

DATACON: UMA PROPOSTA METODOLÓGICA PARA REALIZAÇÃO DE PESQUISAS NA ÁREA EDUCACIONAL

DATACON: A METHODOLOGICAL PROPOSAL FOR RESEARCH IN EDUCATION

DATACON: UNA PROPUESTA METODOLÓGICA PARA LA INVESTIGACIÓN EN EDUCACIÓN

LICENÇA CC BY:

Artigo distribuído sob os termos Creative Commons, permite uso e distribuição irrestrita em qualquer meio desde que o autor credite a fonte original.

Instituto Federal do Amazonas

Iandra Maria Weirich da Silva

Artigo recebido em: 09/09/2022 Aprovado em: 06/06/2023



Resumo: Este artigo evidencia uma proposta metodológica denominada Datacon, criada para auxiliar nas pesquisas da área educacional. Foi desenvolvida à luz dos pressupostos da Linguística de Corpus e da Abordagem Orientada por Dados. Inclui-se a avaliação dessa proposta por estudantes de pósgraduação, em que consideram as principais contribuições e limitações. Os resultados demonstram contribuições relacionadas ao potencial tecnológico e estratégico da metodologia DataCon para realização da busca, compilação e gerenciamento dos dados e limitações referentes às análises, gestão do tempo e habilidades digitais.

Palavras-chave: Educação. Ensino. Metodologia. Pesquisa.

Abstract: This paper presents a methodological proposal called DataCon, which was created to help researchers in the field of education. It was developed based on the premises of Corpus Linguistics and the Data-Driven Approach. This study also includes an evaluation of this proposal by graduate students, who were asked to consider its main contributions and limitations. The evaluations highlighted contributions of the DataCon methodology related to its technological and strategic potential for searching, compiling and managing data, and some limitations related to analyses, time management and digital skills.

Keywords: Education. Teaching. Methodology. Research.

Resumen: Este artículo destaca una propuesta metodológica denominada Datacon, creada para ayudar en la investigación en el área educativa. Se desarrolló a la luz de la Lingüística de Corpus y el Abordaje orientado a datos. Se incluye la evaluación de esta propuesta por parte de estudiantes de posgrado, en la que consideran los principales aportes y limitaciones. Los resultados muestran aportes relacionados al potencial tecnológico y estratégico de la Metodologia DataCon para realizar la búsqueda, recopilación y gestión de datos y limitaciones relacionadas con los análisis, gestión del tiempo y habilidades digitales.

Palabras clave: Educación. Enseñanza. Metodología. Investigación.





Considerações iniciais

Este artigo apresenta uma proposta metodológica, denominada DataCon, desenvolvida para auxiliar estudantes na realização de pesquisas na área educacional. Tem como cenário de fundo, o âmbito da pós-graduação, modalidade que tem entre suas metas,

[...] formar e qualificar para a pesquisa" e para a produção de conhecimento científico, com significado coletivo e social, assim como viabilizar a continuidade, o aprimoramento e [...] a multiplicação dos processos aprendidos para desenvolver e implementar pesquisas (FREITAS; SOUZA, 2018, p. 9).

As aprendizagens envolvidas nessa modalidade formativa visam à preparação de um pesquisador experiente, cujos domínios epistemológicos possam auxiliar no desenvolvimento de projetos de pesquisa autônomos e independentes (FREITAS; SOUZA, 2018). Nessa perspectiva, "a pesquisa exerce papel fundamental, ao desenvolver no aluno, entre outras características, o senso-crítico e a busca pelo autoconhecimento" (PINTO; DE PAULA; ALVES, 2010, p. 21).

Dessa forma, "pesquisa e formação [...] são mutuamente dependentes [...]" (FREITAS; SOUZA, 2018, p. 9), cabendo à pós-graduação fomentar práticas, estimular o uso de ferramentas tecnológicas e consolidar embasamentos epistemológicos necessários para a formação dos estudantes durante o percurso investigativo.

Entre as justificativas que respaldam este estudo, cita-se a importância dessa dimensão investigativa na formação profissional e a necessidade de diminuir as dificuldades enfrentadas pelos estudantes na realização de suas pesquisas (AIELLO-SINDONI, 2009). Entre as principais dificuldades destaca-se a elaboração de projetos e de artigos científicos durante um curso de mestrado, em que a "escolha do problema de pesquisa, coleta de dados e definição da revista ou periódico para envio do trabalho; seguidas de: escolha do tema de pesquisa e delineamento da metodologia de pesquisa" (PINTO; DE PAULA; ALVES, 2010, p. 32) são evidenciados como os principais obstáculos.

De acordo com Freitas e Souza (2018, p. 14), "[...] os desafios propostos para essa formação, assim como a análise sobre as variadas práticas que vêm sendo implementadas [...] ainda têm tido pouca visibilidade no cenário das discussões científicas". Entre esses desafios, destacam-se dimensões relacionadas ao desenvolvimento de pesquisas e decisões epistemológicas que afetam o desenvolvimento dos trabalhos científicos.

Essas "reflexões evidenciam a necessidade de uma nova cultura no fazer científico, com base no uso de recursos tecnológicos e estratégias que possam contribuir para a realização de pesquisas nas áreas de Ensino e Educação, nos âmbitos epistemológico e metodológico" (COELHO, 2022)

Considerando tais argumentos, cabe ao professor permitir que os estudantes conheçam estratégias para que possam superar esses desafios, de forma autônoma (AIELLO-





SINDONI, 2009), além de estimular para que se envolvam "[...] mais profundamente com a pesquisa científica, visando uma formação profissional mais crítica, voltada para a autonomia intelectual" (PINTO; DE PAULA; ALVES, 2010, p. 33).

Nesse sentido, o DataCon é apresentado como uma possibilidade estratégica e metodológica para potencializar as práticas de pesquisa no âmbito educacional, com foco no desenho, gerenciamento e análise de dados automatizados, sob a ótica reflexiva, crítica e autônoma.

O DataCon foi construído à luz dos pressupostos teóricos da Linguística de Corpus (BERBER SARDINHA, 2004) e da Abordagem Orientada por Dados (JOHNS, 1991), que tem o aprendiz como objeto central e o potencial uso de recursos tecnológicos e de programas de concordância. Nesse contexto, a fim de responder à problemática deste artigo – quais as principais contribuições e limitações da proposta Datacon para realização de pesquisas no âmbito educacional –, utiliza-se um questionário aplicado a um grupo de estudantes da pós-graduação, cuja interpretação dos resultados é naturalmente qualitativa, proporcionada pelo método da análise de conteúdo (BARDIN, 2011).

