecimento
Produção - individual ou coletiva
Autoria e colaboração: A produção em Ciências Sociais e Humanas pode ser individual ou coletiva, mas é marcada pela contribuição de autores renomados, cujas obras são citadas ao longo de décadas, reforçando a autoridade acadêmica e a continuidade do debate teórico
Produção - veiculação em publicações
Editoras universitárias e canais de comunicação: As editoras universitárias e os novos formatos digitais de veiculação de conhecimento, como plataformas digitais e agregadores, são meios fundamentais para a disseminação do conhecimento nas CS&H
Padronização de metadados e identificação digital: ISBN, DOI e plataformas como Crossref são importantes para assegurar a rastreabilidade e a indexação das publicações em repositórios digitais, facilitando a disseminação global do conhecimento
Distribuição - motivações para publicar
Razões epistemológicas: O principal objetivo da publicação nas CS&H é o de contribuir para o avanço dos debates teóricos, além de alcançar um público especializado e estimular novas interpretações dentro de campos segmentados
Visão crítica sobre o acesso: A publicação de livros, principalmente em acesso aberto, responde ao desafio de aumentar o alcance de produções científicas para além de pequenas audiências privilegiadas

Distribuição - relevância dos canais de publicação
Diversidade dos canais formais e informais: O uso de livros é essencial para a comunicação científica nas CS&H, onde o debate e a circulação de ideias se dão em diferentes níveis de formalidade, incluindo interações face a face e correspondências informais entre pesquisadores
Distribuição - relevância dos locais de publicação
Editoras de prestígio e controle de qualidade: A marca do editor e a reputação das editoras universitárias influenciam diretamente na escolha dos autores, dada a importância de uma curadoria de qualidade para a legitimação das obras
Crerios de publicação e avaliação: Em CS&H, há maior ênfase em avaliações qualitativas de publicações, levando em conta critérios como originalidade e profundidade teórica, em vez de métricas quantitativas comuns nas ciências exatas
Distribuição - influências do meio digital e físico
Integração de meios: A coexistência do digital e do impresso reflete a transição na forma como o conhecimento é distribuído, com o digital ampliando o alcance e o impresso ainda sendo valorizado por sua permanência e tangibilidade
Digitalização de livros e acesso aberto: O avanço da digitalização e o crescimento de iniciativas de acesso aberto têm permitido maior acessibilidade, especialmente para conteúdos de difícil acesso anteriormente (OAPEN, DOAB)
Distribuição - acesso aberto
Crescimento e desafios do acesso aberto: Iniciativas como OAPEN e DOAB facilitam o acesso global a livros acadêmicos, mas o modelo econômico para a sustentabilidade do acesso aberto nas CS&H ainda enfrenta barreiras financeiras e institucionais
Licenças abertas e direitos autorais: Modelos de licenciamento como o <i>Creative Commons</i> são essenciais para a distribuição em acesso aberto, garantindo a disseminação legal e o respeito aos direitos dos autores
Uso - maneiras de buscar informações
Plataformas digitais e agregadores: Bibliotecas digitais e plataformas como OAPEN e DOAB oferecem novos caminhos para a busca e recuperação de informações, facilitando o acesso para pesquisadores de diversas regiões do mundo
Uso - meio de acesso às informações
Uso de bibliotecas digitais e impressas: As bibliotecas universitárias, tanto físicas quanto digitais, continuam sendo fontes primárias de acesso à informação científica nas CS&H, ainda que o digital esteja ganhando mais espaço
Uso - meio de uso das informações
Interpretação e reutilização de conteúdo: O livro como meio de comunicação nas CS&H facilita o uso crítico e reflexivo das informações, diferentemente do modelo cumulativo comum nas ciências exatas
Uso - contribuições dos ambientes digital e físico
Complementaridade dos formatos: O ambiente digital expande a possibilidade de uso imediato e amplo da informação, enquanto o formato impresso ainda preserva o valor de durabilidade e autoridade acadêmica

Fonte: Elaboração própria a partir dos autores citados

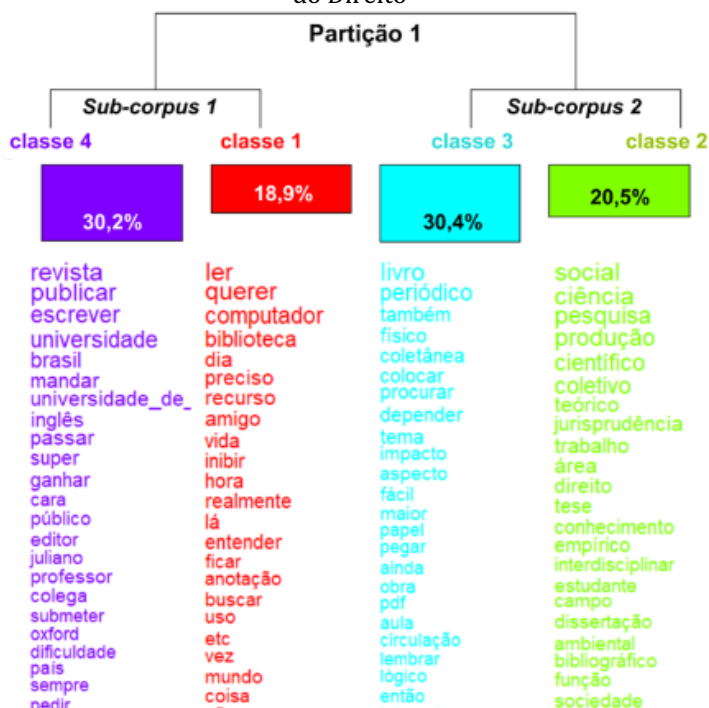
A categorização preliminar de conceitos disposta no Quadro 3 foi utilizada para dar início às análises e categorização aberta da TFD deste estudo, conforme é demonstrado a seguir.

