

O uso do *software* IRAMUTEQ na análise de conteúdo - estudo comparativo entre os trabalhos de conclusão de curso do ProfEPT e os referenciais do programa

The use of the IRAMUTEQ *software* in content analysis - a comparative study between the ProfEPT course completion works and the program references

El uso del *software* IRAMUTEQ en el análisis de contenido - un estudio comparativo entre los trabajos de finalización del curso ProfEPT y las referencias del programa

Recebido: 11/03/2021 | Revisado: 17/03/2021 | Aceito: 23/03/2021 | Publicado: 31/03/2021

Luciana Maria Klamt

ORCID: <https://orcid.org/0000-0001-8561-4762>

Instituto Federal de Educação, Ciência e Tecnologia de Mato Grosso, Brasil

E-mail: lucianaklamt@gmail.com

Vanderley Severino dos Santos

ORCID: <https://orcid.org/0000-0002-0623-2246>

Instituto Federal de Educação, Ciência e Tecnologia de Mato Grosso, Brasil

E-mail: rovander@gmail.com

Resumo

O presente artigo objetiva evidenciar a usabilidade do *software* IRaMuTeq, analisando e comparando os resumos dos TCC, dos estudantes do Mestrado Profissional em Educação Profissional e Tecnológica – ProfEPT, disponíveis na Plataforma Sucupira em dezembro de 2019 com a bibliografia indicada nos exames de acesso ao Programa. O uso do *software* IRaMuTeq mostra-se como uma importante ferramenta de organização de dados para realização da análise de conteúdo, proposta por Bardin (1977), pois possibilita diversas organizações do conteúdo analisado, dentre elas as do tipo Lexográfica, Especificidades, Análise Fatorial de Correspondência (AFC), Método da Classificação Hierárquica Descendente (CHD), Análise de Similitude e Nuvem de Palavras. A aplicação do *software* IRaMuTeq possibilitou realizar estas análises com mais agilidade e demonstrou que há consonância entre os termos usados pelos estudantes do ProfEPT em seus TCCs e as referências teóricas do programa. O *software* mostrou-se bastante útil, principalmente na fase dois da técnica de análise de conteúdo, proposta por Bardin (1977), porém, para se aplicar totalmente esta técnica é imprescindível o protagonismo do pesquisador na fase três, momento de discutir e interpretar os resultados demonstrados pelo *software* livre IRaMuTeq.

Palavras-chave: Análise qualitativa; Ensino; Análise de conteúdo; IRAMUTEQ.

Abstract

This article aims to highlight the usability of the IRaMuTeq *software*, analyzing and comparing the TCC, summaries of the students of the Professional Master in Professional and Technological Education - ProfEPT, available on the Sucupira Platform in December 2019 with the bibliography indicated in the access exams to the Program. The use of the IRaMuTeq *software* proves to be an important tool for organizing data for conducting content analysis, proposed by Bardin (1977), as it allows several organizations of the analyzed content, among them those of the Lexographic, Specificities, Factor Analysis Correspondence (AFC), Descending Hierarchical Classification Method (CHD), Similitude Analysis and Word Cloud. The application of the IRaMuTeq *software* made it possible to perform these analyzes more quickly and demonstrated that there is a consonance between the terms used by ProfEPT students in their TCCs and the theoretical references of the program. The *software* proved to be quite useful, especially in phase two of the content analysis technique, proposed by Bardin (1977), however, to fully apply this technique, the role of the researcher in phase three is essential, when discussing and interpreting the results demonstrated by the free *software* IRaMuTeq.

Keywords: Qualitative analysis; Teaching; Content analysis; IRAMUTEQ.

Resumen

El objetivo de este artículo fue resaltar la empleo del *software* IRaMuTeq, analizando y comparando los resúmenes del TCC, de los alumnos del Máster Profesional en Educación Profesional y Tecnológica - ProfEPT, disponible en la Plataforma Sucupira en diciembre de 2019 con la bibliografía indicada en los exámenes de acceso al Programa. El uso del *software* IRaMuTe resulta ser una herramienta importante para la organización de datos para la realización de análisis de contenido, propuesto por Bardin (1977), ya que habilita varias organizaciones del contenido analizado, entre ellas las del Lexográfico, Especificidades, Análisis Factorial de Correspondencia. (AFC), Método de

Clasificación Jerárquica Descendente (CHD), Análisis de Similitudes y Nube de Palabras. La aplicación del *software* IRaMuTeQ permitió realizar estos análisis de forma más rápida y demostró que existe una consonancia entre los términos utilizados por los estudiantes de ProfEPT en sus TCC y las referencias teóricas del programa. El *software* resultó ser bastante útil, especialmente en la fase dos de la técnica de análisis de contenido, propuesta por Bardin (1977), sin embargo, para aplicar plenamente esta técnica, el rol del investigador en la fase tres es fundamental, a la hora de discutir e interpretar los resultados demostrado por el *software* libre IRaMuTeQ.

Palabras clave: Análisis cualitativo; Enseñando; Análisis de contenido; IRAMUTEQ.

1. Introdução

O presente artigo busca evidenciar a usabilidade do *software* livre IRaMuTeq para organizar dados a serem utilizados no método de análise de conteúdo de um determinado trabalho. A organização de dados a serem contemplados pela análise de conteúdo, quando feita manualmente é bastante trabalhosa e onerosa quanto ao tempo utilizado, pois se vale do conjunto de palavras e suas formas a serem analisadas.

O método de análise de conteúdo constitui-se em um conjunto de técnicas utilizadas na análise de dados qualitativos (Campos, 2004). Para Silva e Fossá (2013), análise de conteúdo são instrumentos metodológicos, que estão em constante aperfeiçoamento e, podem ser usados para analisar conteúdos verbais ou não-verbais.

Segundo Bardin (1977) o objeto da análise de conteúdo é a palavra, em seu aspecto individual e em seu ato de linguagem; a análise de conteúdo trabalha com a palavra, ou seja, o uso da língua por emissores identificáveis, buscando compreender os fatos em um dado momento pelas observações das partes, considerando os seus significados (conteúdos).

Para Bardin (1977) a análise de conteúdo busca encontrar o que está oculto nas palavras, nas “entrelinhas”, busca compreender diferentes realidades por meio das mensagens. Assim, com a semântica, seu material central, a análise de conteúdo estuda o sentido das palavras e seus significados. Com a sociolinguística, estabelece sistematicamente correlações entre estruturas linguísticas e sociais que pode ajudar a descrever características do linguajar e seus grupos sociais. Já a lexicologia e a estatística lexical funcionam como unidades de significações simples que permitem à análise de conteúdo, classificações e a contabilização minuciosas de frequências das palavras em um texto.

A análise de conteúdo permite compreender, o emprego e a aplicação de um certo conteúdo, pois, trata-se de uma leitura “profunda”, determinada pelas condições ofertadas pelo sistema linguístico e tem como objetivo desvendar as relações entre o conteúdo do discurso e os aspectos exteriores (Santos, 2012). Entretanto, conforme destaca Gil (2008), na pesquisa qualitativa as possibilidades para a análise dos dados são múltiplas e depende da capacidade e do estilo do pesquisador.

A capacidade humana apresenta limites e a partir de certo momento pode não ser suficiente para compreender um grande número de dados. Desta forma, a tarefa de analisar dados qualitativos, com apoio de *software*, fica facilitada. E com o propósito de analisar volumes expressivos de textos foi desenvolvido pelo professor Ratinaud (2009), do laboratório de estudos e pesquisas em ciências sociais aplicadas (LERASS), da Universidade de Toulouse o *software* livre IRaMuTeQ (*Interface de R pour les Analyses Multidimensionnelles de Textes et de Questionnaires*).

O IRaMuTeQ é um *software* de análise textual, que funciona ancorado ao programa estatístico R e gera dados, a partir de textos (*corpora* textuais) e tabelas. Os resultados dessas análises demonstram a posição e a estrutura das palavras em um texto, ligações e outras características textuais, que permitem detectar indicadores e, assim, visualizar intuitivamente a estrutura e ambientes do texto a ser analisado.