Enquadramento teórico

A metodologia que se apresenta neste artigo, intitulada DataCon, tem como intuito contribuir para o delineamento do desenho, gerenciamento e análise dos dados. Concretizase a partir de um planejamento filosófico e procedimental que integra três dimensões para o desenho da investigação: a epistemológica, a estratégia geral e a técnica. A epistemológica envolve aspectos como a intencionalidade, problema, as hipóteses e a função da teoria. A dimensão da estratégia geral inclui a seleção do tipo de desenho (quantitativo, qualitativo ou misto), e a seleção de casos, por meio de critérios amostrais, apoiando-se, sempre que possível em noções básicas da representatividade estatística. Por outro lado, a dimensão técnica inclui a seleção e aplicação de técnicas de obtenção, análise e validação das informações empíricas (SIRVENT, 2007).

Esse conjunto de três dimensões enfatiza dois contextos principais: o de verificação, pautado na perspectiva quantitativa e o contexto da descoberta, pautado na perspectiva qualitativa, demonstrando a relevância da integração dessas duas lógicas de análise para o desenvolvimento de um processo de (re)construção de problemas de pesquisa e criação de novos saberes argumentativos, gerados a partir da curiosidade e dos interesses de investigação.

A integração entre teoria e informação empírica, entre descoberta e verificação, se constrói amparada na Linguística de Corpus (BERBER SARDINHA, 2004), uma disciplina empírica que tem como foco a base em dados, cujo preceito básico é a descoberta e na Abordagem Orientada por Dados (JOHNS, 1991), compondo a gênese dessa proposta metodológica.





Nesse processo, considera-se a composição de uma problemática de pesquisa, materializada linguisticamente no formato de uma pergunta, seguida pela composição de um *corpus* intencional, cuja finalidade principal é servir como ponto de partida para a realização da verificação e terreno fértil para as descobertas. Nessa metodologia, "[...] os alunos interagem com o *corpus* por meio de tarefas da pesquisa" (MISHAN, 2004, p. 222).

Para essa fase, toma-se a modalidade investigativa consolidada sob a denominação de Estado da Arte. Trata-se de uma "metodologia de pesquisa qualitativo-documental crítico-interpretativa" (GÓMEZ, GALEANO; JARAMILLO, 2015, p. 423), que visa a auxiliar no mapeamento e discussão de "certa produção acadêmica em diferentes campos do conhecimento, tentando responder que aspectos e dimensões vêm sendo destacados e privilegiados em diferentes épocas e lugares [...]". Essas pesquisas possuem uma abordagem inventariante e descritiva para as quais se deve levar em conta um conjunto de categorias que possam caracterizar o fenômeno a ser analisado.

Dessa forma, busca-se recuperar o conhecimento acumulado sobre o objeto de estudo, tentando identificar os principais questionamentos, contribuições, limitações, lacunas, motivações, categorias de análise, evidências e resultados já indicados previamente. Essa metodologia evita duplicar esforços desnecessários no âmbito da pesquisa, evitar a repetição de estudos que já foram desenvolvidos, buscar a originalidade, encontrar possíveis erros ou limitações relacionadas a um objeto investigável e contribuir cientificamente na identificação de novos problemas de pesquisas e na construção de novos conhecimentos científicos.

Essa recuperação e mapeamento da produção acadêmica relacionada ao objeto de estudo se realiza com auxílio da compilação de um *corpus* específico, de ferramental e de novas estratégias de pesquisa, como a que se apresenta neste artigo com o uso da proposta DataCon. As estratégias visam a auxiliar na recuperação do conhecimento, compilação de amostra, gerenciamento e análise dos dados, em busca de possíveis evidências de aprendizagem, das lacunas deixadas por outros estudos e de novas descobertas que surgem com a proposta de uma pesquisa orientada por dados.

Essas estratégias são guiadas e fundamentadas na hermenêutica, na reflexão, na avalição crítica da qualidade dos estudos que compõem o *corpus*, na compreensão e nas análises qualitativas, podendo integrar-se às métricas, que possibilitam diferentes análises, univariadas, bivariadas ou multivariadas, dependendo da modalidade investigativa.

O uso de estratégias nessa perspectiva integrativa, entre o enfoque hermenêutico e uso de métricas, auxilia na recuperação, organização e interpretação do fenômeno investigado de uma forma diferenciada, com o propósito de interpretar o material bibliográfico e ir além das informações encontradas. Esse processo implica transcender reflexivamente, a partir das análises realizadas à luz de valores epistemológicos e não epistemológicos, tais como os valores culturais, ideológicos, sociais, etc., que podem contribuir para a área educacional, a depender do problema de pesquisa.





Charaudeau (2011, p. 16) ressalta que o *corpus* "propõe um certo caminhar intelectual a partir de hipóteses de base para descobrir "fatos e ideias", isto é, para interpretar". Nesse sentido, "interpretar consiste em uma operação de correlação de fatos observados com outros fatos, para formular, por inferência, hipóteses de sentido".

Lembremos que a lógica da pesquisa se baseia numa abordagem "empírico-dedutiva" de idas e vindas entre, de um lado, a determinação dos objetivos de análise e das hipóteses significantes, e de outro, os métodos de observação dos objetos e de coleta de dados segundo uma instrumentação que deve poder dar conta dos objetivos e das hipóteses previamente estabelecidos. É num vaivém entre esses dois polos que se elaboram conceitos, instrumentos de análise e interpretação dos resultados (CHARAUDEAU, 2011, p. 19, grifo do autor).

Nesse contexto, apresenta-se uma possibilidade de aplicação efetiva do uso de *corpus* para a construção de uma metodologia relevante para auxiliar no processo de desenho do projeto de pesquisa, considerando a utilização de recursos computacionais, *softwares* e ferramentas específicas, de pertinência quantitativa e/ou qualitativa, que auxiliam na coleta, armazenamento, exploração, sondagens provocadas e análise dos dados. À vista disso, propõe-se uma pesquisa atrelada à descoberta, ao questionamento, à busca por evidências e ao manuseio de dados.