Diante da coleta de dados junto aos pesquisadores, a saturação da amostra foi percebida e estabelecida de formas diferentes dentre os dois universos da população delineada (Quadro 1). Para os pesquisadores da Sociologia, a saturação fez-se na 11ª entrevista e para os do Direito, na 6ª entrevista. As categorizações abertas e axiais

demonstraram nesses pontos de corte que não seria mais necessário a coleta de dados com esses indivíduos, já que os conceitos se repetiam e não mais se complementaríamos substancialmente. A análise dos dados passou pelas fases aberta, axial e seletiva (Quadro 2) e foi elaborada conforme rigor metodológico proposto pela TFD, com auxílio de tabelas dinâmicas de planilhas e com a colaboração de categorizações e análises realizadas por meio do *software* Iramuteq.

A Figura 2 demonstra 100% do corpus textual referente aos dados coletados em campo com pesquisadores do Direito, onde 18,9% dos discursos foram aproveitados na classe 1; 20,5% na classe 2; 30,4% na classe 3 e 30,2% na classe 4. A análise desse *corpus* resultou em 1.968 formas ativas (classes de palavras entre adjetivo, substantivo, advérbio, verbo e formas não reconhecidas) e 850 segmentos de texto.

Figura 2 - Classificação Hierárquica Descendente (CHD) e dendrograma da codificação aberta do *corpus* do Direito



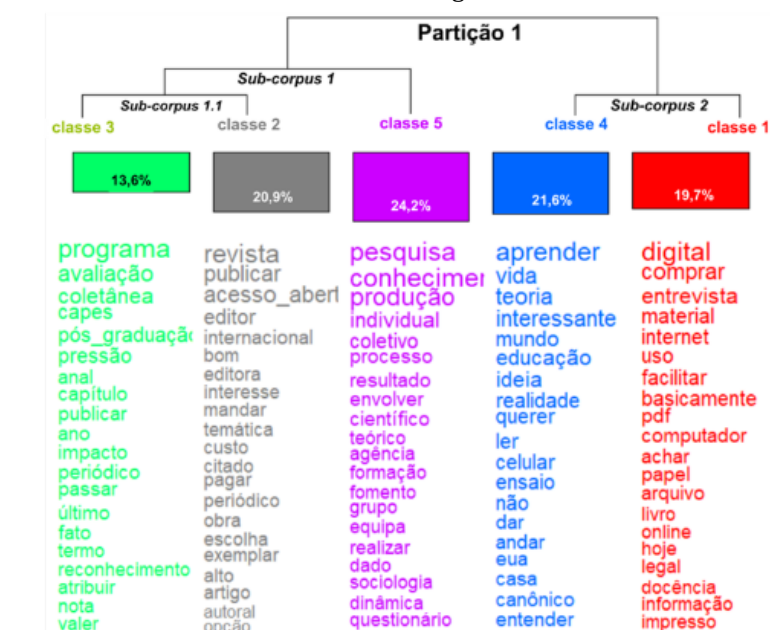
Fonte: Dendrograma gerado pelo Iramuteq a partir do *corpus* do Direito

Descrição da imagem: Dendrograma composto de 4 colunas, com “Partição 1”; seguido do “Sub-corpus 1”, o qual possui duas classes de palavras: “Classe 4” - 30,2%, com um conjunto de palavras organizado em coluna, na cor roxa; e “Classe 1” - 18,9%, com um conjunto de palavras organizado em coluna, na cor vermelha. Logo ao lado vem o “Sub-corpus 2”, composto por duas classes de palavras: “Classe 3” - 30,4%, com um conjunto de palavras organizado em coluna, na cor azul clara; e “Classe 2” - 20,5%, com um conjunto de palavras organizado em coluna, na cor verde clara.

Já a Figura 3 demonstra 100% do *corpus* textual referente aos dados coletados em campo com pesquisadores da Sociologia, a qual resultou em 1.501 segmentos de texto, com 2.930 formas ativas (dentre elas adjetivos, substantivos, advérbios, verbos e formas

não reconhecidas), totalizando 19,7% dos discursos aproveitados na classe 1; 20,9% na classe 2; 13,6% na classe 3, 21,6% na classe 4 e 24,2% na classe 5.