Segundo Salviati (2017) o IRaMuTeQ realiza análises textuais dos tipos: Estatísticas textuais, Classificação Hierárquica Descendente (CHD); Análises de Similitude; Nuvem de Palavras; Análise de Especificidades; e Análise fatorial de correspondência.

O Mestrado Profissional em Educação Profissional em Rede Nacional (ProfEPT), foi idealizado com vistas a necessidade de promover o aperfeiçoamento das práticas educativas e a gestão escolar vinculadas à educação profissional e

tecnológica (EPT). Também foi implementado com a intenção de atender à demanda por qualificação de profissionais da Rede Federal de Educação Profissional, Científica e Tecnológica (RFEPCT) e ao público em geral. Iniciou suas primeiras turmas no segundo semestre de 2017, contando com dezoito Instituições Associadas (IA), no início do mês de dezembro de 2019 já possuía oitenta e sete trabalhos de conclusão de cursos registrados na plataforma Sucupira.

Os exames de acesso ao ProfEPT são regidos por editais, estes indicam os textos sobre os quais versarão as questões da prova de seleção ao programa. Os principais autores que embasam as referências teóricas do ProfEPT, que são citados nos editais são: Frigotto (2009), Frigotto, Ciavatta e Ramos (2005), Manacorda (1990), Moll (2010), Marx (1968), Saviani (1989), Freire (1996, 2005), Kuenzer (2007), Libâneo (1983), Zabala (1998 e 1999), Pacheco (2015) e Paro (2007).

Estes autores dos referenciais bibliográficos indicados no ProfEPT descrevem que a educação deve ser articulada com o mundo do trabalho para atingir uma formação humana integral ou omnilateral; que a educação ocorra em uma perspectiva de emancipação dos sujeitos e tendo o trabalho como princípio educativo e defendem que o ensino médio integrado propicie uma travessia para a politécnia ou educação tecnológica. Desta forma, o ProfEPT apresenta em suas referências subsídios para o debate Educação e Trabalho como bases a uma Educação Profissional Emancipadora.

A formação integral contradita diretamente com a formação meramente técnica que busca apenas a aprendizagem de atividades repetitivas e rotineiras, de baixa complexidade e que traga ao trabalhador que desenvolve aquela tarefa a compreensão apenas de uma parte da atividade/projeto desenvolvido.

Assim, há a busca pela compreensão do todo do processo de produção e/ou criação, em sua complexidade total. Para que essa compreensão seja possível é necessário conhecer muito além das técnicas, é necessário o conhecimento dos princípios que estão na base da atividade/projeto. Para isso a formação precisa ser ampla e irrestrita, que não selecione apenas os conhecimentos necessários àquela profissão ou àquela atividade.

Para que a educação integral seja possível é necessário vencer a dualidade entre ensino para o trabalho manual e ensino para o trabalho intelectual. Dentro da perspectiva da integralidade, o ensino ofertado para todos deve ter como objetivo o desenvolvimento de todas as potencialidades do indivíduo.

A proposta do ensino integrado, como base teórica de referência do ProfEPT, busca propiciar ao estudante a formação que lhe dê perspectivas de transformação social para si mesmo e para o contexto no qual está inserido.

Sendo estas as bases conceituais deste programa, é esperado que os TCCs produzidos no âmbito do mesmo reflitam a mesma necessidade que alguns autores acima descrevem em seus textos, ou seja, um entrelaçamento entre conhecimentos gerais e formação profissional, alicerçado na politécnia e na omnilateralidade que propicie ao estudante o seu desenvolvimento integral.

Desta forma, buscamos saber: Os trabalhos de conclusão de curso, dos estudantes do ProfEPT, seguem a mesma linha dos trabalhos produzidos pelos principais autores que servem de referência ao programa? O conteúdo dos trabalhos produzidos até 2019 contemplam os principais termos encontrados em suas referências? A organização dos dados por meio do *software* livre IRaMuTeQ é suficiente para a utilização da análise de conteúdo?

Assim, com o objetivo de compreender e testar as funcionalidades do *software* IRaMuTeQ e obter conhecimentos para aplicações futuras em pesquisas acadêmicas no âmbito do ProfEPT que envolva análise de conteúdo com aplicação de *software*, este estudo foi desenvolvido para verificar se os trabalhos de conclusão de curso do ProfEPT acompanham a mesma linha de pensamento presentes nos principais referenciais do programa.

2. Metodologia

Este estudo apresenta um tipo de análise de dados qualitativos de um material textual (resumo das dissertações e trechos das referências bibliográficas indicadas nos exames de acesso ao ProfEPT), para isso, quantificou-se a frequência das

palavras presentes nos textos, comparou os grupos de palavras quanto a essas frequências, também se fez análises complexas e multivariadas com uma diversidade de variáveis quantitativas extraídas dos materiais textuais por meio de um software apropriado.

Desta forma, busca-se neste trabalho apresentar um estudo quali-quantitativo, nessa abordagem metodológica os resultados qualitativos são complementados por dados numéricos (Pereira et al. 2018). Quanto a natureza, trata-se de uma pesquisa aplicada que visa solucionar problemas imediatos. Em relação aos objetivos é uma pesquisa descritiva e, do tipo bibliográfica considerando os procedimentos empregados.

Para a execução do trabalho foi realizado inicialmente uma pesquisa bibliográfica sobre o tema “análise de conteúdo” e como fazê-la. Assim, obteve-se o conhecimento que os dados para essa análise podem ser organizados com o auxílio de *software* específico, entre esses, o IRaMuTeq, um *software* livre que vem sendo utilizado como ferramenta de apoio ao processamento de dados na pesquisa com tipologia qualitativa.

O método de análise de conteúdo, trata-se de um conjunto de técnicas aplicadas para a análise de dados qualitativos. Segundo Fonseca (2017), o método de Análise de Conteúdo foi desenvolvido no princípio do século XX. Moraes (1999) afirma que originou no final do século XIX. Conforme Lima, Alongo & Ritter (2021) a Análise de Conteúdo aparece como uma das metodologias mais frequentemente utilizadas nos estudos publicados em dois periódicos Qualis-CAPES A1 no ensino de ciências, demonstrando ser um método de relevância neste campo de pesquisa.

Para Bardin (1977) a análise de conteúdo tem como finalidade a dedução de conhecimentos referentes às condições de produção (ou casualmente, de recepção), dedução está que lança mão de indicadores que podem ser quantitativos ou não.

De acordo com Bardin (1977), a Análise de Conteúdo divide-se em quatro etapas: a primeira é a pré-análise que corresponde à uma espécie de análise exploratória dos dados, é composta por uma leitura flutuante, referenciação dos índices e elaboração de indicadores e a organização do material como a escolha dos documentos e objetivos. Na segunda etapa, denominada codificação, ocorre a exploração do material, trata-se da mais longa e complexa, nela os dados brutos são transformados em representações do conteúdo presente no *corpus*. Na terceira etapa há o inventário e a classificação dos dados, por último ocorre a análise dos resultados por meio da inferência e interpretação lógica dos conteúdos encontrados nos textos.

Seguindo as etapas de análise de conteúdo, durante a pré-análise os materiais denominados *corpus* 1 contendo os resumos dos TCC dos alunos do ProfEPT e o *corpus* 2 contendo trechos da bibliografia indicada nos exames de acesso ao Programa foram analisados e as ideias iniciais foram sistematizadas. Assim, com a leitura flutuante se fez o contato inicial com os documentos, com vistas a conhecer os textos, separar, demarcar o que seria analisado e elaborar hipóteses e objetivos. Na sequência envolveu a exploração do material, consistiu identificação das unidades de registro e na definição de categorias e a exploração do material. Posteriormente, ocorreu o tratamento dos resultados dando o devido destaque para as informações analisadas, finalizando fazendo-se as inferências intuitivas e análise crítica e reflexiva.

Para a análise dos dados da pesquisa foram usadas, estatísticas textuais clássicas, classificação hierárquica descendente, pesquisa de especificidade de grupos de variáveis, análise de similitude e nuvem de palavras, todos construídos com o IRaMuTeq utilizando os dois *corpora* textuais.