Essa concepção desdobra-se na necessidade da identificação de critérios explícitos e compilação de um *corpus* organizado de maneira criteriosa, permitindo a busca de informações e de descobertas, de modo confiável. Para tanto devem ser levados em conta alguns critérios no momento da compilação: (a) a origem: os dados devem ser autênticos; (b) o propósito: o *corpus* deve ter a finalidade de ser um objeto de estudo linguístico; (c) a composição: o conteúdo do *corpus* deve ser criteriosamente escolhido; (d) a formatação: os dados do *corpus* devem ser legíveis por computador; (e) a representatividade: o *corpus* deve ser representativo de uma língua ou variedade e (f) a extensão: o *corpus* deve ser vasto para ser representativo (BERBER SARDINHA, 2000, p. 338).

Essa gestão dos dados também deve considerar questões relacionadas ao conteúdo selecionado, a partir de critérios e circunstâncias rígidas de coleta do *corpus*, cujas escolhas têm relação com o posicionamento teórico e os processos analíticos do fenômeno em estudo. Essas escolhas são realizadas contemplando a finalidade, a representatividade, o tratamento teórico-metodológico e a influência da problemática, aspectos que evidenciam a circularidade do "duplo movimento dedutivo-indutivo da pesquisa em ciências humanas e sociais", que pode ser representado pelas expressões: "Dize-me qual é o teu corpus, eu te direi qual é tua problemática / dize-me qual é a tua problemática, eu te direi qual é o teu corpus" (CHARAUDEAU, 2011, p. 16).

Trata-se de uma abordagem centrada no estudante, que busca compreender o funcionamento de determinado fenômeno, por meio da observação de um *corpus* compilado para tal fim. De forma análoga, a proposta potencializa uma pesquisa orientada por dados,





que tem por objetivo utilizar o *corpus* na construção do Estado da Arte e no fornecimento de dados empíricos (evidências), referente à problemática de pesquisa.

É utilizado, portanto, como uma estratégia metodológica para a realização de pesquisas, visando a tornar o estudante protagonista, procedimento viabilizado pelo processo da descoberta e desenvolvimento de estratégias automatizadas e personalizadas para o gerenciamento e análise dos dados. Dessa forma, o papel do professor é propiciar meios para o desenvolvimento de novas estratégias que possam fomentar o ato de descobrir. Trata-se de um processo de imersão nos dados, em que o computador é considerado um elemento central para a gestão e processo de (re)construção de novos conhecimentos.

A importância do uso desses recursos leva em conta o fato de que "[...] a virada computacional na produção de conhecimento científico não é nada subestimada" e "o digital transforma especialmente as possibilidades de coleta de dados" (SILVA, 2015, p. 2) que podem potencializar novas práticas, frente ao caráter dialético dos fenômenos científicos, especialmente, ao tomar o ensino e a formação docente como objetos investigáveis. Nesse processo, compreende-se a importância dos dados e variáveis que "[...] representam os diferentes níveis de integração da realidade humana e, portanto, da complexidade envolvida que implica fazer recortes no nosso objeto" (BRAVIN; PIEVI, 2008, p. 119).

Metodologia

A implementação da proposta Datacon foi realizada por meio de um desenho qualitativo, com o objetivo de identificar as principais contribuições e limitações encontradas por um grupo de doze estudantes da pós-graduação, de diferentes áreas (física, matemática, engenharia, geografia, computação, pedagogia, língua inglesa e biologia), durante a compilação do *corpus*, gerenciamento e análise. Vale ressaltar que se trata de uma amostra não probabilística, pois a técnica de amostragem não utiliza seleção aleatória (MALHOTRA, 2007, p. 325), sendo selecionada por acessibilidade e por tipicidade (VERGARA, 2003, p. 50).

Os estudantes participaram de aulas práticas, no contexto remoto, para uso do DataCon que integra o uso de um *software* e de uma proposta metodológica que visa a potencializar estratégias de pesquisa com base em dados (oriundos da Web, de *corpus* de aprendizes ou de técnicas de coleta, como entrevistas e questionários). A proposta envolve cinco etapas: 1) Identificar o problema de pesquisa; 2) Selecionar e empregar estratégias de pesquisa; 3) compilar um *corpus* de pesquisa; 4) explorar e analisar os dados de pesquisa; e 5) identificar os resultados de pesquisa (COELHO, 2022).

Os estudantes selecionaram problemáticas e questões norteadoras, de acordo com seus interesses de pesquisa, em uma perspectiva em que o ensino e a formação docente foram tomados como objetos investigáveis. A partir disso, compilaram um *corpus*, seguindo as etapas mencionadas, que permitiu o gerenciamento e análise, em busca das possíveis respostas às questões de pesquisa. Para a aplicação da proposta metodológica tiveram acesso a dois programas (Sphinx iQ2 e AntConc).





A partir desses *softwares*, foi possível explorar os itens, organizar listas de palavras por frequência de uso e analisar as variáveis da pesquisa por meio das concordâncias, sob uma abordagem analítica, orientada pelos dados e pautada nos dois contextos que fundamentam a gênese da proposta, a descoberta e a verificação. Vale destacar que apesar de o AntConc não oferecer possibilidades para análises estatísticas mais complexas, como o Sphinx iQ2, foi utilizado para visualização e exploração dos dados, com técnicas quantitativas de contagem de palavras e de frequência.

Para dar conta de responder a problemática deste artigo, foi utilizado um questionário, cujos dados foram interpretados mediante a técnica de análise de conteúdo (BARDIN, 2011), de acordo com três etapas: 1) pré-análise; 2) exploração do material; e 3) tratamento dos resultados e interpretação. Essa análise também contou com auxílio do *software* Sphinx iQ2 para organização, identificação da frequência de palavras e constituição das unidades de registro que foram categorizadas de maneira progressiva, com devido compromisso com os participantes e rigor metodológico, amparado nos parâmetros da Linguística de Corpus (BERBER SARDINHA, 2004).

Cumpre ainda mencionar as questões éticas relacionadas à realização desta pesquisa, especialmente, envolvidas na definição, descrição do problema, na coleta e interpretação de dados. Diante do período pandêmico, não houve a possibilidade de revisão ética da pesquisa, realizada por um Comitê. Contudo, considerando as diretrizes regulamentadoras de pesquisas envolvendo seres humanos, na área de Educação. Os participantes receberam e assinaram um Termo de Livre Consentimento Esclarecido, detalhando os procedimentos, a importância da pesquisa, os benefícios e potenciais riscos envolvidos, bem como a finalidade e a abrangência. As questões éticas relacionadas à coleta também envolveram a obtenção de permissão institucional, a privacidade dos participantes, a curadoria, armazenamento adequado dos dados e publicação, garantindo o anonimato e a confidencialidade (CRESWELL, 2007).