Figura 3 - Classificação Hierárquica Descendente (CHD) e dendrograma da codificação aberta do *corpus* da Sociologia



Fonte: Dendrograma gerado pelo Iramuteq a partir do *corpus* da Sociologia

Descrição da imagem: Dendrograma composto de 5 colunas, com “Partição 1”; seguido do “Sub-corpus 1”, o qual possui o “Sub-corpus 1.1”, composto de duas classes de palavras: “Classe 3” - 13,6%, com um conjunto de palavras organizado em coluna, na cor verde clara; e “Classe 2” - 20,9%, com um conjunto de palavras organizado em coluna, na cor cinza. Logo ao lado, e ainda dentro do “Sub-corpus 1”, apresenta-se a “Classe 5” - 24,2%, com um conjunto de palavras organizado em coluna, na cor lilás. Logo ao lado vem o “Sub-corpus 2”, composto por duas classes de palavras: “Classe 4” - 21,6%, com um conjunto de palavras organizado em coluna, na cor azul escura; e “Classe 1” - 19,7%, com um conjunto de palavras organizado em coluna, na cor vermelha.

Conforme Camargo e Justo (2016), para que a CHD seja eficaz na classificação de dados textuais e qualitativos, é preciso reter ao menos 75% dos segmentos de texto. No caso do *corpus* dos pesquisadores de Direito e da Sociologia, as retenções foram respectivamente de 83,65% e 88,21%, superando o mínimo necessário, o que confirma a viabilidade da análise com o uso do Iramuteq e do pacote estatístico R integrado ao *software*.

Os vocábulos gerados pela CHD que representam todas as classes das repartições nos dendrogramas dos respondentes do Direito (Figura 2) e da Sociologia (Figura 3), escolhidos pelos autores de acordo com as dimensões da literatura da área (Quadro 3) e objetivo da pesquisa (Quadro 1), estão representados nos Quadros 4 e 5 e contribuem para a primeira fase e início de codificações da TFD.

A análise lexicográfica dos vocábulos pode ser observada na aba “perfis” do Iramuteq dentro da CHD gerada a partir dos *corpus* do Direito e da Sociologia, que, dentre outras informações, demonstram a *pourcentage*, ou seja, a proporção de ocorrência da palavra nos segmentos de texto de cada classe em comparação com sua ocorrência no *corpus* completo. Também indica o χ^2 de cada vocábulo, o χ^2 de associação da palavra com a classe, demonstrando o ponto de corte para que uma palavra seja importante para aquela classe ($\chi^2 > 3,84$) (Camargo; Justo, 2016). Esses vocábulos evidenciam as análises e temas trabalhados em cada classe para que a codificação aberta da metodologia escolhida pudesse ser iniciada.

Quadro 4 - Repartições e classes oriundas da CHD do *corpus* do Direito - codificação aberta

Repartição 01: Corpus Direito					
Dimensão e categoria preliminar	Sub-corpus	Classe	Análise lexicográfica		
			Vocabulo	x ²	%
Uso: maneiras de buscar informações Uso: meio de acesso às informações Uso: meio de uso das informações Uso: contribuições dos ambientes digital e físico	1: Uso e distribuição	1	ler	45,6	60,53
			querer	44,72	49,25
			computador	33,57	90
			biblioteca	29,96	52,63
Distribuição: motivações para publicar Distribuição: relevância dos canais de publicação Distribuição: relevância dos locais de publicação Distribuição: influências do meio digital e físico Distribuição: acesso aberto		4	revista	50,75	60,4
			publicar	45,08	57,27
			escrever	40,08	64,62
			universidade	39,53	82,76
Produção: como acontece Produção: individual ou coletiva Produção: veiculação em publicações	2: Produção e distribuição	3	livro	54,09	51,31
			periódico	39,33	90,91
			físico	26,75	84,21
			coletânea	26,36	92,86
Distribuição: motivações para publicar Distribuição: relevância dos canais de publicação Distribuição: relevância dos locais de publicação Distribuição: influências do meio digital e físico Distribuição: acesso aberto		2	social	61,82	94,44
			ciência	61,82	94,44
			pesquisa	54,55	55,22
			produção	54,13	81,88

Fonte: Elaboração própria, com base na análise qualitativa feita pelo Iramuteq

O Iramuteq possibilitou a continuidade do processo de codificações do *corpus* do Direito dentro do contexto das codificações axial e seletiva, a partir da codificação aberta descrita no Quadro 4, facilitada pelo *software*. A partir disso, chegou-se à Teoria Substantiva pertinente às práticas e percepções relacionadas à produção, distribuição e

ao uso de livros no Direito, resumida a seguir. A categoria central identificada é a interação entre práticas e percepções contemporâneas e tradicionais sobre o papel do livro, em torno da qual se organizam as demais categorias emergentes.

1. Uso da Informação e Acesso: pesquisadores do Direito recorrem tanto ao ambiente digital quanto ao físico para buscar e utilizar informações. Termos como "ler", "querer", "computador" e "biblioteca" refletem essa dinâmica. O computador permite velocidade e acesso amplo, enquanto a biblioteca física mantém o valor da leitura reflexiva e da imersão. A convivência entre esses espaços mostra que o uso da informação é híbrido, adaptando-se aos objetivos e necessidades do pesquisador.

2. Produção Científica: Teoria e Autoria: a produção teórica é fortemente valorizada, sendo o livro o principal suporte para o aprofundamento conceitual e analítico. A produção individual ainda predomina, refletindo a tradição jurídica e a importância da autoria singular na construção doutrinária. No entanto, há um crescimento da produção coletiva, influenciada por políticas de fomento (Fundação Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - CAPES, CNPq), pelo movimento interdisciplinar e pelas exigências de editais e projetos.