Um *corpus* textual é um conjunto de textos ou um único texto, nessa pesquisa, o primeiro *corpus* corresponde a 87 resumos dos TCC dos estudantes do ProEPT que estavam disponíveis na Plataforma Sucupira no dia seis de dezembro de 2019 e, outro *corpus* contendo 18 trechos extraídos somente das referências bibliográficas, disponibilizadas nos editais que regem os exames de acesso ao programa. Assim, os 87 resumos e os 18 textos deram origem ao que é denominado de Unidade de Contexto Inicial (UCI).

Para a preparação do primeiro *corpus* textual, conforme indicado em Camargo e Justo (2013) e Salviati (2017), foi

usado um editor de textos, os 87 resumos de TCC foram organizados em um único arquivo. Cada resumo é iniciado com uma linha de comando numerada sequencialmente da seguinte forma (**** *resumo_1 até o **** *resumo_87). Posteriormente, foram excluídos dos textos os seguintes caracteres: aspas ("), apóstrofo ('), hífen (-), cifrão (\$), porcentagem (%), reticências (...), e asterisco (*). Este só foi usado nas linhas de comando que antecedem cada texto (resumo). Em seguida o arquivo foi salvo como documento de texto com a codificação de caracteres no padrão cp 1252 - Western Europe.

Da mesma forma foi criado o segundo *corpus* textual contendo trechos dos arquivos de dezoito referências bibliográficas indicadas para o exame nacional de acesso ao ProfEPT, para servir de comparação com as análises provenientes dos resumos.

Após a preparação dos textos, os dados denominados *corpus 1* e *corpus 2* foram importados para o *software* Iramuteq e aplicado os procedimentos e técnicas para a análise de conteúdo.

3. Resultados e Discussão

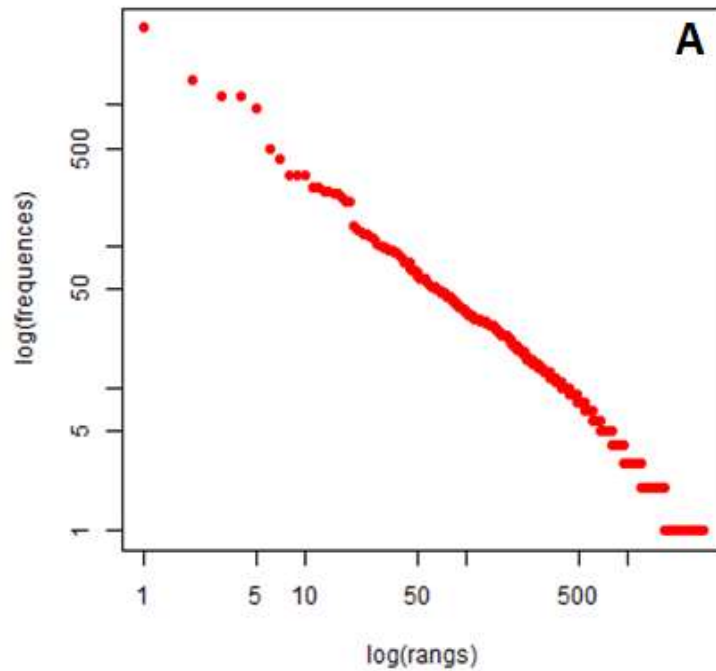
Com a importação dos *corpora* para o Iramuteq verificou-se no *corpus* textual com os resumos dos TCC, que o *software* reconheceu os 87 textos, reclassificou-os em 28.708 ocorrências (segmentos de textos) e em 2.913 formas, encontrou 1.232 hapax (palavras com única frequência) e em média 330 palavras em cada texto no *corpus 1*. No *corpus 2* (bibliografia) foram reconhecidos os 18 textos, que foram reclassificados em 10.101 segmentos de texto e em 1.767 formas, foram encontrados 918 hapax e detectado uma média de 561 palavras por texto.

O primeiro processamento ao qual os dois *corpora* textuais foram submetidos denomina-se estatísticas textuais clássicas. Nessa primeira análise estatística descritiva obteve-se o diagrama de Zipf, ilustração gráfica da distribuição de frequências que representam o comportamento das frequências de todas as palavras presentes nos *corpora* (Figuras 1 e 2). O eixo frequences (y) demonstra quantas vezes uma palavra e suas formas associadas (derivadas) aparecem, enquanto no eixo rangs (x) mostra a quantidade das mesmas.

Assim, na Figura 1 (*corpus 1*) observa-se que uma única palavra foi citada com uma frequência superior de mil vezes, em consulta a planilha gerada no *software* Iramuteq foi verificado que se trata da palavra 'de' e suas formas associadas (da/do/dos/das). Outras 5 palavras (a/e/em/o/que) e suas respectivas formas associadas foram repetidas com uma frequência superior a 500 vezes. Também, observando o eixo x da Figura 1, se infere que cerca de 1.200 formas apareceram nos TCC apenas 1 vez, aproximadamente outras 500 formas foram repetidas 2 vezes e que em torno de 260 formas foram citadas 3 vezes no *corpus 1*.

Na Figura 2, observa-se também que, a palavra 'de' e suas formas associadas (da/do/dos/das) foram citadas com uma frequência superior a 1.000 vezes. Nessa mesma figura é verificado que uma segunda forma foi usada um pouco mais de 500 vezes (a/à/á/as/às/aos/àqueles). Observando a Figura 2 no sentido do eixo x, verifica-se que cerca de 900 palavras apareceram apenas uma vez no *corpus 2*, outras 300 aparecem 2 vezes e aproximadamente 150 formas foram citadas 3 vezes.

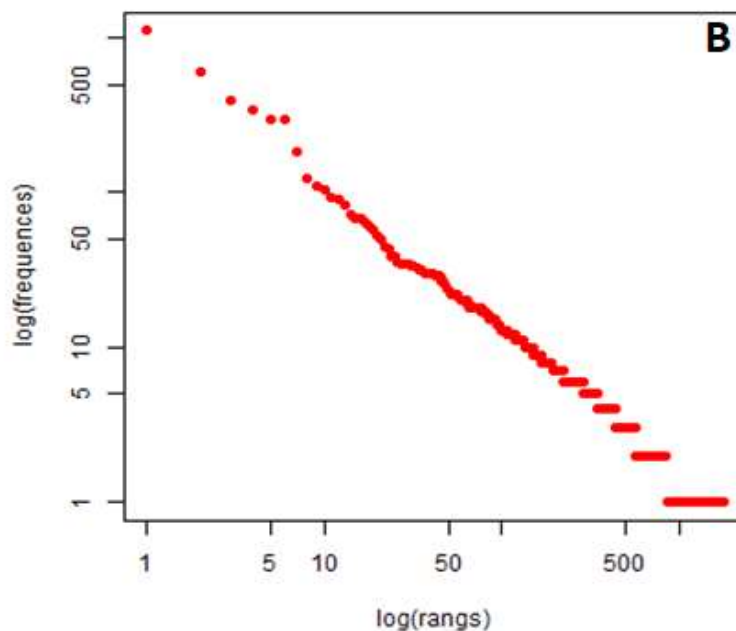
Figura 1 - Diagrama de Zipf de frequência das palavras (*corpus 1* - TCC).



Fonte: Autores.

Os gráficos apresentam a relação entre a frequência das formas (y) e a quantidade de formas (x). Observando no sentido de y pode-se afirmar que nas Figuras 1 e 2, há sete formas (cinco em 1 e duas em 2) que são mais frequentemente utilizadas, superando 500 vezes, enquanto cerca de 3.300 formas foram usadas com menor frequência, menos de três vezes. Assim, nos diagramas de Zipf observamos que existem muitas formas de palavras que repetem pouco e pouquíssimas que se repetem com muita frequência. Tendência também verificada em análises efetuadas em Carmo et al. (2018) e Santos et al. (2018).

Figura 2 - Diagrama de Zipf de frequência das palavras (*corpus 2* - bibliografias).



Fonte: Autores.