Resultados e Discussões

Foi realizado um agrupamento progressivo das categorias, no qual foram identificadas 40 categorias iniciais relacionados às contribuições (Quadro 1) e 20 categorias iniciais relacionadas às limitações (Quadro 2). As categorias de análise emergentes foram agrupadas, a partir das lentes de quatro domínios: teórico-metodológico, tecnológico, operacional/estratégico e analítico/simbólico.

Quadro 1: Categorias referentes às contribuições da Proposta DataCon

Categorias iniciais

Facilidade na compreensão do desenho do projeto e identificação do problema

Conhecimento de uma nova ferramenta para estudo e pesquisa





Delimitação e refinamento de perguntas e de variáveis

Identificação do problema de pesquisa de forma mais clara

Seleção específica de trabalhos

Estabelecimento de critérios baseados em números e dados

Organização e sistematização dos dados de pesquisa

Maior capacidade de selecionar dados para a pesquisa e melhoramento das análises

Exploração simplificada para os parâmetros a serem analisados

Pesquisas com múltiplas variáveis

Categoria final: Domínio teórico-metodológico

Compilação de um corpus de pesquisa, com uso de banco de dados, em formato eletrônico

Construção de categorias e suas variáveis proporcionando uma pesquisa mais detalhada e com maior grau de precisão

Otimização da exploração e análise dos dados de pesquisa

Possibilidade de realizar análise lexical

Facilita a leitura de concordâncias

Resposta aos questionamentos da pesquisa

Velocidade

Agilidade na coleta de dados

Facilidade no acesso dos arquivos

Clareza

Eficiência

Praticidade

Objetividade na busca e seleção

Retroalimentação entre pesquisa e o problema

Facilidade na organização dos dados

Acesso gratuito ao software de pesquisa

Gerenciamento

Otimização do tempo para elaboração do problema de pesquisa, buscas e análise do corpus

Categoria Final: Domínio Tecnológico

Seleção e emprego de estratégias de pesquisa

Uso de estratégias de busca e adoção para uma perspectiva do ensino como um objeto investigável e principalmente a escolha do recurso metodológico

Potencial uso de repositórios e buscadores





Uso de estratégias para compilação de corpus

Possibilidade de combinações para extração de dados

Uso operadores lógicos e booleanos que aceleram o processo de busca

Criação de novas estratégias de pesquisa visando à busca na web

Fragmentação do todo em variáveis e comparação de dados

Emprego de diferentes estratégias

Categoria final: Domínio operacional/estratégico

Ressignificação de um campo de estudo

Abertura de um novo horizonte para as estratégias de pesquisa

Fomento à autonomia

Categoria final: Domínio analítico/simbólico

Fonte: Elaboração própria.

No Quadro 2 são apresentadas as principais limitações associadas à aplicação da proposta DataCon, organizadas nos quatro domínios já mencionados.

Quadro 2: Categorias referentes às limitações da proposta DataCon

Categorias iniciais

Necessidade de conhecimento da Linguística de Corpus

Dificuldades iniciais para compreensão da proposta

Categoria final: Domínio teórico-metodológico

Habilidades para uso do software para realização das análises

Desconhecimento das ferramentas de pesquisa

Conhecimento técnico dos repositórios

Dificuldades atreladas ao uso dos softwares

Desconhecimento de grande parte das funcionalidades do software usado

Problemas como internet, ferramental e dispositivos

Categoria Final: Domínio Tecnológico

Dificuldades para montar string de busca e encontrar palavras-chave

Tempo curto para exploração das bases

Limitações relacionadas ao tempo disponível para a pesquisa

Necessidade de otimização do tempo

Falta de habilidade para realizar a análise de dados

Dificuldades na exportação dos dados

Tratamento de arquivos que demandam muito tempo e solução de problemas

Trabalho para preparar os textos (PDF/TXT)





Necessidade de um ferramental atualizado, um bom dispositivo de trabalho e uma internet razoável

Categoria final: Domínio operacional/estratégico

Processo cansativo

Dificuldade de vislumbrar o "todo" do processo

Reconhecer-se como pesquisador

Categoria final: Domínio analítico/simbólico

Fonte: Elaboração própria.

Com base nessa categorização, o domínio teórico-metodológico compreende os processos relacionados à abordagem metodológica adotada e suas implicações. De acordo com os resultados, entre as principais contribuições atreladas a esse domínio, destacam-se o potencial uso da proposta DataCon para a criação do desenho da pesquisa, a identificação e o refinamento das perguntas, possiblidades investigativas com uso de múltiplas variáveis, seleção objetiva de trabalhos, estabelecimento de critérios, resultados baseados em dados, organização e sistematização dos dados, facilidade na compilação de um corpus e seu gerenciamento, exploração simplificada e melhoria das análises.

Os excertos, a seguir, representam algumas das respostas dos participantes da aplicação que fazem referência ao domínio teórico-metodológico:

- i) A proposta abriu um novo horizonte para as estratégias de pesquisa (E3).
- ii) Contribui para construção de um problema novo, ou, a tratar de uma nova forma um velho problema existente (E4).
- iii) Acredito que as contribuições são inúmeras, porém, vou atribuir a que chamou minha atenção: a forma de estabelecer critérios, baseados em números, para construção de uma pesquisa (E5).

Mediante o exposto pelos estudantes, de maneira geral, as contribuições foram visíveis quanto ao desenho da investigação, a adoção de estratégias, emprego de novos formatos de gerenciamento e análise dos dados. Os parâmetros e percurso adotado podem contribuir para o uso de procedimentos parcialmente sistemáticos de coleta e tratamento, conforme apontado por E5, ao destacar o uso de critérios e estatística que podem complementar as análises qualitativas. Essa constatação evidencia a relevância do uso dos pressupostos da Linguística de Corpus que podem impactar na formação para a pesquisa no âmbito educativo e possibilitar novos horizontes de pesquisa.