3. Distribuição do Conhecimento: a distribuição híbrida do conhecimento combina o alcance imediato das publicações digitais com o prestígio e a durabilidade do formato impresso. O livro permanece um veículo central de autoridade acadêmica, mesmo com o crescimento dos artigos científicos e da pressão por visibilidade em periódicos internacionais. O papel das editoras renomadas e o prestígio simbólico do livro impresso são fundamentais para a legitimação da produção na área.

4. Motivação e Relevância das Publicações: a publicação no Direito tem motivações epistemológicas e de reconhecimento acadêmico. A originalidade e a profundidade teórica são mais valorizadas que métricas quantitativas. Livros e periódicos cumprem funções distintas, mas complementares: o livro aprofunda, enquanto o periódico atualiza e circula mais rapidamente. A escolha dos canais de publicação é influenciada tanto pela reputação das editoras e revistas, quanto pela busca por impacto e circulação.

5. Acesso Aberto e Mercado Editorial: o acesso aberto é visto como um vetor de democratização do conhecimento, embora enfrente desafios de sustentabilidade. Modelos como o *Creative Commons* e revistas de código aberto são estratégias reconhecidas para ampliar a circulação. Ainda assim, o mercado editorial e a lógica de

prestígio das editoras continuam influenciando as decisões dos pesquisadores, mostrando uma interdependência entre ciência e mercado.

6. Teoria Substantiva: a teoria construída revela que os pesquisadores do Direito operam em um campo híbrido, navegando entre: o modelo tradicional, centrado no livro impresso, na autoria individual e na produção doutrinária; e o modelo contemporâneo, que valoriza a colaboração, o digital, o acesso aberto e a visibilidade rápida. Essa flexibilidade é necessária para atender às pressões institucionais, aos novos formatos editoriais e às transformações tecnológicas, sem romper com a centralidade do livro como suporte para o pensamento jurídico profundo e estruturado.

Quadro 5 - Repartições e classes oriundas da CHD do *corpus* da Sociologia - codificação aberta

Repartição 01: Corpus Sociologia					
Dimensão e categoria preliminar	Sub-corpus	Classe	Análise lexicográfica		
			Vocabulo	x ²	%
Produção: como acontece Produção: individual ou coletiva Produção: veiculação em publicações	1: Produção, distribuição e uso	3	programa	133,08	88,89
			avaliação	85,82	74,07
			coletânea	65,24	52
			Capes	64,21	68
Distribuição: motivações para publicar Distribuição: relevância dos canais de publicação Distribuição: relevância dos locais de publicação Distribuição: influências do meio digital e físico Distribuição: acesso aberto		2	revista	146,72	59,86
			publicar	100,51	45,06
			acesso aberto	87,84	85,29
			editora	65,36	84,62
Uso: maneiras de buscar informações Uso: meio de acesso às informações Uso: meio de uso das informações Uso: contribuições dos ambientes digital e físico		5	pesquisa	113,74	55,03
			conhecimento	109,52	76,81
			produção	99,1	71,43
			individual	58,02	81,25
Produção: como acontece Produção: individual ou coletiva Produção: veiculação em publicações	2: Produção e uso	4	aprender	55	83,33
			teoria	40,94	77,27
			educação	34	77,78
			ler	29,5	46,67
Uso: maneiras de buscar informações Uso: meio de acesso às informações Uso: meio de uso das informações Uso: contribuições dos ambientes digital e físico		1	digital	105,64	64,94
			comprar	70,44	82,14
			internet	47,02	87,5
			uso	46,66	63,16

Fonte: Elaboração própria, com base na análise qualitativa feita pelo Iramuteq

Da mesma maneira para o *corpus* da Sociologia o Iramuteq possibilitou a continuidade do processo de codificações dentro do contexto das codificações axial e

seletiva, a partir da codificação aberta descrita no Quadro 5, facilitada pelo *software*. Na área da Sociologia, a produção, distribuição e uso do livro refletem tensões e equilíbrios entre práticas tradicionais e exigências contemporâneas. O livro, seja impresso ou digital, é compreendido como elo entre temporalidades, articulando o legado teórico clássico à necessidade de inovação metodológica, institucional e tecnológica.

1. Produção: entre autonomia intelectual e coletivização institucional: a produção na Sociologia se organiza em torno de um eixo articulador: o compromisso com a investigação empírica e com os problemas sociais contemporâneos. A pesquisa pode ser individual e autoral, conforme relatos que destacam a centralidade da autonomia intelectual, mas também se estrutura em redes colaborativas, impulsionadas por editais, grupos de pesquisa e exigências de fomento. A CAPES e o CNPq influenciam essas dinâmicas por meio de critérios avaliativos que, apesar de valorizarem periódicos qualificados, nem sempre contemplam adequadamente livros e coletâneas, gerando tensões no reconhecimento dessas formas de produção.