O segundo processamento realizado no Iramuteq é denominado de Nuvem de Palavras, trata-se do agrupamento e organização gráfica das palavras em função da frequência que aparecem nos textos, o resultado pode ser visualizado nas Figuras 3A e 3B. Esse método é visto como uma análise lexical simples, nesse processamento foram usadas apenas as 125 formas ativas com mais frequência nos resumos dos TCC (*corpus 1*) e as 125 formas ativas mais frequentes na bibliografia (*corpus 2*). Assim, considera-se que as palavras com fonte maiores tornam-se as mais relevantes porque foram utilizadas mais vezes em cada *corpus*. Vilela, Ribeiro e Batista (2020) descrevem a utilização da técnica Nuvem de Palavras na análise de dados qualitativos, usada para identificar desafios colocados aos mestrandos profissionais em ensino na saúde.

As palavras pesquisa, educação, profissional, ensino, produto e formação se destacam na Figura 3A, por serem bastante utilizadas nos resumos dos TCC, enquanto as palavras trabalho, educação, formação, social, educação profissional e ensino são as que mais se destacam na Figura 3B, por serem mais citadas nas referências do ProfEPT (*corpus 2*). As palavras Educação, Formação, Ensino, Trabalho e Prático são as que estão presentes nos dois *corpora* e se encontram entre as de mais destaque.

Figura 3 – Nuvem de palavras (A – *corpus 1* - TCC e B – *corpus 2* - bibliografias).



Fonte: Autores.

A palavra ‘educação’ é utilizada 349 vezes nos dois *corpora*. Nos resumos é citada isoladamente ou composta a palavra profissional e também é utilizada antes das palavras omnilateral, socialista, básica, articulada, politécnica. Nos textos da bibliografia do ProfEPT aparece associada mais frequentemente com as palavras trabalho, unitária, integral e inclusiva.

Nos dois *corpora* a palavra educação compõe frases e conceitos que traduzem o sentimento de que a educação deve ser articulada com o mundo do trabalho, promover a superação da ‘subordinação ao poder econômico’ e combater a ideologia de escola como aparelho que garanta a reprodução das relações sociais vigentes.

Outra forma utilizada para analisar os dados foi por meio da análise de similitude que representa através de indicadores estatísticos as ligações existentes entre as palavras em um *corpus*. A análise de similitude, ou de semelhanças, tem suas bases na teoria dos grafos, parte da matemática que trata das relações que ocorrem entre os objetos em um conjunto, ela possibilita identificar as ocorrências entre palavras (Salviati, 2017).

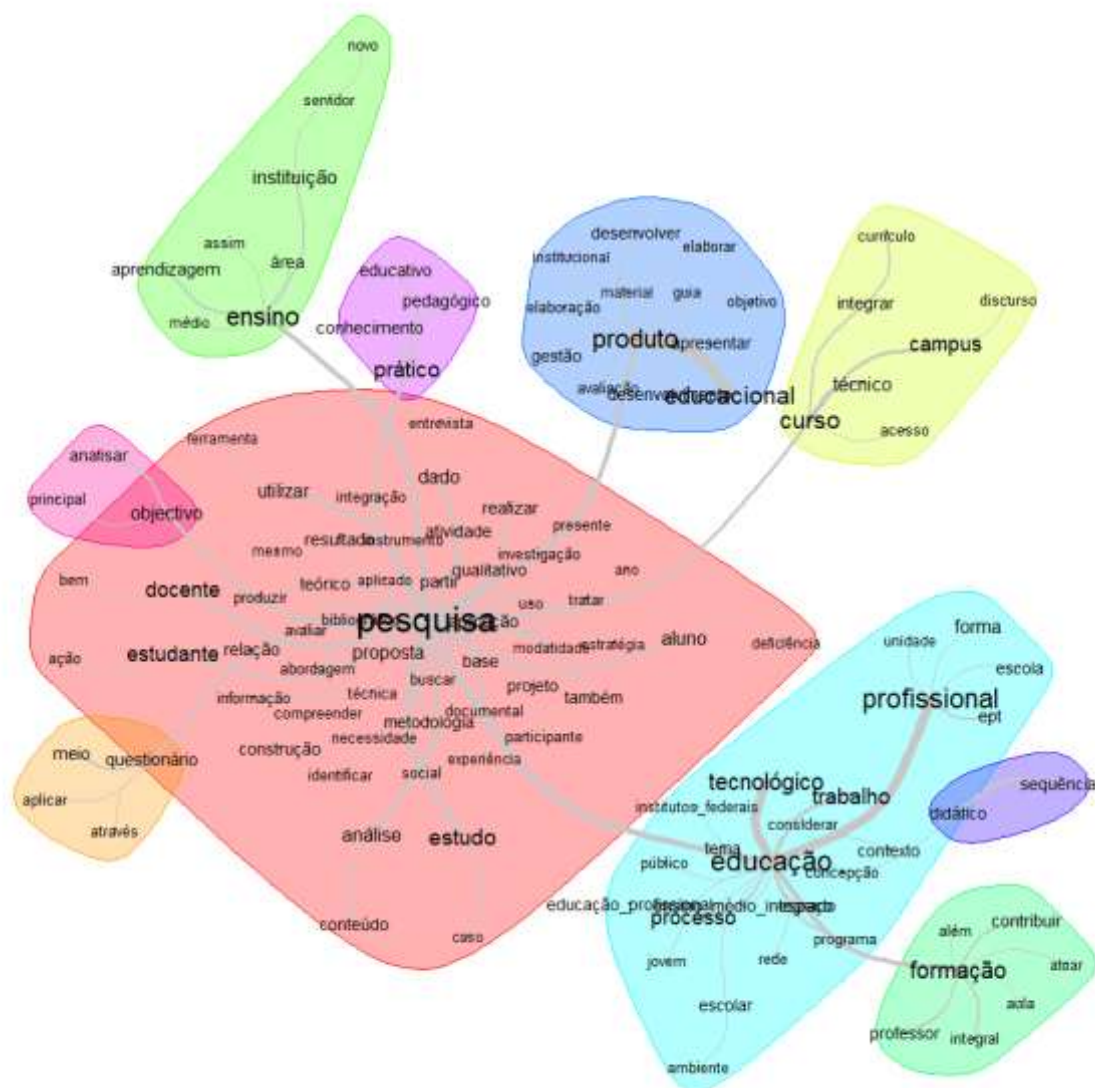
A análise de similitude construída com as 125 formas ativas mais frequentes nos resumos dos TCC (Figura 4) e nos

trechos das bibliografias do ProfEPT (Figura 5), nessas figuras é possível observar a conexidade existente entre as palavras presentes nos textos, nessa análise é importante observar na figura o tamanho da fonte, a espessura das linhas que ligam as palavras e o polígono a qual pertencem.

Na Figura 4 verifica-se as principais cocorrências entre as palavras e a conexidade entre os termos presentes nos textos do *corpus* 1, a palavra pesquisa conecta-se praticamente com todos os subgrupos. Conforme a árvore de cocorrência os resultados indicaram, que entre os pares de associação as relações são mais fortes entre as palavras: produto - educacional, educação - profissional, educação - tecnológico, pesquisa - educação, pesquisa - ensino e pesquisa - produto.

A análise de similitude na Figura 4 gerou 10 grupos, sendo que os mais expressivos revelaram uma forte relação entre os termos “pesquisa” e “educação”. Em um dos grupos fica bem descrita a necessidade dos mestrandos em desenvolver e ou elaborar os seus respectivos produtos educacionais o trecho a seguir foi extraído do *corpus* 1 e demonstra esta necessidade: “O objetivo deste estudo foi desenvolver e aplicar um produto educacional denominado de Sequência Didática na Técnica de Massagem Reflexologia Podal como instrumento de orientação docente para o processo ensino”.

Figura 4 – Análise de Similitude entre as palavras do *Corpus* 1 – TCC.