De acordo com os resultados, as estratégias associadas ao uso dos recursos tecnológicos podem auxiliar na fase de imersão nos dados, etapa fundamental no processo de investigação, em busca de respostas à problemática, amparadas nos dados de pesquisa. Nesse processo, o computador e o *corpus*, constituído em meio digital, são considerados elementos centrais para o desenvolvimento de análises automatizadas, com possibilidade para (re)análise e (re)uso dos dados por diferentes atores sociais.





de busca, podendo observar e selecionar as palavras mais representativas, por meio da quantificação do vocabulário evocado, além de ampliar as análises, retomando o contexto das palavras para extrair significados dos enunciados (por meio do uso da estratégia das linhas de concordância).

Ao buscar um determinado termo, é possível observar, em tela, os principais resultados da busca e ir além, ampliando a observação das significâncias inscritas nos textos, em função da problemática, a fim de ampliar as descobertas e gerar novas informações e conhecimentos apoiados na análise e na síntese. Com isso, possibilita-se maior objetividade e inovação no processo de interpretação dos dados textuais, por meio de um processo que integra níveis quantitativos e qualitativos (hermenêutica e métrica).

Dando continuidade à discussão, diferentes contribuições integram o domínio tecnológico: a compilação de um corpus de pesquisa com uso de banco de dados em formato eletrônico, a construção de categorias e suas variáveis, proporcionando uma pesquisa mais detalhada e com maior grau de precisão, a otimização e o gerenciamento, a exploração dos dados de pesquisa, a potencial leitura de concordâncias e a realização de diferentes análises, por exemplo, a lexical.

Ainda relacionado a esse domínio, foi mencionado um conjunto de elementos que fazem menção à relevância da proposta: velocidade, agilidade na coleta de dados, clareza, facilidade no acesso dos arquivos, eficiência, praticidade, objetividade na busca e seleção, otimização do tempo, retroalimentação entre pesquisa e problema, facilidade na organização dos dados, bem como a possiblidade de ter acesso gratuito ao *software* de pesquisa. Alguns exemplos são destacados a seguir:

- iv) Através das estratégias de compilação dos dados é possível construir um projeto de acordo com uma lógica sequencial amparada na análise linguística, essa contribuição é fundamental para construção do problema de pesquisa (E6).
- v) Com base nas explicações da professora, ficou mais claro o norte a seguir dos objetivos preciso do desenho da pesquisa e pude perceber os seguintes: Técnicas a serem implementadas para coleta de dados, método aplicado para a análise dos dados coletados. Tipo de metodologia de pesquisa, problemas na investigação e medição da análise (E9).
- vi) Agilidade; praticidade; retroalimentação entre pesquisa e o problema (E10).

O domínio operacional/estratégico faz referência à seleção e emprego de estratégias de pesquisa, busca e adoção de métodos iniciais de produção de dados. Entre as contribuições ressaltadas, estão: o uso de estratégias que podem ser utilizadas no sistema de busca, seleção, coleta e gerenciamento de dados para responder à problemática, fomento à exploração das variáveis, ampliação da visão do contexto da pesquisa e das variáveis de análise, potencial uso de repositórios e buscadores, uso de operadores *booleanos* que aceleram o processo de busca avançada, assim como o emprego e a criação de novos procedimentos de pesquisa.

Vale ressaltar que uma das contribuições mais destacadas pelos participantes faz referência à possiblidade de observação de diferentes variáveis, considerando que podem





"influenciar no que se está estudando" (AIELLO-SINDONI, 2009, p. 151). Nesse contexto, o processo de identificar variáveis é uma das principais estratégias da proposta para a realização das análises, com o intuito de evocar dados, sob uma lógica quantitativa e qualitativa, de forma integrada.

Esse processo analítico destaca a primazia do dado, possibilitando um processo que permite encontrar resultados para os objetos investigáveis e os "achados de pesquisa", por meio das concordâncias e do cruzamento de variáveis que não foram evocadas inicialmente, mas que foram identificadas, durante o processo analítico. Dessa forma, o *corpus* caracterizase como um produto que também tem por finalidade potencializar descobertas de novas problemáticas.

- vii) Uma vez que o corpus foi constituído conforme as orientações, as evidências relacionadas à problemática são coerentes e fáceis de identificar (E1).
- viii) Para mim foi revolucionário, estava acostumado a baixar inúmeros artigos, criar critérios como ano, revista e a partir daí realizar leituras e fichamentos. Usando o ferramental tudo ficou mais objetivo (E4).
- ix) As estratégias, amparadas na metodologia da Linguística de *Corpus*, observei que o planejamento e a execução de uma pesquisa compreendem, principalmente, as estratégias de pesquisa para Formulação do problema; Revisão de literatura na *web*; Metodologia e, principalmente, utilizar a *web* como ferramenta para encontrar determinadas variáveis (E9).

As estratégias apontadas pelos estudantes evidenciam um processo que permite, com base nos pressupostos da Linguística de *Corpus*, fragmentar o objeto de pesquisa em possíveis variáveis, com o intuito de otimizar a exploração de categorias e unidades de sentido que estão ligadas às questões norteadoras.

Apesar de reconhecer que em educação "[...] os fatos acontecem de maneira tão imbricada que fica difícil isolar as variáveis envolvidas" (ABREU; ALMEIDA, 2008, p. 77), vale ressaltar que o processo de fragmentar o fenômeno educacional em variáveis, nessa proposta, tem como objetivo permitir um estudo analítico capaz de elucidar possíveis categorias responsáveis por determinados efeitos.

O próximo domínio "analítico/simbólico" agrupa alguns aspectos que fazem referência a uma proposta de ressignificação de estratégias de pesquisa. Essa dimensão remete ao protagonismo e à autonomia do estudante na condução de sua pesquisa, por meio de uma proposta em que os dados e as respectivas análises definem a dinâmica do processo investigativo e evocam uma dimensão simbólica que pode transcender os espaços formativos.

Esse percurso evidencia um caráter dinâmico que emerge da relação entre os sujeitos da pesquisa, as problemáticas investigadas, os dados, as análises que são realizadas e a síntese dos resultados, alcançadas por meio de um olhar crítico e reflexivo, que deve permanecer no horizonte epistêmico das ciências sociais e humanas, levando em conta que "[...] a pesquisa é, também, um estudo pessoal, pois carrega em si marcas, inferências





e atitudes investigativas de quem a faz" (ABREU; ALMEIDA, 2008, p. 75). Seguem alguns excertos para elucidar tais argumentos.