2. Distribuição: tensões entre prestígio acadêmico e capilaridade digital: as formas de distribuição do conhecimento sofrem impactos diretos da lógica produtivista, das métricas de impacto e das condições de financiamento. O artigo científico, por seu custo reduzido, velocidade de circulação e impacto em *rankings*, torna-se o principal canal de disseminação, especialmente em contextos de escassez de recursos. No entanto, o livro segue ocupando um lugar de prestígio simbólico e epistemológico, sobretudo pela profundidade que proporciona aos debates teóricos. A reputação das editoras, especialmente as universitárias, é um critério relevante, em oposição às editoras comerciais que operam com lógica de mercado e práticas predatórias.

3. Uso: práticas híbridas e transformação digital: o uso do livro na Sociologia contemporânea transita entre os formatos físico e digital. O meio digital favorece o acesso, a reutilização crítica e a democratização do conhecimento, sobretudo em tempos de pandemia, quando a digitalização foi intensificada. Ainda assim, o impresso é valorizado por seu simbolismo acadêmico e sua presença nas tradições disciplinares. Os livros são utilizados tanto para pesquisa quanto para ensino, embora o segundo tenha se tornado mais recorrente na aquisição recente de materiais.

4. Formação e aprendizagem: intersubjetividade e confronto de ideias: a leitura e a aprendizagem são compreendidas como processos coletivos, marcados por trocas, interdisciplinaridade e confrontos teóricos. A leitura de livros clássicos permanece como

base obrigatória da formação, mas as ideias novas emergem da interação entre diferentes campos e sujeitos. O livro, nesse sentido, é uma mediação entre o saber acumulado e as novas formulações críticas, contribuindo para a formação de uma sociologia reflexiva e ética.

5. Ambivalência linguística e internacionalização: a questão do idioma representa uma tensão entre a valorização da produção nacional e a necessidade de internacionalização. A escolha entre publicar em português ou em línguas estrangeiras está relacionada tanto às exigências institucionais quanto ao desejo de manter a especificidade cultural e teórica da produção brasileira.

6. Acesso aberto, plataformas digitais e bibliotecas: as práticas informacionais são atravessadas pela presença de plataformas digitais, bases abertas e repositórios institucionais. O acesso aberto é defendido como um compromisso com a sociedade e com a ciência pública. Teses, dissertações, ensaios e outros formatos circulam amplamente em meio digital, substituindo parcialmente o papel tradicional das bibliotecas físicas. A migração para bibliotecas digitais particulares e o uso intensivo de repositórios exemplificam esse deslocamento.

A análise da Sociologia revela uma teoria substantiva em que o livro funciona como um elo entre temporalidades, articulando a tradição intelectual da disciplina com as transformações contemporâneas da ciência. Ele atua como mediador entre: o pensamento clássico e as demandas atuais de inovação; a produção autoral e a coletivização científica; o formato impresso e o digital; o reconhecimento simbólico e os critérios técnico-burocráticos de avaliação; o rigor acadêmico e o impacto social da ciência.

Nesse cenário, o livro resiste como um instrumento de profundidade e reflexão, mesmo diante da pressão por dinamicidade, capilaridade e produtividade imposta pelos artigos e pelas métricas de impacto. Ele se adapta às transformações digitais e institucionais, mantendo-se como recurso valioso na formação, pesquisa e circulação do conhecimento sociológico.

5 DISCUSSÃO

A utilização do Iramuteq nesta pesquisa foi decisiva para a organização, categorização e interpretação dos dados qualitativos, coletados por meio de entrevistas com cientistas sociais e humanistas atuantes em instituições federais de ensino superior brasileiras. Mais do que um *software* de apoio técnico, o Iramuteq revelou-se uma

ferramenta metodológica com potencial epistemológico, ao viabilizar uma análise aprofundada das narrativas, alinhada aos pressupostos da TFD e da análise de conteúdo. Neste estudo, optou-se pela CHD, uma das formas de processamento de dados oferecidas pelo Iramuteq, o que permitiu a identificação de categorias lexicais e a construção de classes temáticas a partir das regularidades linguísticas do *corpus* textual.

Ao organizar o material empírico com base na frequência e ocorrência de palavras, a CDH favoreceu a emergência de categorias que se mostraram coerentes com as dimensões teóricas da pesquisa e com os objetivos propostos. Essa técnica permitiu a leitura sistemática de padrões discursivos compartilhados entre os participantes, revelando percepções, práticas e tensões relacionadas à produção, distribuição e uso da informação científica nas CS&H. Assim, o Iramuteq, operando como uma ponte entre o dado bruto e a formulação teórica, foi essencial para a elaboração das teorias substantivas delineadas nas disciplinas do Direito e da Sociologia.

A abordagem proposta por Shannon e Weaver em 1949, relida por Berlo (2003), já apontava para a necessidade de modelos comunicacionais específicos, adaptados aos distintos domínios científicos. A operacionalização do Iramuteq neste estudo contribuiu para a identificação dessas especificidades no contexto das CS&H, especialmente ao evidenciar a centralidade do livro como veículo de comunicação científica e as tensões entre tradição e inovação no ecossistema informacional dessas áreas. Nesse sentido, os achados desta pesquisa corroboram e expandem os modelos teóricos de Garvey e Griffith (1967), Hurd (2000) e Costa (2000), atualizando-os à luz dos desafios impostos pela digitalização da ciência, pelas políticas de avaliação acadêmica e pelas novas dinâmicas de acesso à informação.