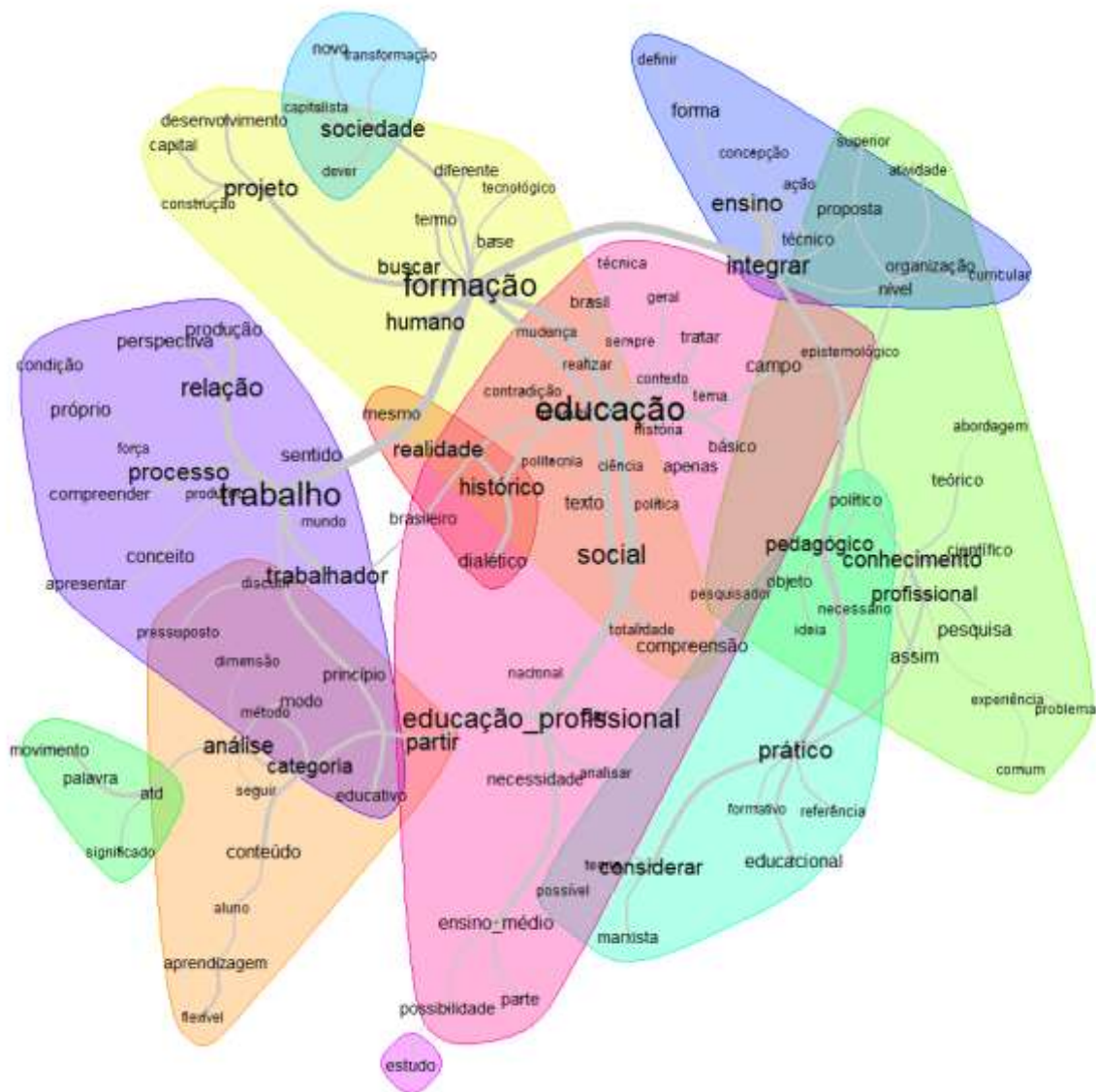


Fonte: Autores.

A análise de similitude permite entender a estrutura de construção do texto e temas de relativa importância, mostra as palavras próximas e distantes umas das outras, ou seja, forma uma árvore de palavras com suas ramificações a partir das relações guardadas entre si nos textos. Moura et al. (2012) utilizarão a análise de similitude para identificar os fatores associados à queda de idosos e, verificaram que os riscos estão relacionados com chão irregular, com as condições físicas e de saúde dos idosos.

Conforme indicado na árvore de coocorrência na Figura 5, produzida com os trechos da bibliografia do ProfEPT, a palavra educação tem papel importante na interligação dos grupos e subgrupos. Nessa figura observa-se que entre os pares de associação uma relação mais forte entre as palavras educação - educação - profissional, educação - formação, formação - integrar, formação - trabalho e prático - pedagógico.

Figura 5 – Análise de Similitude entre as palavras no *corpus 2* – Bibliografia.



Fonte: Autores.

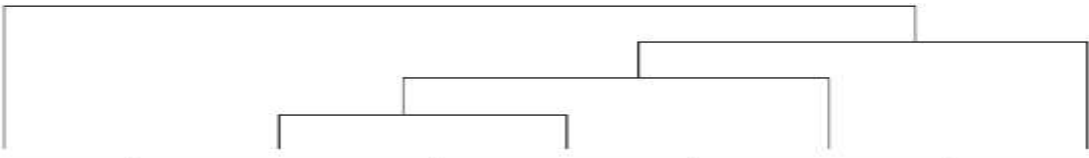
A análise de similitude do *corpus 2* mostrada na Figura 5, também apresentou 10 grupos, sendo que os mais expressivos revelaram uma forte relação entre os termos principais “trabalho” “formação” e “educação”. Um dos grupos nos

remete à concepção, forma, organização, propostas do ensino técnico integrado, nesse grupo observamos que essa preocupação se expressa na seguinte frase em um dos textos do corpus 2: “Refletir sobre o que é ou que pode vir a ser a formação integrada é o objetivo deste texto. O que é integrar? É tornar íntegro, tornar inteiro, o que?”

Outra forma de organizar compreender os dados pode ser feita empregando a Classificação Hierárquica Descendente (CHD) ou método de Reinert, este método apresenta a relação entre as classes de segmento de texto (st). Cada classe de segmentos de texto apresenta vocabulário semelhante entre si, e ao mesmo tempo, diferente do vocabulário dos segmentos de texto das outras classes. As Unidades de Contexto Elementar (UCE), ou segmentos de texto que compõem cada classe, são obtidas a partir das Unidades de Contexto Inicial (UCI).

O dendograma apresenta o resultado do processamento dos 87 textos do corpus 1. No Quadro 1, é possível visualizar o percentual de uso (%), o resultado do teste Qui-quadrado (X^2) e as principais palavras que formam as cinco classes semânticas identificadas no corpus 1. O dendograma demonstra a ligação entre as palavras que estão associadas entre si, isso permite interpretar as formações de cada classe, assim como, nos leva a compreender as aproximações e afastamentos entre as classes criadas.

Quadro 1 - Dendograma da Classificação Hierárquica Descendente (CHD) com a classificação do conteúdo do corpus 1



CLASSE 5 19.27%			CLASSE 2 21.71%			CLASSE 1 13.3%			CLASSE 3 28.29%			CLASSE 4 17.43%		
Palavra	%	X ²	Palavra	%	X ²	Palavra	%	X ²	Palavra	%	X ²	Palavra	%	X ²
questionário	89.66	202.73	educação	59.03	151.27	federal	72.22	55.71	produto	49.07	27.55	acesso	76.0	61.95
análise	78.12	158.01	tecnológico	63.92	119.35	oferta	100.0	52.78	sequência	70.0	26.97	ingresso	90.91	41.97
dado	75.76	150.61	profissional	55.12	103.47	jovem	59.26	51.57	prático	49.44	22.72	comunicação	80.0	41.76
qualitativo	90.0	137.05	profep	100.0	70.56	proejar	75.0	40.35	geral	81.25	22.68	tilsp	100.0	38.36
bibliográfico	86.21	87.42	mestrado	100.0	66.74	institutos_federais	51.85	36.29	curricular	84.62	20.75	glossário	100.0	38.36
documental	86.21	87.42	educativo	58.54	34.89	educação_profissic	45.95	36.24	avaliação	65.52	20.73	sistema	90.0	37.16
coleta	91.67	83.96	ept	55.56	32.56	adulto	72.73	34.26	educacional	47.0	20.38	tecnologia	68.42	35.35
utilizar	63.79	81.13	formação	43.43	32.38	adolescente	100.0	32.84	pedagógico	56.52	19.44	sala	70.59	34.27
entrevista	77.42	70.75	pós_graduação	100.0	29.2	compromisso	100.0	32.84	coletivo	90.0	19.07	competência	100.0	33.52
caso	83.33	65.75	inst_fed_farroupilha	81.82	23.78	básico	85.71	32.17	sugerir	100.0	17.94	animal	88.89	32.38
metodologia	60.47	50.23	pressuposto	100.0	21.83	expansão	85.71	32.17	componente	100.0	17.94	digital	81.82	32.23
instrumento	66.67	45.42	referencial	75.0	20.42	ano	55.56	28.65	projeto	60.61	17.89	socialização	81.82	32.23
aplicar	72.73	41.83	papel	72.73	17.13	rede	46.43	27.83	atividade	57.89	17.43	tomar	75.0	28.15
abordagem	65.52	41.73	autor	85.71	17.05	ifpr	75.0	26.73	interesse	88.89	16.52	informação	53.85	24.95
coletado	91.67	41.2	perspectiva	61.11	16.9	promoção	100.0	26.23	construção	56.41	16.17	tarefa	100.0	23.87
conteúdo	60.0	39.45	programa	53.85	16.45	trajetória	83.33	25.75	integrar	56.76	15.67	economia	100.0	23.87
aberto	100.0	38.24	integral	54.17	15.44	nível	57.14	23.84	significativo	76.92	15.47	ambientação	100.0	23.87
observação	90.91	36.92	contribuição	64.29	15.25	criação	52.94	23.78	conceitual	76.92	15.47	intuito	72.73	23.77
analisado	84.62	36.42	teórico	46.15	14.58	lei	66.67	22.53	teoria	76.92	15.47	uso	54.17	23.36
pesquisa	33.51	33.67	formal	100.0	14.51	administração	71.43	20.73	conhecimento	53.33	14.94	consciente	77.78	23.09
descritivo	90.0	32.67	especialmente	100.0	14.51	contextualizar	100.0	19.64	didático	55.56	13.96	facilitar	85.71	22.92