- x) Minha pesquisa foi totalmente beneficiada com o método. Usarei em todas daqui para frente (E1).
- xi) Com aprimoramento do processo aprendido na disciplina, sinto-me capaz e motivado para realizar novas pesquisas (E11).
- xii) O método + ferramenta, diferentemente do que imaginava no meio do processo, colaborarão muito nas minhas perspectivas para torna-me um pesquisador cada vez mais autônomo e ativo (E3).

Os resultados corroboram o uso das estratégias, no sentido de potencializar a autonomia do aprendiz no direcionamento de seu projeto de pesquisa, assim como a motivação para a aplicação em diferentes contextos, visando a desenvolver habilidades crítica e analítica, aspectos considerados significativos no âmbito acadêmico.

A partir desse momento, são evidenciadas as categorias relacionadas às limitações. Com relação ao domínio teórico-metodológico, o principal aspecto apontado referese à necessidade de conhecimento da Linguística de *Corpus* e a dificuldade inicial para compreender a proposta metodológica apresentada.

Por outro lado, o domínio tecnológico abrange um conjunto de limitações subjacentes à necessidade de desenvolvimento de habilidades digitais para uso dos *softwares* para otimização das análises, das ferramentas e suas funcionalidades, conhecimento técnico dos repositórios, além dos problemas relatados como o acesso à internet, conforme alguns dos excertos que seguem.

- xiii) Entre as limitações: problemas como *internet*, ferramental, dispositivos, na minha concepção, é um fato limitador, é uma questão de infraestrutura que permite ou não o desenvolvimento da proposta DataCon (E4).
- xiv) Entre a limitação verifica-se a necessidade de possuir habilidades para trabalhar com *softwares* de análise de dados (E5).

Vale ressaltar que a aplicação da proposta foi desenvolvida em contexto remoto, não sendo possível o uso da infraestrutura institucional para a aplicação em laboratório de informática, com devido acompanhamento presencial do docente. Esse contexto, segundo os estudantes, ampliou as dificuldades.

- xv) As aulas remotas ainda possuem algumas limitações que propiciam uma baixa na qualidade da aula, por exemplo, na hora de tirar as dúvidas, o processo se tornava um pouco lento, mas a professora se esforçava para sanar todas. As quedas de acesso à internet também me causaram transtornos nas últimas aulas, mas estas dificuldades foram contornadas assistindo aos vídeos das aulas gravadas (E1).
- xvi) Nas limitações, acredito que temos muito conteúdo para pouco tempo. Acredito que poderíamos ter um pouco mais de tempo para as aulas práticas. No quesito tecnológico, eu tive algumas dificuldades por ser de forma *online*, mas compreendo a situação peculiar que estamos vivendo (E7).





A adaptação da dinâmica das aulas remotas, especialmente, quando se faz necessário o uso de laboratórios, recursos tecnológicos e aulas práticas é desafiador. Os participantes destacam questões relacionadas à escassez de tempo, falta de ambiente adequado, necessidade de internet compatível e de treinamento para uso da tecnologia.

Na sequência da análise, o domínio operacional/estratégico engloba limitações referentes às dificuldades que os estudantes tiveram para construção de uma *string* de busca e identificação das palavras-chave. Soma-se a isso, a necessidade do desenvolvimento de habilidades para realização das análises, dificuldades na exportação dos dados, demanda de tempo para tratamento de arquivos, preparação dos textos (PDF/TXT) e solução de problemas que surgem durante o processo investigativo.

As principais limitações subjacentes a esse domínio, apontadas pelos estudantes, referem-se ao gerenciamento de tempo (considerando que o *corpus* compilado e as análises empreendidas foram realizadas em um período de 30 horas) e aos problemas tecnológicos decorrentes da aplicação da proposta em formato remoto.

A última lente a ser mencionada refere-se ao domínio analítico/simbólico, cujas limitações fazem menção ao cansaço e à dificuldade de vislumbrar o "todo" do processo, conforme excertos que seguem. Nesse âmago, perceberam-se os principais desafios causados, de acordo com os excertos:

- xvii) Conteúdo (a dificuldade de vislumbrar o "todo" do processo, deixou muito turva a caminhada) âmbito tecnológico (tratamento de arquivos demandaram muito tempo para solucionar os vários problemas que aconteceram) (E2).
- xviii) O fator negativo é somente o fator cansaço, todo esse processo é cansativo, mas recompensador (F8)
- xix) A minha principal limitação não fora no âmbito tecnológico e sim em me enxergar como pesquisador (E7).

Tais resultados ressaltam a necessidade de refletir sobre os fatores que influenciam no processo investigativo, que vão além dos saberes e habilidades tecnológicas, incluindo as atitudes frente à pesquisa. Essa atitude implica o reconhecimento dos constantes desafios que demandam a necessidade de adaptações rápidas, busca de novas aprendizagens e formas para condução da pesquisa.

Essas demandas ficaram ainda mais evidentes durante o período pandêmico, devido às novas exigências e adaptações na condução e aplicação das pesquisas e ampliação das estratégias de coleta de dados "[...] para contemplar adaptações e novos recursos que permitam a continuidade das pesquisas" (SCHMIDT; PALAZZI; PICCININI; 2020, p. 961).

Nessa perspectiva, "fator cansaço" e "reconhecer-se como pesquisador", apesar de terem sido apontados uma única vez, evidenciam a necessidade de reflexão sobre a relevância do desenvolvimento de uma nova cultura de investigação e compreensão das





complexidades que envolvem o processo investigativo. Esse processo também envolve a autoavaliação e a responsabilidade do estudante, em relação à sua pesquisa e processo formativo.

Vale ainda ressaltar que os resultados dos participantes relacionados às dificuldades teóricas e conceituais, operações práticas e decisões concretas enfrentadas no trabalho de investigação qualitativa, principalmente, na etapa de análise de dados, também foram apontados por outras pesquisas (AIELLO-SINDONI, 2009; MEJÍA NAVARRETE, 2011).

Com relação à particularidade da análise qualitativa, Mejía Navarrete (2011) destaca que se trata de um processo flexível, cujas etapas se inter-relacionam.