Os dados extraídos com o auxílio do Iramuteq também reforçam a importância de considerar os aspectos socioculturais e institucionais que atravessam a comunicação científica. A relação entre os pesquisadores e as agências de fomento, a valorização da publicação de artigos em detrimento dos livros, a pressão por internacionalização e as barreiras linguísticas foram recorrentes nas falas analisadas. Esses elementos, embora muitas vezes implícitos, foram sistematizados com precisão pelo *software*, o que demonstra sua capacidade de apoiar leituras teóricas mais densas e comprometidas com os contextos de produção do conhecimento.

Outro aspecto relevante diz respeito à articulação entre a análise assistida por *software* e o trabalho interpretativo do pesquisador. Embora o Iramuteq ofereça

visualizações e segmentações iniciais do *corpus*, sua maior potência reside na capacidade de subsidiar reflexões teóricas embasadas. A CDH não se limita à frequência de termos, mas orienta a interpretação dos sentidos atribuídos pelos sujeitos à sua prática científica. Alinhado ao paradigma qualitativo, tal uso reforça a dimensão hermenêutica da pesquisa em Ciência da Informação e aponta caminhos promissores para o uso crítico e criativo de ferramentas computacionais no campo.

Por fim, é importante destacar que o uso do Iramuteq neste estudo contribuiu para uma leitura epistemologicamente comprometida com a complexidade das CS&H. A estruturação dos dados empíricos possibilitou o diálogo com a fundamentação teórica e com os resultados da RSL, fortalecendo a coerência interna da pesquisa e ampliando sua capacidade de contribuição para a compreensão dos fluxos comunicacionais e das práticas informacionais dos cientistas sociais e humanistas. Assim, o uso do Iramuteq não foi apenas um recurso instrumental, mas um componente estratégico do percurso metodológico, com implicações diretas na construção de conhecimento no campo da comunicação científica.

6 CONCLUSÃO

Este artigo teve como objetivo descrever o uso do Iramuteq e uma de suas formas de processamento de dados qualitativos, destacando sua integração com a TFD no contexto da investigação sobre a comunicação científica nas CS&H. A experiência metodológica relatada demonstrou que o *software* se configura como uma ferramenta robusta para auxiliar na organização, sistematização e visualização de grandes volumes de dados textuais, contribuindo para análises mais aprofundadas e fundamentadas teoricamente.

A utilização do Iramuteq, sobretudo por meio da CHD, mostrou-se eficiente para identificar categorias emergentes nos relatos dos pesquisadores das disciplinas do Direito e da Sociologia, permitindo a formulação de teorias substantivas que evidenciam práticas, desafios e percepções relacionados à produção, distribuição e uso do livro acadêmico nessas áreas. Ao associar as potencialidades computacionais do *software* a uma abordagem qualitativa rigorosa, foi possível não apenas enriquecer o processo analítico, mas também fortalecer a reflexividade do pesquisador diante dos dados.

Contudo, é importante reconhecer as limitações inerentes do uso do Iramuteq nesta pesquisa. Embora a ferramenta ofereça múltiplas funcionalidades para o

tratamento e análise de dados textuais, optou-se por um recorte metodológico específico, centrado em uma de suas formas de análise, o que implicou na não exploração de outros módulos disponíveis que poderiam revelar aspectos complementares dos dados. Essa delimitação metodológica visou manter o foco no objetivo geral do estudo e evitar a extrapolação dos resultados, mas também restringiu o aprofundamento de outras possibilidades analíticas. Além disso, a utilização do *software* exige do pesquisador não apenas domínio técnico, mas também discernimento teórico e metodológico para interpretar criticamente os resultados gerados - um aspecto que, embora valorizado, pode representar um desafio para quem está em fase inicial de familiarização com a ferramenta.

Além de explorar aspectos técnicos e epistemológicos do uso do Iramuteq, este estudo contribui para a divulgação de sua aplicabilidade em pesquisas qualitativas na área da Ciência da Informação. Ainda que o *software* seja reconhecido por sua aplicação na análise de similitude e na nuvem de palavras, seu uso em abordagens mais interpretativas, como a TFD, ainda carece de maior disseminação na literatura nacional. Descrever esse uso em detalhe, como feito neste trabalho, é um passo importante para qualificar a prática investigativa e ampliar o repertório metodológico da comunidade científica.

Ressalta-se, contudo, que o domínio da ferramenta exige não apenas familiaridade com seu funcionamento técnico, mas também profundidade teórica e sensibilidade analítica. O pesquisador continua sendo o principal agente interpretativo do processo investigativo, e sua atuação crítica é imprescindível para que os dados processados pelo *software* resultem em conhecimento relevante e socialmente comprometido.

Diante dos resultados alcançados, recomenda-se que futuras pesquisas explorem outras formas de análise disponíveis no Iramuteq, assim como comparações entre diferentes ferramentas de apoio à análise qualitativa, com vistas a fortalecer o diálogo entre métodos digitais e epistemologias críticas. Tais investigações poderão contribuir para consolidar um ecossistema de pesquisa mais plural, acessível e alinhado às especificidades das CS&H.