Fonte: Autores.

No relatório de processamento do software foi verificado que a Classe 1 contém 97 segmentos de texto correspondendo a 13,11% do total; a classe 2 com 141 segmentos (19,05%); a classe 3 com 208 (28,11%) abriga a maior parte do corpus; a classe 4 com 146 segmentos (19,73%) e a classe 5 com 148 segmentos, correspondendo a 20% (Quadro 1) dos 740 segmentos de texto aproveitados.

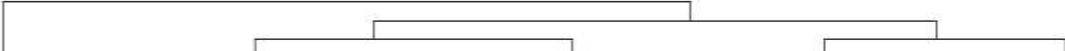
Com relação às classes geradas para o *corpus* em análise, observa-se no Quadro 1 duas ramificações, uma delas solitária, contendo apenas a classe 5 e outra, contendo as demais classes. Na primeira subdivisão encontra-se isolada a classe 4, a outra subdivisão se ramifica com a classe 3 também solitária e com as classes 1 e 2 numa mesma ramificação.

Os textos presentes nas classes 1 e 2 são os que apresentam mais aproximações entre si e, ao mesmo tempo, são mais distantes frente às demais classes. Pois, quanto mais afastado na ramificação da CHD, menores serão as relações entre as palavras. Quanto mais aproximadas, maior a afinidade contextual entre as classes. Nas classes 1 e 2 verifica-se formas vinculadas ao contexto dos Institutos Federais, educação profissional e tecnológica (EPT). Nas classes 3 e 4, fica demonstrado as preocupações dos mestrandos com o produto educacional exigido para a conclusão do mestrado. Silva & Abreu (2018) utilizando a CHD identificaram o conteúdo lexical e os segmentos de texto característicos de cada classe e nomearam-nas de acordo com a frequência das palavras que apareceram nas reportagens do site Exame.com considerando sua relevância com as ações de responsabilidade social corporativa das empresas.

O teste Qui-quadrado (χ^2) apontou maior grau de significância estatística para as palavras na classe 5. Indicativo de que esses termos possuem forte relação com esta classe. Assim a classe 5 foi a mais prevalente, nessa classe as palavras estão mais relacionadas com os métodos usados nas pesquisas. Na tabela de processamento do Iramuteq se apurou que a classe 5 é caracterizada pelas palavras: questionário ($\chi^2 = 202,7$), análise ($\chi^2 = 158,0$), dado ($\chi^2 = 150,6$) e qualitativo ($\chi^2 = 137,0$).

O dendograma (Quadro 2) apresenta o resultado do processamento dos 18 textos do *corpus* 2. Pela CHD, foram identificadas cinco classes semânticas no *corpus* 2. Classe 1 com 43 segmentos de texto correspondendo a 20,19%; classe 2 com 40 segmentos, representando 18,78%; classe 3 com 40 segmentos de texto (18,78%); classe 4 com 47 segmentos, indicando 22,07%; e classe 5 com 43 segmentos que significam 20,19% dos 213 segmentos de texto aproveitados.

Quadro 2 - Dendograma da Classificação Hierárquica Descendente (CHD) com a classificação do conteúdo do *corpus* 2



CLASSE 4 22.07%			CLASSE 5 20.19%			CLASSE 2 18.78%			CLASSE 3 18.78%			CLASSE 1 20.19%		
Palavra	%	X ²	Palavra	%	X ²	Palavra	%	X ²	Palavra	%	X ²	Palavra	%	X ²
teórico	71.43	21.23	pedagógico	82.61	62.35	partir	65.0	30.92	termo	90.91	39.56	força	100.0	28.61
teoria	87.5	20.7	ensino	66.67	46.82	educativo	81.82	30.22	tema	90.0	34.89	sociedade	60.0	27.86
possível	87.5	20.7	integrar	62.5	41.84	categoria	62.5	21.68	politecnia	100.0	26.7	homem	100.0	20.24
âmbito	100.0	18.08	ensino_médio	68.75	25.32	sistematizar	85.71	21.26	tornar	85.71	21.26	contraditório	100.0	16.12
objeto	85.71	17.05	prático	51.61	22.24	educação_profissional	48.28	19.15	enfrentamento	100.0	17.63	mudança	83.33	15.28
estudo	85.71	17.05	técnico	72.73	19.87	princípio	66.67	19.12	humano	46.15	14.55	modo	63.64	13.59
comum	77.78	16.96	organização	85.71	19.29	objetivo	100.0	17.63	sugerir	100.0	13.16	política	80.0	11.37
dialético	66.67	14.71	possibilidade	70.0	16.16	escolar	100.0	17.63	capacidade	71.43	13.15	presente	80.0	11.37
metodológico	100.0	14.4	procedimento	100.0	16.12	dualidade	83.33	16.87	ação	60.0	11.69	momento	80.0	11.37
senso	100.0	14.4	integrador	100.0	16.12	epistemológico	80.0	12.58	pensar	62.5	10.42	análise	47.37	9.57
acumular	100.0	14.4	finalidade	100.0	16.12	permitir	80.0	12.58	educação	35.71	9.84	mundo	62.5	9.24
problema	75.0	13.54	indicação	100.0	16.12	campo	50.0	12.56	ver	66.67	9.28	realizar	62.5	9.24
materialismo	75.0	13.54	favorecer	100.0	16.12	marxista	62.5	10.42	divisão	66.67	9.28	transformação	62.5	9.24
conhecimento	50.0	13.44	nível	61.54	14.69	geral	62.5	10.42	trabalho	33.33	8.61	capitalista	62.5	9.24
experiência	66.67	10.87	ideia	66.67	12.6	projeto	40.74	9.78	objectivo	75.0	8.45	trabalho	35.42	8.92
pesquisador	100.0	10.75	comprometer	100.0	12.03	tender	66.67	9.28	afirmar	75.0	8.45	capital	54.55	8.5
sujeito	100.0	10.75	integração	71.43	11.79	sistema	75.0	8.45	vínculo	75.0	8.45	defender	75.0	7.6
entender	100.0	10.75	conteúdo	58.33	11.48	acumulação	75.0	8.45	reconhecer	75.0	8.45	espaço	75.0	7.6
linguagem	100.0	10.75	curricular	80.0	11.37	regime	75.0	8.45	colocar	75.0	8.45	sentido	46.67	7.02
perceber	100.0	10.75	médio	80.0	11.37	educação	33.33	7.27	relação	37.04	6.76	contradição	57.14	6.13
mestrado	100.0	10.75	decreto	80.0	11.37	concepção	57.14	6.98	social	33.33	6.02	produção	43.75	5.96

Fonte: Autores.