É flexível porque a análise se adapta, se molda e emerge segundo a dinâmica da investigação concreta dos dados. A integração dos componentes da análise é em espiral (redução, análise descritiva e interpretação), se influenciam uns com outros ao mesmo tempo, são processos paralelos e o mais importante, é que assim que acaba uma etapa, se passa a seguinte para, com frequência, voltar novamente à fase anterior e reiniciá-la com uma informação mais acabada e profunda (MEJÍA NAVARRETE, 2011, p. 48).

Sendo assim, levando em conta que uma das "[...] limitações da investigação qualitativa refere-se ao desenvolvimento de metodologias específicas que permitam avançar na elaboração de ferramentas e instrumentos investigativos" (MEJÍA NAVARRETE, 2011, p. 49), destaca-se a relevância do uso da proposta e do ferramental aqui apresentado.

Nesse sentido, acredita-se que a imersão nos dados, a verificação, a identificação de "achados de pesquisa" relacionados ao fenômeno, a possível fragmentação de textos em variáveis, o uso de estratégias, a utilização de recursos tecnológicos e a implementação de análises realizadas em espiral contínuo, com idas e voltas e constantes reflexões sobre o objeto em estudo, podem contribuir para a implementação de uma nova cultura investigativa no âmbito educacional, especialmente, no sentido de sanar as limitações apontadas pelos estudantes, atreladas aos saberes, habilidades e atitudes a serem desenvolvidos durante processo investigativo.

Para tais desafios, acredita-se que o DataCon pode contribuir para: i) desenvolver novas estratégias para auxiliar na análise e interpretação dos dados, utilizando amostras adequadas; ii) potencializar o desenvolvimento de habilidades cognitivas avançadas, visando a ampliar a criatividade; iii) desenvolver a capacidade de trabalhar no gerenciamento do tempo, permitindo sua otimização; iv) fomentar o desenvolvimento da autonomia, a independência e capacidade de tomar decisões em relação ao próprio direcionamento durante a coleta, gerenciamento e análise; v) potencializar a construção de habilidades analíticas para o trabalho com dados, permitindo organizar de forma lógica, diferentes informações; vi) promover estratégias pra desenvolver o pensamento crítico, por meio de decisões pautadas em dados observáveis; e vii) formular estratégias que permitam fomentar a curiosidade, elemento essencial de um bom pesquisador, permitindo expandir





as descobertas.

Antes de concluir, é importante ressaltar que a proposta DataCon se caracteriza como um caminho, um ponto de partida para o qual será necessário expandir os horizontes de investigação. Por meio dessa proposta busca-se contribuir para elevar a qualidade das pesquisas, ampliar o rigor e a efetividade da atividade de busca, gerenciamento e análises, fomentar as descobertas dos vazios recorrentes nas investigações que envolvem o processo de ensino e formação docente.

Soma-se a isso, o potencial desenvolvimento de um trabalho intelectual autônomo, crítico e criativo, pautado na verificação, nas descobertas e na construção de novas atitudes de trabalho científico, responsabilidade com o gerenciamento da pesquisa e com a construção de novos conhecimentos que podem gerar impactos positivos para a área.

Considerações Finais

Os resultados, evidenciados a partir das lentes de quatro domínios: teórico-metodológico, tecnológico, operacional/estratégico e analítico/simbólico, demonstram contribuições e limitações do uso da proposta. Entre as principais contribuições, destacamse: o potencial uso de técnicas, estratégias e recursos tecnológicos para auxiliar na busca, a relevância da compilação de um *corpus* organizado de forma parcialmente sistemática que permite o gerenciamento, a (re)análise e (re)uso dos dados, a possiblidade de articulações entre o material coletado e a problemática.

Entre as principais limitações destacam-se as dificuldades para realização das análises, a necessidade de desenvolvimento de habilidades digitais para uso dos recursos tecnológicos, a gestão do tempo e dificuldades referentes ao conhecimento técnico dos repositórios, obstáculos na exportação dos dados e problemas relacionados ao acesso à Internet.

Os resultados evidenciam a emergência em fomentar novas práticas investigativas, no âmbito da pesquisa, envolvendo dimensões técnicas e de conteúdo, que exigem o desenvolvimento de saberes, habilidades e atitudes subjacentes à investigação científica. A relação com o conhecimento, sob uma perspectiva epistemológica, visa à compreensão dos aspectos cognitivos do desenho da pesquisa e seus desdobramentos. As habilidades, sob um viés metodológico e de relação com a prática, permitem desenvolver as capacidades para a análise e interpretação dos dados. E as atitudes, sob uma perspectiva da dimensão ontológica, visa a potencializar o desenvolvimento da criatividade, a capacidade de otimização do tempo, o reconhecimento das prioridades durante o trabalho de investigação, o fomento à autonomia e independência, a resolução de problemas técnicos e teóricos, a efetiva tomada de decisões durante a coleta, gerenciamento e análise, assim como o fomento à curiosidade e ao desenvolvimento do pensamento crítico e reflexivo, elementos importantes referentes à natureza da realidade do sujeito pesquisador.



Com base na discussão apresentada, a proposta pode contribuir com uma



abordagem menos baseada em impressões subjetivas e mais apoiada em dados de pesquisa (evidências), consubstanciada por um percurso parcialmente sistemático de coleta e seleção dos dados. Esse procedimento metodológico pode ser útil na realização das pesquisas no âmbito educacional, podendo potencializar a discussão e reflexões sobre experiências de aprendizagens e identificação dos avanços na área.

Nesse sentido, considerando que os programas de pós-graduação profissionais se constituem relevantes espaços de aplicação, geração e avaliação de metodologias "[...] que vão além da compreensão dos fenômenos, por incorporarem a perspectiva da aplicação, da intervenção ou do desenvolvimento (FIALHO; HETKOWSKI, 2017, p. 19), ressalta-se a necessidade de fomentar novas pesquisas aplicadas que possam evidenciar os principais obstáculos tecnológicos, implicações e impactos atrelados ao ecossistema do uso dos softwares.

Dessa forma, entre os estudos futuros, destaca-se a necessidade de potencializar debates referentes ao estímulo cognitivo para compreensão de novos formatos de investigação, buscando romper com estruturas de pesquisas convencionais, na tentativa de enfrentar os desafios persistentes no processo investigativo na área educacional e inovar nas práticas e percursos metodológicos.

Investigações vindouras também podem contemplar novas aplicações desta proposta (incluindo estudantes de graduação) e a validação da proposta por um comitê de especialistas, com o intuito de evidenciar itens relacionados à pertinência e à clareza da proposta.