REFERÊNCIAS

ARAÚJO, Paula Carina; MIGUEL, Sandra. Motivações dos discentes do Programa de Pós-Graduação em Direito da Universidade Federal do Paraná (UFPR) para publicar em periódicos científicos no domínio do direito. *Perspectivas em Ciência da Informação*, Belo Horizonte, v. 22, n. 1, p. 38-56, 2017. Disponível em: <http://dx.doi.org/10.1590/1981-5344/2710>. Acesso em: 28 dez. 2024.

BERLO, D. K. *O processo de comunicação: introdução à teoria e à prática*. 10. ed. São Paulo: Martins Fontes, 2003.

BRASIL. CONSELHO NACIONAL DE SAÚDE (CNS). Resolução no 466, de 12 de dezembro de 2012. Aprova diretrizes e normas regulamentadoras de pesquisas envolvendo seres humanos. Brasília, 13 jun. 2013. Disponível em: <https://www.gov.br/conselho-nacional-de-saude/pt-br/aceso-a-informacao/legislacao/resolucoes/2012/resolucao-no-466.pdf>. Acesso em: 13 out. 2024.

BRASIL. CONSELHO NACIONAL DE SAÚDE (CNS). Resolução no 510, de 07 de abril de 2016. Dispõe sobre normas aplicáveis a pesquisas em Ciências Humanas e Sociais. Brasília, 24 maio 2016. Disponível em: <https://www.gov.br/conselho-nacional-de-saude/pt-br/aceso-a-informacao/legislacao/resolucoes/2016/resolucao-no-510.pdf>. Acesso em: 13 out. 2024.

CAMARGO, Brígido Vizeu; JUSTO, Ana Maria. *Tutorial para uso do software IRaMuTeQ (Interface de R pour les Analyses Multidimensionnelles de Textes et de Questionnaires)*. Florianópolis: [s. n.], 2016. Disponível em: <http://www.iramuteq.org/documentation/fichiers/tutoriel-en-portugais>. Acesso em: 24 set. 2024.

CANUTO, Angela *et al.* Critic aspects of CAQDAS usage in qualitative research: an empiric comparison of Alceste and Iramuteq as digital tools. *New Trends in Qualitative Research*, [s. l.], v. 3, p. 199–211, 2020. Disponível em: <https://doi.org/10.36367/ntqr.3.2020.199-211>. Acesso em: 24 set. 2024.

CARVALHO, Kátia *et al.* Aspectos gerenciais da política científica brasileira: um olhar sobre a produção científica do campo da Sociologia face aos critérios de avaliação do CNPq e da CAPES. *Em Questão*, Porto Alegre, v. 19, n. 1, p. 187–212, 2013. Disponível em: <https://seer.ufrgs.br/index.php/EmQuestao/article/view/36908>. Acesso em: 28 dez. 2024.

CASSIANI, Silvia Helena De Bortoli; ALMEIDA, Ana Maria de Almeida. Teoria fundamentada nos dados: a coleta e análise de dados qualitativos. *Cogitare Enferm.*, Curitiba, v. 4, n. 2, p. 13–21, 1999. Disponível em: <https://revistas.ufpr.br/cogitare/article/view/44840/27269>. Acesso em: 14 jun. 2024.

CASSIANI, Silvia Helena de Bortoli; CALIRI, Maria Helena Larcher; PELÁ, Nilza Teresa Rotter. A teoria fundamentada nos dados como abordagem da pesquisa interpretativa. *Rev. Latino-Am. Enfermagem*, Ribeirão Preto, v. 4, n. 3, p. 75–88, 1996. Disponível em: <https://doi.org/10.1590/S0104-11691996000300007>. Acesso em: 14 jun. 2024.

CHARMAZ, K. *A construção da teoria fundamentada: guia prático para análise qualitativa*. Porto Alegre: Artmed, 2009.

COSTA, Sely Maria de Souza. Changes in the information dissemination process within the scholarly world: the impact of electronic publishing on scholarly communities of academic social scientists. In: *Anais*, 2000, Rússia. ELPUB CONFERENCE ON ELECTRONIC PUBLISHING, 4. Rússia: ICC Press, 2000. Disponível em: <https://repositorio.unb.br/handle/10482/1063>. Acesso em: 14 jun. 2021.

DANTAS, Claudia de Carvalho *et al.* Teoria fundamentada nos dados - aspectos conceituais e operacionais: metodologia possível de ser aplicada na pesquisa em enfermagem. *Rev Latino-am Enfermagem*, Ribeirão Preto, v. 17, n. 4, 2009. Disponível em: <https://doi.org/10.1590/S0104-11692009000400021>. Acesso em: 16 jun. 2024.

GAMA, Merabe Carvalho Ferreira da; ZANINELLI, Thais Batista Zaninelli; SANTOS NETO, João Arlindo dos. A fundamentação teórico-metodológica da grounded theory nas pesquisas da ciência da informação. In: *Anais*, 2023, Aracaju. (ANCIB, Org.) 23o Encontro Nacional de Pesquisa e Pós-graduação em Ciência da Informação. Aracaju: ANCIB, 2023. Disponível em: <https://www.ancib.org.br/enancib/index.php/enancib/xxxiiienancib/paper/viewFile/1552/1008>. Acesso em: 14 jun. 2024.