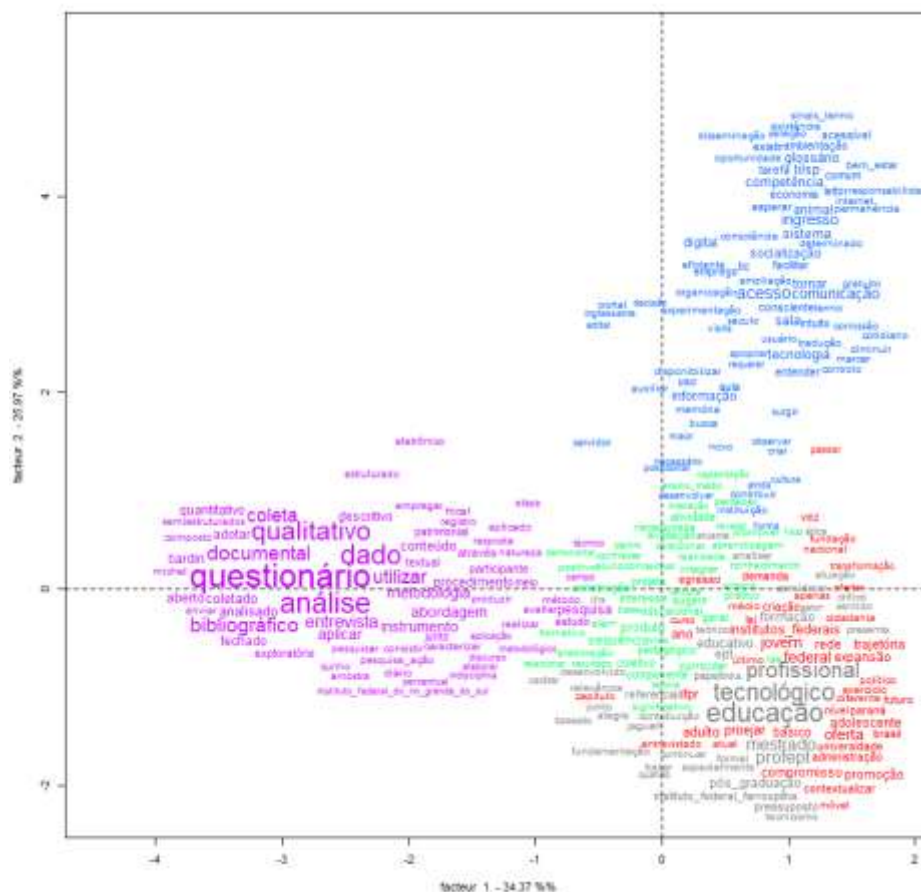
No *corpus* 2, o teste Qui-quadrado (χ^2) também apontou maior grau de significância estatística para as palavras na classe 5. Indicativo de que esses termos estão mais relacionados com esta classe do que com as outras. Assim a classe 5 foi a

mais prevalente. Na classe 5 as palavras estão mais relacionadas com o processo de ensino, no Quadro 2 observa-se que é caracterizada pelas palavras: pedagógico ($\chi^2 = 62,3$), ensino ($\chi^2 = 46,8$), integrar ($\chi^2 = 41,84$) e ensino médio ($\chi^2 = 25,3$).

Nas Figuras 6 e 7, são apresentados os aspectos da classificação hierárquica descendente numa visão bidimensional, por meio da análise fatorial de correspondência (AFC). Nesse plano as aproximações e ou distanciamentos entre as classes podem ser identificados com precisão de acordo com a disposição nos quadrantes.

Na Figura 6 observamos maior aproximação entre as classes 1(vermelho), 2 (cinza) e 3 (verde) e o afastamentos das classes 5 (rosa) e 4 (azul). No plano cartesiano, ficam demonstradas as palavras formadoras de cada classe e o grau de importância de cada uma, com destaque para as palavras da classe 5 (questionário, análise, qualitativo e dado) e na classe 2 as palavras educação, tecnológico e profissional.

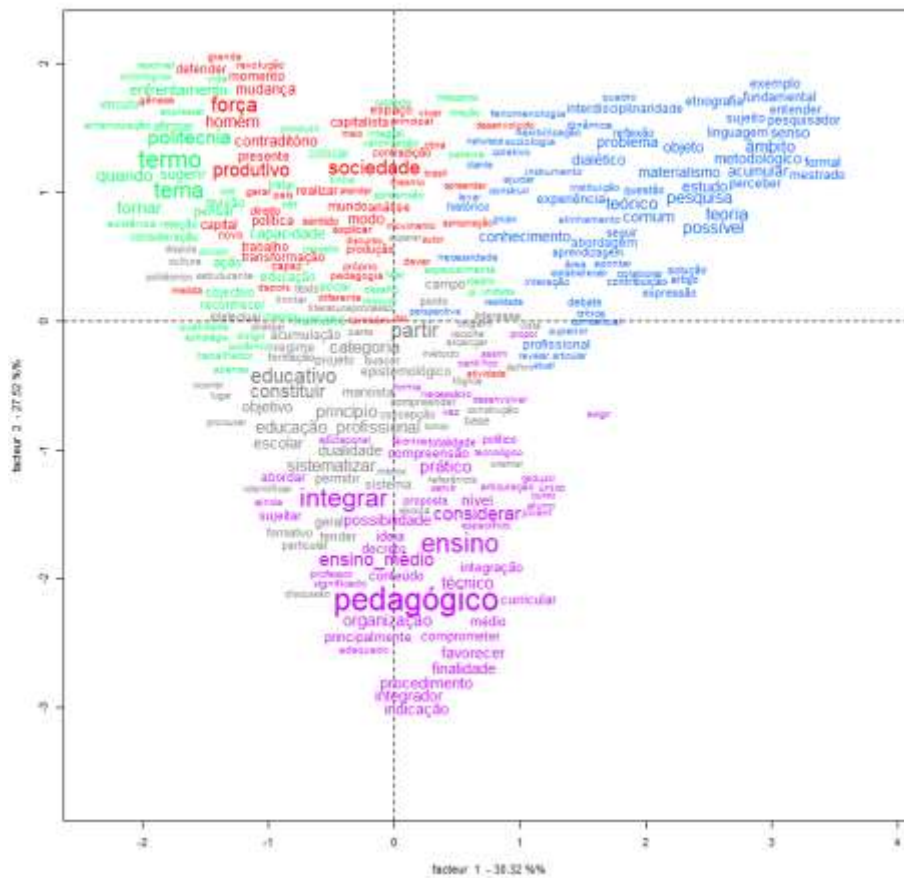
Figura 6 – Análise fatorial de correspondência (AFC) das palavras ativas mais frequentes em cada uma das classes lexicais obtidas na classificação hierárquica descendente no *corpus* 1 (TCC).



Fonte: Autores.

Na Figura 7 observamos maior aproximação entre as classes 1(vermelho) e 3 (verde), também entre as classes 2 (cinza) e 5 (rosa) e o afastamento da classe 4 (azul). Nesse plano destacam as palavras da classe 5 (pedagógico, ensino e integrar) e na classe 2 as palavras educação, tecnológico e profissional.

Figura 7 – Análise fatorial de correspondência (AFC) das palavras ativas mais frequentes em cada uma das classes lexicais obtidas na classificação hierárquica descendente no *corpus 2* (bibliografias).



Fonte: Autores.

4. Considerações Finais

A aplicação do *software* IRaMuTeQ possibilitou realizar vários tipos de análises, dentre elas as do tipo Lexográfica, Especificidades, Método da Classificação Hierárquica Descendente (CHD), Análise fatorial de correspondência (AFC, Análise de Similitude e Nuvem de Palavras.

Os resultados das análises dos dados selecionados, *corpus 1*– TCC e *corpus 2* - referências teóricas, realizadas com aplicação do *software* livre IraMuTeq demonstram que as dissertações produzidas no âmbito do ProfEPT aqui analisadas apresentam conteúdos que convergem para os mesmos temas encontrados na bibliografia que contém as bases conceituais da educação profissional e tecnológica no Brasil.