Antes de concluir, vale ressaltar que os resultados relacionados às contribuições e limitações do DataCon são considerados dentro de seu contexto, considerando a amostra analisada, não podendo ser extrapolados para o todo. Portanto, qualquer generalização dos resultados encontrados deve ser tomada com certa cautela, devido às características dos participantes e situação de pesquisa, em um contexto de ensino remoto, o que leva a uma série de implicações, tais como a ausência do uso de laboratório, falta das práticas presenciais e falta de infraestrutura adequada para o uso dos recursos tecnológicos.



Referências

ABREU, Roberta Melo de Andrade; ALMEIDA, Danilo Di Manno de. Refletindo sobre a pesquisa e sua importância na formação e na prática do professor do ensino fundamental. **R. Faced**, n. 14, p. 73-85, 2008. Disponível em: https://repositorio.ufba.br/ri/bitstream/ri/131/2655.pdf. Acesso em: 10 jul. 2022.

AIELLO-SINDONI, María. Dificultades en el aprendizaje de la metodología de la investigación. **Magis Revista Internacional de Investigación en Educación**, v. 2, n. 3, 2009, p. 141-155. Disponível em: https://www.redalyc.org/pdf/2810/28102158008.pdf. Acesso em: 10 jul. 2022.

BERBER SARDINHA, Tony. Linguística de corpus. Barueri, São Paulo; Editora: Manole, 2004.

BERBER SARDINHA, Tony; SHEPHERD, Tania Maria Granja. O Corpus, novas tecnologias e mídias no ensino de inglês: para uma pedagogia do terceiro milênio. *In*: BERBER SARDINHA, Tony *et al*. **Tecnologias e Mídias no Ensino de Inglês**: o corpus nas receitas. São Paulo: Macmillan, 2012, p. 6-14.

BRAVIN, Clara; PIEVI, Néstor. **Documento Metodológico Orientador para la Investigación Educativa**. Coordinación de Investigación del INFD, Argentina, 2008.

CHARAUDEAU, Patrick. Dize-me qual é teu corpus, eu te direi qual é a tua problemática. **Revista Diadorim,** v. 10, 2011, p. 1-23. Disponível em: https://revistas.ufrj.br/index.php/diadorim/article/view/3932/15637. Acesso em: 10 jul. 2022.

COELHO, landra Maria Weirich da Silva Coelho. Desenvolvimento de pesquisas educacionais: implicações teórico-metodológicas, propostas e desafios da gestão de dados científicos. **Revista Exitus**, Vol. 12, p. 1-25, e022069, 2022. Disponível em: http://ufopa.edu.br/portaldeperiodicos/index. php/revistaexitus/article/view/1762/1282. Acesso em: 10 jul. 2022.

CRESWELL, John. **Projeto de pesquisa:** métodos qualitativo, quantitativo e misto. Tradução de Luciana Oliveira da Rocha. 2. ed. Porto Alegre: Artmed, 2007.

FIALHO, Nadia Hage; HETKOWSKI, Tânia Maria. Mestrados Profissionais em Educação: novas perspectivas novas perspectivas da pós-graduação no cenário brasileiro. **Educar em Revista**, Curitiba, Brasil, n. 63, p. 19-34, 2017. Disponível em: https://www.scielo.br/j/er/a/7Fw3HtVgNXg5TZcGgYzGfP H/?format=pdf&lang=pt. Acesso em: 10 jul. 2022.

GÓMEZ, Maricelly; GALEANO, Catalina; JARAMILLO, Dumar Andrey. El estado del arte: una metodología de investigación. **Revista Colombiana de Ciencias Sociales**, 6(2), 2015, p. 423-442. Disponível em: redalyc.org/pdf/4978/497856275012.pdf. Acesso em: 12 jul. 2022.

JOHNS, Tim. Should you be persuaded: two samples of data-driven learning materials. *In*: JOHNS, T.; KING, P. Classroom concordancing. **English Language Research Journal**, v. 4, p. 1-16, 1991.

MEJÍA NAVARRETE, Julio. Problemas centrales del análisis de datos cualitativos. **Revista Latinoamericana de Metodología de la Investigación Social,** n. 1, ano 1, 2011, p. 47-60. Disponível em: http://jbposgrado.org/icuali/An%E1lisis%20de%20datos%20cualitativos%20problemas%20centrales.pdf. Acesso em: 14 jul. 2022.

MISHAN, Freda. Authenticating corpora for language learning: a problem and its resolution. In: **ELT Journal**, vol. 58/3, 2004, p. 219-227.

PINTO, Patrycia Scavello Barreto; DE PAULA, Melisa Maia; ALVES, Francisco José dos Santos. A relação entre as experiências com a pesquisa científica nos cursos de graduação e as dificuldades para elaboração de artigos no mestrado acadêmico. **SINERGIA,** 14 (1): 21-33, 2010. Disponível em: http://repositorio.furg.br/bitstream/handle.pdf?sequence=1. Acesso em: 10 jul. 2022.





SCHMIDT, Beatriz; PALAZZI, Ambra; PICCININI, Cesar Augusto. Entrevistas online: potencialidades e desafios para coleta de dados no contexto da pandemia de COVID-19. **REFACS**, v. 8, n. 4, 2020, p. 960-966. Disponível em: http://seer.uftm.edu.br/revistaeletronica/index.php/refacs/article/view/4877. Acesso em: 10 jul. 2022.

SILVA, Tarcízio Silva. Pesquisa baseada em Dados Sociais Digitais: mapeamento de ferramentas e táticas de coleta de dados no Intercom. **Razón y palavra**, n. 90, 2015, p. 1-19. Disponível em: https://revistarazonypalabra.org/index.php/ryp/article/view/348. Acesso em: 22 jul. 2022.

SIRVENT, María Teresa. **El proceso de investigación**. Las dimensiones de la metodología y la construcción del dato científico. Ficha I. Buenos Aires: Facultad de Filosofía y Letras, Universidad Nacional de Buenos Aires, 1995.

SIRVENT, María Teresa. **Cuadro comparativo entre lógicas según dimensiones del diseño de investigación.** Ficha III. Buenos Aires: Facultad de Filosofía y Letras, Universidad Nacional de Buenos Aires, 1999.

SIRVENT, María Teresa. **El proceso de investigación**. Investigación y Estadística Educacional I. Ficha de Cátedra, 2007.