GARVEY, W D; GRIFFITH, B C. Scientific communication as a social system: the exchange of information on research evolves predictably and can be experimentally modified. *Science*, [s. l.], v. 157, n. 3792, p. 1011–1016, 1967. Disponível em: <https://doi.org/10.1126/science.157.3792.1011>. Acesso em: 3 fev. 2021.

GLASER, Barney; STRAUSS, Anselm. *The discovery of grounded theory: strategies for qualitative research*. Nova Iorque: Aldine Transaction, 1967.

HURD, Julie M. The transformation of scientific communication: a model for 2020. *Journal of the American Society for Information Science and Technology*, [s. l.], v. 51, n. 14, p. 1279–1283, 2000. Disponível em: [https://doi.org/10.1002/1097-4571\(2000\)9999:9999%3C::AID-ASI1044%3E3.0.CO;2-1](https://doi.org/10.1002/1097-4571(2000)9999:9999%3C::AID-ASI1044%3E3.0.CO;2-1). Acesso em: 2 fev. 2021.

JOKIĆ, Maja; MERVAR, Andrea; MATELJAN, Stjepan. Comparative analysis of book citations in social science journals by Central and Eastern European authors. *Scientometrics*, [s. l.], v. 120, n. 3, p. 1005–1029, 2019. Disponível em: <https://doi.org/10.1007/s11192-019-03176-y>. Acesso em: 28 dez. 2024.

KAMA, Ana Flávia Lucas de Faria. *O lugar do livro na produção, distribuição e uso do conhecimento científico das Ciências Sociais e Humanas*. 2025. Tese - Universidade de Brasília, Brasília, 2025. Disponível em: <https://repositorio.unb.br/handle/10482/52108>. Acesso em: 11 maio 2025.

KAMA, Ana Flavia Lucas Faria; LEITE, Fernando Cesar Lima. Produção, distribuição e uso de livros digitais de acesso aberto nas ciências sociais e humanas: uma revisão sistematizada da literatura. *RDBCI: Rev Digit Bibl e Cienc Inf*, Campinas, v. 21, 2023. Disponível em: <https://doi.org/10.20396/rdbci.v21i00.8674715>. Acesso em: 13 nov. 2024.

KVALE, S.; BRINKAMNN, S. *Interviews: learning the craft of qualitative research interviewing*. Thousand Oaks (CA): Sage, 2009.

LINDHOLM-ROMANTSCHUK, Ylva; WARNER, Julian. The role of monographs in scholarly communication: an empirical study of Philosophy, Sociology and Economics. *Journal of Documentation*, [s. l.], v. 52, n. 4, p. 389–404, 1996. Disponível em: <https://doi.org/10.1108/eb026972>. Acesso em: 28 dez. 2024.

MOURA, Erick de Freitas. *Diferentes configurações de serviços de telessaúde voltadas à promoção do acesso a saúde especializada em áreas vulneráveis: análise do modelo de negócio de uma startup social no Brasil*. 2024. Dissertação - Unicamp, Limeira, 2024. Disponível em: <https://hdl.handle.net/20.500.12733/17038>. Acesso em: 24 set. 2024.

MOURA, Mariluce. Universidades públicas respondem por mais de 95% da produção científica do Brasil. [S. l.], 2019. Disponível em: <http://www.abc.org.br/2019/04/15/universidades-publicas-respodem-por-mais-de-95-da-producao-cientifica-do-brasil/>. Acesso em: 10 out. 2024.

RAMOS, Maurivan Güntzel; LIMA, Valderez Marina do Rosário; AMARAL-ROSA, Marcelo Prado. IRAMUTEQ software and discursive textual analysis: interpretive possibilities. In: COSTA, António Pedro; REIS, Luís Paulo; MOREIRA, António (org.). *Computer supported qualitative research: new trends on qualitative research*. Switzerland: Springer, 2019. p. 58–72. Disponível em: <https://link.springer.com/book/10.1007/978-3-030-01406-3>. Acesso em: 24 set. 2024.

RATINAUD, Pierre. IRAMUTEQ: *Interface de R pour les analyses multidimensionnelles de textes et de questionnaires (computer software)*. [S. l.: s. n.], 2025. Disponível em: <http://www.iramuteq.org/>. Acesso em: 6 abr. 2025.

REINERT, Max. Alceste une méthodologie d'analyse des données textuelles et une application: Aurelia De Gerard De Nerval. *Bulletin de Méthodologie Sociologique*, [s. l.], v. 26, n. 1, p. 24–54, 1990. Disponível em: <https://doi-org.ez54.periodicos.capes.gov.br/10.1177/075910639002600103>. Acesso em: 8 out. 2024.

SOUZA, Marli Aparecida Rocha *et al.* O uso do software IRAMUTEQ na análise de dados em pesquisas qualitativas. *Revista da Escola de Enfermagem, São Paulo*, v. 52, p. e03353, 2018. Disponível em: <http://dx.doi.org/10.1590/S1980-220X2017015003353>. Acesso em: 24 set. 2024.

STRAUSS, Anselm; CORBIN, Juliet. *Pesquisa qualitativa: técnicas e procedimentos para o desenvolvimento de teoria fundamentada*. 2. ed. Porto Alegre: Artmed, 2008.

Recebido em: 12 de maio de 2025
Aprovado em: 27 de dezembro de 2025
Publicado em: 27 de dezembro de 2025