Assim, podemos concluir que os trabalhos de conclusão de curso do ProfEPT, produzidos até dezembro de 2019 seguem a mesma linha dos estudos produzidos pelos principais autores que servem de referência ao programa, pois aqueles contemplam em diversos momentos os principais termos utilizados nas referências teóricas do programa.

O *software* livre IRaMuTeq mostrou-se bastante eficiente em organizar os dados coletados e demonstrar os resultados conforme o objetivo de cada análise, se de similitude ou classificação hierárquica, ou frequência no uso dos termos, conforme cada análise feita, ou seja, o *software* auxilia sobremaneira na demonstração dos resultados.

A técnica de análise de conteúdo vai além da organização e demonstração dos dados, ela apresenta-se em etapas, conforme Bardin (1977), a primeira que abrange uma pré-análise com leitura fluente e organização do material; a segunda,

mais longa e complexa é a codificação das fontes, a terceira trata da classificação e por fim, trata da discussão dos resultados, por meio da inferência e interpretação.

O *software* IRaMuTeq é um grande aliado para a realização da etapa dois, que é bastante trabalhosa e, por isso, demanda muito tempo, se feita manualmente. Já na etapa final, quando os resultados demonstrados na etapa dois são discutidos e interpretados, é primordial o protagonismo do pesquisador. É ele que deverá, a luz das suas leituras e estudos, inferir, discutir e interpretar os dados organizados e demonstrados pelas diversas análises resultantes da submissão dos dados ao IRaMuTeq.

Portanto, a utilização do *software* livre IRaMuTeq é de grande valor para o desenvolvimento de uma pesquisa que se utilize da análise de conteúdo, mas ele por si só, não realiza a tarefa de discutir e interpretar os resultados obtidos, esta é prevalente do pesquisador.

Sugerimos que sejam realizados estudos complementares que explorem as demais possibilidades, de uso do IRaMuTeq e de análise dos produtos gerados por ele.

Referências

- Bardin, L. (1977). *Análise de Conteúdo*. Edições 70.
- Camargo, B. V. & Justo, A. M. (2013). *Tutorial para uso do software de análise textual IRAMUTEQ*. Laboratório de Psicologia Social da Comunicação e Cognição, Universidade Federal de Santa Catarina. Recuperado de: <http://www.iramuteq.org/documentation/fichiers/tutoriel-en-portugais>
- Campos, C. J. G. (2004). Método de análise de conteúdo: ferramenta para a análise de dados qualitativos no campo da saúde. *Revista Brasileira de Enfermagem*, 57(5), 611-614. <https://doi.org/10.1590/S0034-71672004000500019>
- Carmo, H., Braum, L., & Sá, E. (2018). Descontinuidade tecnológica em patentes envolvendo o uso de cinzas de carvão: análise baseada em conteúdo textual. *Revista Metropolitana de Sustentabilidade (ISSN 2318-3233)*, 8(2), 49-66. <http://revistaseletronicas.fmu.br/index.php/rms/article/view/1273>
- Fonseca, P. A. A. da. (2017). Poder e Discurso: Uma análise de conteúdo do discurso de posse dos Ministros das Relações Exteriores do Brasil (2003-2016). 9º Congresso Latino-Americano de Ciência Política, organizado pela Associação Latino-americana de Ciência Política (ALACIP). Montevidéu, 26 ao 28 de julho de 2017.
- Freire, P. (2005). *Pedagogia do Oprimido*. Paz e Terra.
- Freire, P. (1996). *Pedagogia da autonomia: saberes necessários à prática educativa*. Paz e Terra.
- Frigotto, G. (2009). A polissemia da categoria trabalho e a batalha das ideias nas sociedades de classe. *Revista Brasileira de Educação*, 14(40), 168-194. <https://doi.org/10.1590/S1413-24782009000100014>
- Frigotto, G., Ciavatta, M., & Ramos, M. (orgs.). (2005). *Ensino médio integrado: concepção e contradições*. Cortez.
- Gil, A. C. (2008). *Métodos e técnicas de pesquisa social*. Atlas.
- Kuenzer, A. Z. (Org.). (2007). *Ensino Médio: construindo uma Proposta para os que vivem do Trabalho*. São Paulo: Cortez.
- Libâneo, J. C., (1983). Tendências pedagógicas na prática escolar. *Revista da Ande*, 3(6), 11-19. https://praxistecnologica.files.wordpress.com/2014/08/tendencias_pedagogicas_libaneo.pdf
- Lima, F. O., Alonço, M., & Ritter, O. M. S. (2021). A análise de conteúdo como metodologia nos periódicos Qualis-CAPES A1 no Ensino de Ciências. *Research, Society and Development*, 10(3), e43110313378. <https://doi.org/10.33448/rsd-v10i3.13378>
- Manacorda, M. A. (1990). *O princípio educativo em Gramsci*. Artmed.
- Marx, Karl. (1968). *O Capital*. Civilização Brasileira.
- Moll, J. et al. (2010). *Educação profissional e tecnológica no Brasil contemporâneo: desafios, tensões e possibilidades*. Artmed.
- Moraes, R. (1999). Análise de conteúdo. *Revista Educação*, 22(37), 7-32.
- Moura, S. R. B., Marques Junior, M. A. S. S., Rocha, A. K. L., Vieira, J. P. P. N., Mesquita, G. V., & Brito, J.N.P.O. (2015). Análise de similitude dos fatores associados à queda de idosos. *Revista Interdisciplinar*, 8(1), 167-173.
- Pacheco, E. (20015). Fundamentos Político-Pedagógicos dos Institutos Federais: diretrizes para uma educação profissional e tecnológica transformadora. Natal: IFRN. <https://memoria.ifrn.edu.br/handle/1044/1018>
- Paro, V. H. (2007). *Gestão Escolar, Democracia e Qualidade do Ensino*. Editora Ática.
- Pereira, A. S. et al. (2018). *Metodologia da Pesquisa Científica*. https://repositorio.ufsm.br/bitstream/handle/1/15824/Lic_Computacao_Metodologia-Pesquisa-Cientifica.pdf?sequence=1

Ratinaud, P. (2009). IRAMUTEQ: Interface de R pour les Analyses Multidimensionnelles de Textes et de Questionnaires [Computer *software*]. <http://www.iramuteq.org>

Saviani, D. (1989). *Sobre a concepção de politecnia*. Fiocruz.

Salviati, M. E. (2017). *Manual do aplicativo Iramuteq*. Planaltina. <http://www.iramuteq.org/documentation/fichiers/manual-do-aplicativo-iramuteq-par-maria-elisabeth-salviati>

Santos F. M. (2012). Análise de conteúdo: a visão de Laurence Bardin. *Rev eletrônica Educ [on-line]*. 6(1). 383-387. <http://www.reveduc.ufscar.br/index.php/reveduc/article/viewFile/291/156>

Santos, A. L. S. dos, Pereira. I. V., Ferreira. A. M. R. & Palmeira. I. P. (2018). Percepções de portadores de hanseníase sobre as reações hansênicas e o cuidado de si. *Revista Pan-Amazônica de Saúde*, 9(4), 37-46. <https://dx.doi.org/10.5123/s2176-62232018000400004>

Silva, F. M. C. da & Abreu, J. C. A. de. (2018). A responsabilidade social corporativa analisada em Exame.com. *DESENVOLVE: Revista de Gestão do Unilasalle*, 7(2), 09-22. <http://dx.doi.org/10.18316/desenv.v7i2.4389>.

Silva, A. H. & Fossá, M. I. T. (2013). Análise de conteúdo: exemplo de aplicação da técnica para análise de dados qualitativos. 2013. In: IV Encontro de Ensino e Pesquisa em Administração e Contabilidade. Brasília, DF.

Vilela, R. B., Ribeiro, A. & Batista, N. A. (2020). Nuvem de palavras como ferramenta de análise de conteúdo: uma aplicação aos desafios do ensino no mestrado profissional. *Millenium*, 2(11), 29- 36. <https://doi.org/10.29352/mill0211.03.00230>.

Zabala, A. (1998). *A prática educativa: como ensinar*. ArtMed.

Zabala, A. (1999). *Como trabalhar os conteúdos procedimentais em aula*. (2a ed.), ArtMed.